REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author OAUEM M

A thesis submitted for the degree of Doctor of Philosophy of the University of London

DATA-DRIVEN METHODS FOR PROCESS ANALYSIS

by

Margret Bauer

Department of Electronic & Electrical Engineering

University College London

London

August 2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The thesis is concerned with the development of data-driven methods for fault diagnosis of plant-wide disturbances. Industrial plants producing large quantities of liquid or gaseous chemicals run continuously and under tight cost, safety, quality and environmental constraints. Any unwanted variability in the form of disturbances in the production process affects one of these constraints. Worse still, disturbances can spread and cause large parts of the process to be upset. Detection and diagnosis of disturbances is therefore an important subject for chemical companies. Chemical processes are well equipped with modern instrumentation technology so that measurements of process variables such as flow, temperature or pressure are abundantly available.

The research has used time series analysis for process measurements in a novel way. In particular, it focuses on measures to decide about cause and effect of processes variables to address the question whether A causes B or B causes A. Knowing the causal relationship finds the fault propagation path in case of a disturbance and eventually traces the disturbance back to the root cause. Three different measures are investigated and developed for the application to chemical process data: one straight forward algorithm based on the cross-correlation function and two statistics based on nearest neighbours methods and transfer entropy. Together with the automatic generation of causal maps these approaches lead to a breakthrough in fault diagnosis. Guidelines for the parameters of the methods are tailored to signatures caused by disturbances common in chemical processes. A significance level is introduced for automatic implementation of industrial applications.

Case studies with process disturbances, in particular those from a three months placement with Eastman Chemical Company, are analysed with the developed tools. The results are compared and recommendations of choosing the best method for a data set are generalised from results of the case studies.

Acknowledgements

I am most grateful to my supervisor, Professor Nina Thornhill. Her invariably constructive ideas and endless energy provided me with constant support. I had the necessary freedom to make my own choices but also my own mistakes. Nina introduced me to the process control community so I had close contact to industry which is essential for all engineering work. I hold her in high regard as a researcher, teacher and friend.

During the PhD I spent two months at Eastman Chemical Company in Kingsport, Tennessee. I owe a major debt of gratitude to everyone in Jim Downs' group: John Cox, Michelle Caveness, Ernie Vogel, Chip Anderson, Jacob Cheshier, Jerry Cole, Laura Johnson, Brian Joyner, Steve Miller, Ken Yount, and John Twork. Thank you for all your time and patience in explaining real chemical processes to an electrical engineer. At BP Chemicals, I am grateful for the discussions with Adrian Meaburn. Alexander Horch at ABB asked challenging questions and confronted me with the problem of two interacting controllers. I benefited greatly from the University College London Graduate School through a generous scholarship and numerous high quality courses. The Institute of Electrical Engineers sponsored my field work at Eastman Chemical Company.

Friends and colleagues challenged and contributed to my work. At UCL, I thank Ioannis Papakonstantinou for the times we discussed the really important things over coffee. Michael Düser gave me essential backup, always at the right moment. Amine Houyou, Lamia Benabbas, Kai Wang and Teruyuki Kataoka made room 708 a great place. I am indebted to Izzat Darwazeh for his invaluable advice. Professor Sirish Shah and Shoukat Choudhury at the University of Alberta gave me extremely helpful and constructive feedback. Thanks to Mo Bill Chen at Imperial College for long discussions about predictability.

Many friends were also PhD candidates and we lived through the same highs and lows, among them: Melanie Etherton, Jürgen Biela, Christian Forster, Catherine Gegout, Nishita Hathi, Cynthia Kinnan, Priya Kissoon, Dario Milo and Uta Protz. I could not have done this work without your constant encouragement and friendship. Thanks for innumerable dinners and fantastic events at Goodenough College: Alisha, Arkaja, Barry, Benedetto, Carina, Catalina, Chiro, Collin, Elisabetta, Francesco, Jürg, Mandira, Marco, Nadia, Nikos, Nisha, Noriko, Pamela, Peter, Roeb, Saadia, Sofia, Tina and more. Renate and David Widdup spurred my love for London and helped me out in all crucial situations.

'No excellent soul is exempt from a mixture of madness.' Faizel Ismail turned my life upside down (north to south) and from normal to extraordinary.

My greatest debts, however, are to my parents, for their love and support throughout my studies. Ngiyabonga (I thank) my brother Wolfhart and sister Annette for being there whenever I need you.

I dedicate this thesis to the strong women of my family.

Contents

List of Figures

List of Tables

Nomenclature

A	Amplitude of sine wave; area of tank; area of pipeline
CompDev	Compression deviation blanket
CompMax	Maximum compression time
CompMin	Minimum compression time
D_i	Constructional function for nearest neighbours method
F	Flow rate
$G(X)$	Entropy
$G(X\|Y)$	Conditional Entropy
$G_c(s)$	Transfer function
H	Nearest neighbours causality measure
H^o	Nearest neighbours causality measure (OSAP)
$I(X,Y)$	Mutual information
I_{ent}	Differential entropy
K, K_1	Number of nearest neighbours, Kernel function
K_c	Filter gain
L	Time interval
N	Number of samples
$N_{\min}$	Minimum number of samples
N_p	Number of samples per cycle
N_s	Number of surrogates
N^*	Number of embedded vectors
$P(\cdot)$	Probability distribution
R	Resistance of flow
$\mathbf{R}$	Matrix containing indices of nearest neighbours $r_{i,j}$
R_{ent}	Entropy ratio
$\mathbf{S}$	Matrix containing indices of nearest neighbours $s_{i,j}$
$T(X\|Y)$	Transfer entropy
T_1	Time interval over which function x is defined
T_d	Dead time
T_p	Period of oscillation
V	Volume
X, Y	Random variable
a_ν, b_ν	Filter coefficients
c	Constant
d	Derivative
$f(\cdot)$	Functional relationship
f_n	Discrete frequency
f_q	Quantisation factor
h	Prediction horizon
$\mathbf{h}$	Causality matrix
$\overline{h}$	Histogram bin width; Kernel function width
h_l	Level in tank
$\overline{h}_{\text{opt}}$	Optimal width $\overline{h}$
$h_{x\to y}$	Causality measure (nearest neighbours)
i	Discrete time index
j	Nearest neighbours parameter; index for amplitude bin
k, l	Embedding dimension (transfer entropy)
l_p	Length of pipeline

m	Embedding dimension (nearest neighbours)
n	Number of discrete amplitude bins (histogram)
n_f	Filter order
n_j	Number of samples per amplitude bin (histogram)
p	Number of process variables
$p(\cdot)$	Probability density function
$\hat{p}(\cdot)$	Estimated probability density function
r^{comp}	Compression ratio
$r_{i,j}$	Index of nearest neighbours
r_{xy}	Linear correlation coefficient
$r^{\max}$, r_{thresh}	Correlation index in CCF algorithm and threshold value
s	Laplace continuous frequency
$s_{i,j}$	Index of nearest neighbours
$s_{x\to y}$	Significance level
t	Continuous time index
$t_{x\to y}$	Transfer entropy causality measure
u	Input time series
v	Velocity
w_{in}, w_{out}	Inflow and outflow
x	Process variable measurement and time series
$\mathbf{x}$	Embedded vector
$\overline{x}$	Linear mean
x^{quant}	Quantised time trend
x^{restored}	Recovered compressed time trend
x_{surr}	Surrogate time series of x
y	Process variable measurement and time series
y_{surr}	Surrogate time series of y
Δ	Causality matrix (positive values only)
$\Delta\phi$, $\Delta\kappa$	Parameters for CCF algorithm by Horch [45]
Θ	Step function
α	Probability
δ	Delta-function; delta in delta-epsilon method
ϵ	Epsilon in delta-epsilon method
ϵ^{comp}	Compression error
ϵ^{quant}	Quantisation error
η	Index of process variables
κ, κ_1	Discrete time delay
$\kappa^{\max}$, $\kappa^{\min}$	Indices for CCF algorithm in this thesis
λ	Detected time delay
λ_0, λ_j	Index for hypothesis testing
μ	Linear mean
ν	Filter coefficient index
ρ	Index of process variables
ϱ, $\Delta\varrho$	Proximity radius of delta-epsilon method, threshold
σ, σ^2	Standard deviation, variance
τ, τ_1	Continuous time delay
τ_D, τ_I	Derivative and integral dynamic time constant
ϕ_{xx}, ϕ_{xy}	Autocorrelation function, cross-correlation function
$\phi^{\max}$, $\phi^{\min}$	Indices for CCF algorithm in this thesis
χ	Discrete amplitude grid (Kernel estimation)
ψ, ψ_{thresh}	Oscillation index in CCF algorithm and threshold value

Chapter 1

Introduction

This chapter gives an introduction to the subject matter of data-driven methods in process analysis. The motivation for data-driven analysis of process disturbances is given first. Existing approaches and techniques are discussed thereafter. Lastly, the contributions of this thesis are outlined and the organisation of the thesis is laid out.

1.1 Motivation

Industrial chemicals are produced in large quantities in plants that operate continuously over weeks and months. During operation, the process conditions are sought to be constant. For example, the temperature in a reaction tank should be 60 °C because the ingredients react best at this temperature and not at 59.9 °C or 60.1 °C. The most important physical quantities of the process are therefore regulated by an automatic controller. However, changes in process conditions such as the outside temperature or equipment wear might affect the process conditions and cause disturbances in the physical quantities of a process that the controller cannot tackle. Research and development in the process industries is focused on the fast and systematic elimination of disturbances and thus reduction of variability. Reduced variability is in most cases linked to the economic performance of the process.

Physical quantities such as temperature, flow, level and pressure are measured automatically at fixed time instants. The measured value may be used to adjust a control element, for example a flow valve or a pump. Also, the value can be used to monitor the physical quantity and raise an alarm if it is above or below a specified threshold. All measured data samples are recorded automatically in a centralised data base. Thus, time trends of the physical quantities, also referred to as process variables in the following, are readily accessible.

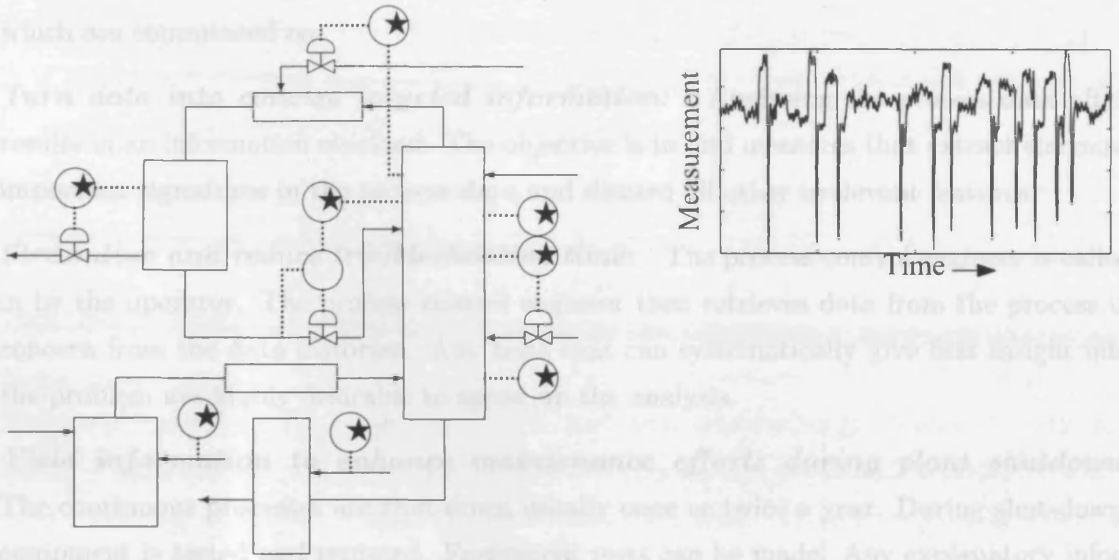


Figure 1.1: A disturbance affecting a number of process measurements in an industrial case study.

The first step to eliminate the disturbance is detection. If only one variable is affected then the root cause of the disturbance can be eliminated immediately. On frequent occasions, however, the disturbance travels with the process flow and affects a number of measurements. The task then becomes a plant-wide problem. For plant-wide disturbances, fault diagnosis is required because the variable at which the disturbance was first detected might not be the root cause. The left hand panel of Figure 1.1 illustrates the process schematic of a chemical process[1]. The circles indicate the position at which a measurement is taken. The stars inside the circles indicate that these measurements are affected by the same disturbance and a sample time trend of the disturbance is shown on the right hand panel of Figure 1.1. The time trend is clearly not constant and will affect the performance of the process. Root cause analysis is hence concerned with finding the process measurement that is closest to the originating point of the disturbance.

A number of different root causes can lead to a plant wide disturbance. Desborough and Miller [21] at Honeywell split the causes in the categories of process or constraint problems, controller tuning problems and valve problems. Valve problems and tuning problems have been addressed in numerous previous publications but only few approaches exist for process and constraint problems. A specific issue in the category of process and constraint problems is controller interaction in recycles. Adrian Meaburn, a process control engineer at BP Chemicals, estimates the severity of the this issue as follows:

> "In my experience about 30% of chemical processes with recycles have suffered from controller interaction problems."

The objective is now to employ the abundantly available process data to detect and diagnose disturbances and identify the root cause. John Cox at Eastman Chemical Company has provided the following motivations for data-driven disturbance analysis at the outset

[1] How to interpret a process schematic is explained in Section 2.2.

which are commented on.

Turn data into concise targeted information: Analysing the process data often results in an information overload. The objective is to find measures that extract the most important signatures in the process data and discard all other irrelevant features.

Streamline and reduce troubleshooting time: The process control engineer is called in by the operator. The process control engineer then retrieves data from the process in concern from the data historian. Any tests that can systematically give first insight into the problem are highly desirable to speed up the analysis.

Yield information to enhance maintenance efforts during plant shutdown: The continuous processes are shut-down usually once or twice a year. During shut-down, equipment is tested and replaced. Equipment tests can be made. Any explanatory information about plant problems helps to focus the maintenance.

Discover problems not found with traditional "fight today's fire" approaches: Problems in the plant might exist of which neither the operator or the process engineer is aware of. Methodologies that can pinpoint disturbances which have escaped the standard investigated tools are extremely useful.

1.2 Introduction to Data-Driven Fault Analysis

A brief review of the work in the area of data-driven methods for process analysis that is most relevant for this thesis is outlined in this section. More detailed literature reviews will be given in Part I of the thesis. Approaches to fault detection and diagnosis of plant-wide disturbances are split in two branches: model-based and data-driven. Model-based techniques compare model parameters against measurements, whilst data-driven methods are based on historical process data. An overview of qualitative model based methods is given in Section 3.4. The focus of this work is on data-driven methods since in the chemical industries only about 5% of the processes[2] are modelled. The organisation of process monitoring and fault analysis techniques is discussed in Section 3.2 while data-driven methods are described in more detail in Section 3.3.

The methods investigated in this thesis are usually grouped in the area of multivariate statistics and higher order statistics. A recent review article on data-driven methods for process analysis is by Venkatasubramanian et al. [143]. Kourti and MacGregor gave an early tutorial and overview on multivariate statistics for process monitoring and diagnosis [66, 67]. A textbook on multivariate statistics for fault diagnosis is by Chiang et al. [15] based on the PhD thesis by Chiang [14]. Choudhury focuses in his PhD thesis on higher order statistics for fault diagnosis [19]. A further PhD thesis that has been consulted for

[2]Estimated by Andrew Ogden-Swift, IEEE APC Seminar 2005.

this work is on condition monitoring of control loops by Horch [46]. The topic is in this case particulary relevant because Horch focuses on the analysis of historical process data for evaluating the performance of a control loop using data-driven methods.

Fault and disturbances are at the centrepoint of the thesis and these terms are therefore frequently used. Broadly, fault is understood as a the malfunctioning of some process equipment. As a consequence, a disturbance is caused which affects a large number of process measurements causing the process to be less efficient as it would be without the presence of the fault. The time scale of the disturbance varies. Some disturbances are *long-term* and persist for weeks and months such as a heat exchanger failure or a sticking valve. A sensor failure, on the other hand, is an example of a *persistent* disturbance which lasts for hours or days. *Transient* disturbances are short deviations from the normal operation such as a compressor trip.

In this work, the root cause of a plant-wide disturbance is argued by the fault propagation path. The disturbance is caused at a specific point in the process and then travels through the process equipment. If the fault propagation path can be determined then the origin of the disturbance helps to identify the root cause. Model-based approaches that have been previously developed to identify the propagation path include signed digraphs. Signed digraphs have been constructed from expert knowledge [52] or from differential equations [80, 81]. The fault propagation path can also be modelled by a simplified version of the signed digraph, called causal map. Chiang and Braatz [16] derived a causal map from the process schematic and verified the fault propagation path using a data-driven entropy measure. The relevance to this work is that the results of the data-driven methods will be used to construct a causal map.

1.3 Contributions of Thesis

Methods and algorithms that form the main part of this thesis are concerned with identifying the fault propagation path of a plant-wide disturbance. They include the following:

- A simple method using cross-correlation to argue cause and effect between two process variables from the presence of dead time (Chapter 5);
- A causality method based on embedded vectors and nearest neighbours based on predictability of one process variable by a second variable. Five variations in the algorithm are proposed and compared. Guidelines for the parameters are established in Chapter 6;
- Application of the recently proposed method of transfer entropy to fault diagnosis in chemical processes. A significance level is introduced and parameter guidelines are established (Chapter 7);

- An algorithm for automatic generation of a causal map from causality measures in Section 4.4;
- Three new industrial case studies and the application of the causality measures to these case studies (Sections 4.2, 9.2 and 9.3);
- A decision tree for selecting the best of the three proposed causality measures depending on the data characteristics in Section 11.1.3.

All methods have been implemented with Mathworks Matlab Release 12[3]. A software license agreement has recently been signed between the author and University College London (UCL) which will allow the methods developed in this work to be incorporated in a plant-wide disturbances assessment tool developed by an ABB/UCL joint venture.

Publications

The work reported in this thesis has so far led to the following submissions and publications.

The transfer entropy method described in Chapter 7 has been published in

- Bauer M., Thornhill, N.F. and Meaburn, A., 2004. Specifying the directionality of fault propagation paths using transfer entropy. *Proceedings of DYCOPS 7*, Boston MA, 5-7 July 2004.

The concept of the nearest neighbours method in Chapter 6 has been published in

- Bauer, M., Cox, J.W. and Thornhill, N.F., 2005. Measuring cause and effect of process variables. *IEEE Advanced Process Control Applications for Industry Workshop*, Vancouver, Canada 9-11 May 2005 (Best Paper Award).

A comparison of the cross-correlation method in Chapter 5 and the transfer entropy method applied to the case study in Section 9.3 has been published in

- Bauer, M., Thornhill, N.F. and Meaburn, A., 2005. Cause and effect analysis of a chemical process analysis of a plant-wide disturbance. *IEE Seminar on Control Loop Assessment and Diagnosis*, London, UK, 16 June 2005.

The concept of transfer entropy using the case studies from Sections 4.2 and 9.2 has been submitted as

- Bauer, M., Cox, J.W., Caveness, M., Downs, J.J. and Thornhill, N.F., 2005. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Transactions on Control Systems Technology* (forthcoming).

[3]For more information on Matlab visit the website http://www.mathworks.com/products/matlab/, July 2005.

1.4 Organisation of Thesis

The thesis is divided into four parts and consists of twelve chapters. In the first part, the background literature is discussed. Chapter 2 gives an introduction to chemical processes, process control and the faults and disturbances that frequently occur in these processes. Chapter 3 explains the stages of process monitoring, that is, fault detection, identification and isolation, and gives a literature review of data-driven process monitoring methods. The data-driven features that are implemented in commercial monitoring systems are also described in Chapter 3.

In Part II, causality measures for fault diagnosis retracing the fault propagation path are developed. Chapter 4 outlines the framework of data-driven causality measures and some graphical representation tools. A reference case study is introduced together with an algorithm that automatically generates a causal map. Both case study and the automated causal map generation are used throughout the thesis. Also in Part II, three causality measures are derived. Chapter 5 describes an approach using the cross-correlation function, Chapter 6 introduces the concept of nearest neighbours and Chapter 7, the concept of transfer entropy.

The developed methods are applied to simulated and industrial data in Part III. Chapter 8 studies the impact of fault propagation effects on the causality measures from Part II. The fault propagation effects are dead time, low pass filtering and additive noise, as introduced in Section 2.4. The causality measures are applied to further industrial case studies in Chapter 9. Two case studies from processes at Eastman Chemical Company and BP sites are investigated and the root cause is identified in each. Chapter 10 studies the effect of quantisation and compression on the causality measures.

Part IV summarises the results of the application of the three causality measures to industrial and simulated data introduced in Part III. In Chapter 11, a framework is developed to find the best method for the analysed process characteristic. All steps required for the causality analysis are summarised in a causal analysis toolbox. Finally, Chapter 12 gives a critical discussion of the work, identifies open issues and highlights the main achievements.

Part I

Background

PART I - BACKGROUND

In this part, the terminology of control and monitoring approaches used for the industrial production of chemicals is defined. The focus of this work is the use of historical process data for monitoring techniques in continuous processes. In Chapter 2, control setup and instrumentation for chemical plants as well as faults and disturbances in these plants are described. A brief introduction to process control is given highlighting the use of PID controllers. A study at Honeywell [21] in 2001 estimated that 97% of all control setups are PID controller. The study also gave insight into common root causes: 50% of the disturbances observed were due to process or constraint problems, 40% due to controller tuning and 10% due to valve problems. Typical examples of data trends from these disturbances are also discussed in Chapter 2 as well as the mechanisms from which the propagation path of a disturbance can be identified.

Chapter 3 discusses previous and current work about process monitoring techniques. Two stages of monitoring are identified: fault detection and fault diagnosis. Fault diagnosis can be divided in two steps, fault identification and fault isolation. While identification is concerned with determining the kind and size of the fault, isolation is concerned with finding the root cause as well as the propagation to the current situation and the identification of the cause and effect relationships. This thesis focuses on these aspects of fault isolation. Thus, the emphasis of the literature survey on monitoring techniques is firstly on data-driven methods and secondly on methods that represent the propagation path, that is, qualitative models in the form of signed digraphs and causal maps. In the last section of Chapter 3, the state-of-the-art of commercial process monitoring systems is discussed focusing on the use of data-driven systems.

Chapter 2

Chemical Processes

This chapter gives an introduction to chemical processes and an overview of state-of-the-art control strategies in these processes. The control and fault terminology used in this thesis is defined. The cause of common faults and disturbances in the processes and their resulting data characteristics will be given.

Most chemicals are produced in continuous processes with a constant inflow and outflow that run continuously 24 hours a day. Seborg et al. [116] summarise the definition of chemical processes as used in the chemical industry:

> "A process is the conversion of feed materials to products using chemical and physical operations."

before adding: "In practice, the term process tends to be used for both the process operation and the processing equipment."

A number of constraints have to be regarded when producing chemicals on a large scale. Important constraints are safety, profitability, quality and environmental regulations. Most physical properties, also referred to as process variables, such as flow, level, temperature and pressure, should therefore be constant in order to achieve the desired chemical reaction. However, process dynamics, caused by basic principles of fluid mechanics, reaction kinetics, thermodynamics, heat transfer or mass transfer, often intervene and require intervening control action. On top of the wanted dynamics outside disturbances caused by faulty equipment, changes in process conditions or unwanted interaction can upset the process.

Eastman Chemical Company is a representative example of a large industrial production site for chemicals as it produces 1,200 chemicals, two basic fibres and three types of

Rank	Company	Rev. US$b (2003)	Country
1	Dow Chemical	32.6	US
2	BASF	30.8	Germany
3	DuPont	30.2	US
4	Bayer	21.6	Germany
5	Total	20.2	France
6	Exxon Mobil	20.2	US
7	BP	16.1	UK
8	Royal Dutch / Shell	15.2	UK / Netherlands
9	Mitshubishi Chemical	13.2	Japan
10	Degussa	12.9	Germany
⋮	⋮	⋮	⋮
34	Eastman Chemical Company	5.8	US

Table 2.1: Revenues of top 10 chemical companies in 2003.

plastics. The chemicals are produced as coatings, adhesives for tapes, labels and packaging, paints, sealants, printing ink, agriculture products, fibres, food, beverages, photographics, pharma polymers, medical devices, electrical connectors, personal care and cosmetics. More than 15,000 employees work for Eastman Chemical Company and half of them in Kingsport, Tennessee, USA. The chemical industry is large and continuously growing. Sales of chemicals in the US climbed from around US$ 200 billion in 1980 to US$ 390 billion in 1998. Sales in Europe were US$ 430 billion in 1998[1]. Table 2.1[2] lists the major chemical production companies and their revenues. The majority of the field work that forms the basis of the development of data-driven methods was undertaken at Eastman Chemical Company in Kingsport. During the placement at Eastman Chemical Company, the author of this thesis learnt control strategies in chemical plants, principles of chemical reactions, process representations and the nature of the task of process systems engineers.

The structure of this chapter is as follows. The first section reviews the differences in production types such as batch and continuous processes. In Section 2.2, graphical representation of chemical processes and process equipment are introduced which will be used in case studies in the reference case study in Section 4.2 and in Part III. Controlling the process maintains it at desired operating conditions. Control loops are introduced in Section 2.3 and their instrumentation as well as control strategies are reviewed. The major interest of this work is the analysis of faults. Faults and disturbances that can occur in chemical processes, their causes and data characteristics are given in Section 2.4.

[1] http://www.technology.gov/Reports/chemicals/chemical.pdf: US Department of Commerce, Office of Technology Policy. Meeting the challenge: US industry faces the 21st century; http://www.chemicalprocessing.com, February 2005.

[2] Chemical & Engineering News, Business, July 19 2004, Volume 82(29), p.12.

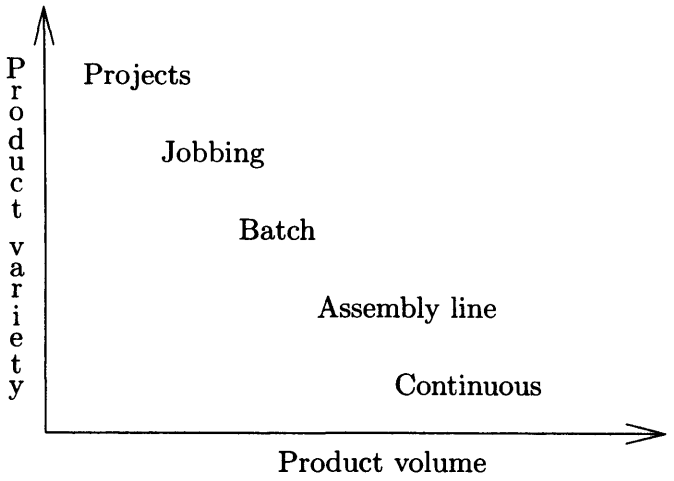


Figure 2.1: Process types as functions of product volume and product variety.

2.1 Process Types

This section discusses the types of processes in the chemical industry and in other industries. In the operations management literature, which provides managerial tools to design and assess processes on a high management level, five process types can be identified [122]: project, jobbing, batch, assembly line and continuous processes. Process types are chosen according to the quantity and variety of a product. Figure 2.1 illustrates the relationship between product volume, variety and process type by categorising processes into projects, jobbing, batch, assembly line and continuous. A general statement for all production types is that lower volumes imply higher variety. The process type therefore depends on volume and variety rather than on the nature of the product.

Projects: Products produced in projects are produced in low quantity and high variety. They are usually very complex and highly customised. The manufacturing process is a long cycle extending over several weeks or months. Projects always have a defined start and finish and some interval is taken between two projects. The resources for manufacturing - raw material, equipment and work force - are assigned to only one project. Examples of projects include ship building, construction sites and drilling oil wells.

Jobbing: Jobbing processes like projects also produce low volume and high variety products. The manufacturing process is highly flexible to product changes. The difference between jobbing and project processes is mainly that jobbing processes use shared resources for several jobs. The items processed are usually smaller and less complex than projects. Precision engineers and small requests in the printing industry are examples of jobbing processes.

Batch: Batch processes can deal with a wider range of product volume with a lower variety. Always more than one item is manufactured at a time, repeating the operation process. The process consists of a sequence of discrete steps. The application of batch

processes is widely spread in various industries: wafer handling in the semiconductor industry, component parts of automobiles, fermentation processes of pharmaceutical or frozen pizzas in the food industry.

Assembly lines: Assembly lines are employed for a high product volume and a relatively narrow variety. The processes are repetitive and largely predictable. The difference to batch processes is that only one discrete item at a time is manufactured. The most common example of an assembly line is car assembly where the basis is identical and variety is only introduced through different colour or engine type. Assembly lines can also be found in the food industry for instance in the form of beer bottling.

Continuous: Continuous processes operate at a highest possible volume and allow almost no product variety. The process runs over a long period of time and sometimes seems like an endless flow. Continuous processes are very inflexible and usually very capital intensive. Halting the production is always extremely costly and sometimes means permanent plant shut-down. Examples of products from continuous processes are petrochemicals and refineries, steel making and the paper industry.

The most common process types in the chemical industry are batch and continuous processes which will be explained further in the following sections with reference to the chemical industry. All processes that were analysed within the frame of this work were continuous processes. There is, however, no reason why the methods developed here should not be applied to data from batch processes with appropriate selection of variables and time parameters.

2.1.1 Continuous Process Monitoring and Control

Chemical products are most commonly produced in a continuous process since the production costs are low for large quantities. Production is run 24 hours a day and process shutdowns are extremely costly. Continuous processes are halted once or twice a year for general inspection of instrumentation and mechanical equipment. Common features are constant inflows and outflows. Seborg et al. [116] give a number of example of continuous processes:

- Tubular heat exchanger: A process fluid on the tube side is cooled by cooling water on the shell side;
- Continuous stirred tank reactor: Two or more reactants are fed into a vessel and continuously stirred;
- Thermal cracking furnace: Crude oil is broken down ("cracked") into a number of lighter petroleum fractions by the heat transferred from a burning fuel/air mixture;
- Multi-component distillation column: Two or more components are separated into a lighter and a heavier mixture.

When continuous processing methods were first used in industrial plants disturbance propagation was a serious issue that was initially reduced by placing surge vessels between process units [116]. Surge tanks are used to buffer changes in process variable such as flow or temperature. In modern plants the extra cost that surge tanks imply through set-up and operation are avoided but the result is a more tightly coupled process in which faults can propagate more easily.

Data measurements for continuous processes are recorded at fixed time points for all process measurements, resulting in so called *time trends*. Analysis can be carried out by applying methods to a single time trend (univariate) or by gathering a number of time trends (multivariate). A difficult part in the analysis is then the selection of the appropriate process variables. The non-invasiveness of any monitoring or fault detection method for continuous processes is important since no changes can be made to the process during operation. As a results, continuous plants in the chemical industry are only shut down once or twice a year.

When referring to chemical process control the term is usually equivalent to the control of continuous processes and numerous text books are available, see for example [76, 116, 118, 126].

2.1.2 Batch Process Monitoring and Control

Batch or semi-batch processing is the most common process type in the chemical industry after continuous processes. A sequence of steps is performed for a tank of defined quantity. Such a sequence can be for a simple blending process as described by the following steps [116]:

1. Transfer amount of material from tank A to tank R,
2. Transfer amount of material from tank B to tank R,
3. Agitate material in tank R for a period of time after the feeds are added,
4. Discharge the product to storage tank C.

Batch processes are often a compromise between the accuracy of a laboratory scale reaction and the volume of continuous processes. Some steps of the process may run in a continuous fashion in which disturbances similar to a continuous process can occur. Disturbances common in batch processes can be caused by fouling of vessel walls and heat transfer surfaces or raw material impurities.

Measurements are taken for each process step and each batch so that the data variations between batches rather than between time intervals are analysed. In this way, the data is available in a three dimensional way: batch number versus process variable versus process time [79]. For most batches, the process time for each step differs so that the length of

the data set for each batch differs as well. This must be taken into account when applying data driven methods. Furthermore, the data is in a three dimensional form which is impractical for most methods. The data can be transformed into a two dimensional set in by rearrangement in a number of different ways. Most commonly, the data is decomposed into a two dimensional array where each row represents a single batch run and comprises individual process measurements at each time point. The number of batches might also be limited, thus restricting the use of methods which requires a large amount of data.

An introduction to batch process control is given in the latest edition of the textbook by Seborg et al. [116] describing the control of batch reactors, the production management and run-to-run control incorporating modifications of the process steps from one run to the next; for control during batch operation, for example see Bonvin [8].

2.2 Graphical Process Representations

For most processes, several graphical representations of the process exist. Examples are the process and instrumentation diagrams (P&ID), the flow diagrams of the control software and process schematics as on overview tool. The P&ID diagram is a regulatory requirement and thus exists for all plants and units. However, for an overview and understanding of the process a process schematic is often produced by the control engineer. Process schematics are also used by the operators in the controller software. In some cases, these controller software flow diagrams are used by the process engineer for discussing the process flow and the controller interaction.

A common feature of all graphical representations is that the direction of intended flow is indicated by vectors and that the top and bottom of the sheet corresponds to top and bottom of the equipment. Unlike in electrical diagrams, in which resistors or capacitors can be placed at will, tanks, columns or pumps will never be shown upside down. Piping, however, might be not drawn to scale so that a tank and a pump on a process representation appear next to each other but are actually situated on other ends of different floors in the building.

Process and Instrumentation Diagrams: P&IDs are used for installation, process adjustments and trouble shooting by process engineers and operators alike. They are created by special computer added design (CAD) software and follow strict norms[3]. In the right hand lower corner, information about contractor and process owner companies as well as management, engineering and contractor approvals is kept. On the main section, tags and sizing of all equipment, exact position of sensors, outflows and inflows, connecting pipes, electrical and pneumatic signals are drawn. Also, physical information of location,

[3]See P&ID design handbook, http://www.engineeringtoolbox.com/p&id-piping-instrumentation-diagram-44_466.html, February 2005.

for example building number and floor level might be indicated. Space is left for notes and comments about changes in the equipments and connection of flows to other sheets are indicated. All information on the P&ID is additionally stored in related databases. P&IDs show process equipment and piping in great detail and for a single plant can easily exceed a hundred pages of A3 sheets.

Process Schematic: The process schematic is a formless graphical representation of all major process flows and equipment drawn by hand or with standard office software such as Microsoft Word or Excel. There are variations in the representation of equipment. There is an unspoken convention that the process schematic should not exceed a single page (letter, A4 or A3 formats) for convenience in handling, especially when going out to the plant. All control loops and major equipment should be included although often only a part of the complete process is represented. Commonly used representations of the important equipment is shown in Figure 2.2. Heat exchangers heat or cool streams of gas or liquid and can induce a phase change from liquid to gas by vaporisation or gas to liquid by condensing [126]. Pumps are used to transport the product from one piece of equipment to the next. They ensure a certain pressure and flow and can be also used to control levels. In a distillation column, two or more components are separated from each other based on their different vaporisation points. A feed enters the column and the lighter distillate exits the column at the top while the heavier product exits at the bottom. Trays or solid packaging, which are in some rare cases indicated in the process schematic, are inserted in the column to ensure a large contact surface for the reaction to happen. Separation in distillation columns is the chemical operation that is most often used in the chemical industry[4]. The most common control element in the chemical industry is the control valve which is explained in more detail in the next section. The measurement points are abbreviated depending on the type of measurement, that is, F for flow, L for level and so on. Furthermore, a differentiation is made between a controlled measurement (C) and an instrument (I). The two letters are then combined. For example, TC stands for a controlled temperature.

2.3 Process Control Loop

To achieve desired operating conditions control mechanisms are added to the process in the form of control loops. A distinguishing feature of feedback control is that the controlled variable is measured and the measurement is then used to adjust the manipulated variable. The block diagram structure of a feedback control loop is shown in Figure 2.3. The process dynamics act on the process variables (PV) which are measured by a sensor. Process variables are physical quantities such as flow, level, temperature or pressure. The measured values are compared against a given setpoint (SP) and the resulting deviation

[4]See http://chemistry.about.com/od/chemicalengineerin1/, March 2005.

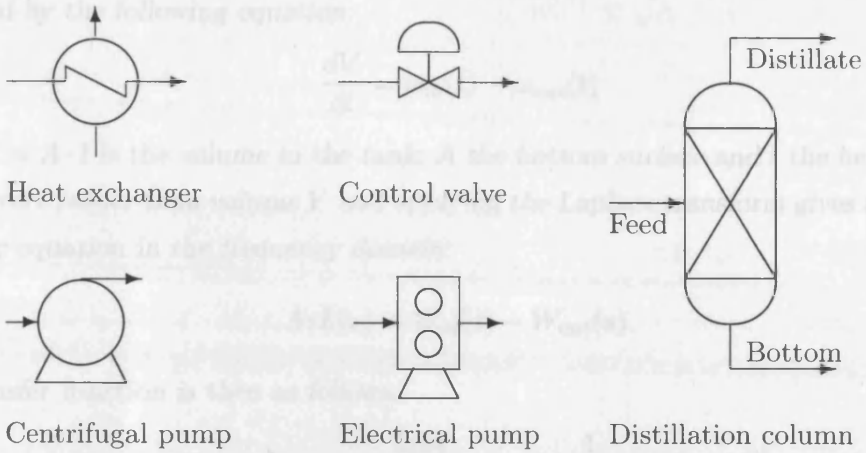


Figure 2.2: Equipment frequently represented in process schematics

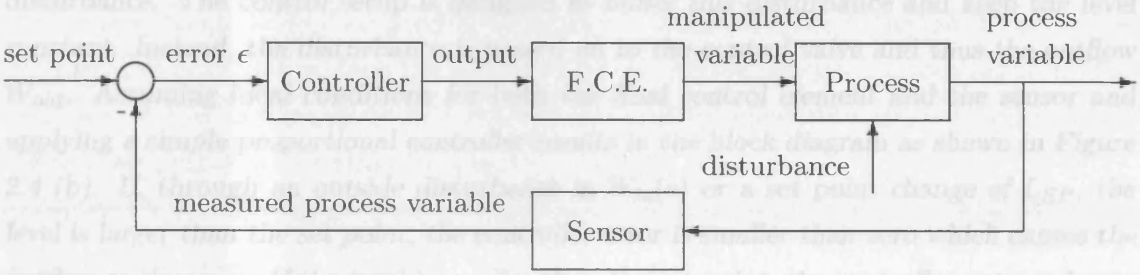


Figure 2.3: Block diagram of feedback control loop including final control element (F.C.E) and measurement facility (sensor).

from the setpoint is an input signal to the controller. The controller output (OP) signal acts on a final control element, in most cases a control valve, which then in turn adjusts the input variables of the process through the manipulated variable (MV). Disturbances enter the process at any point but are not measured in feedback control. The corrective action occurs via the feedback regardless of the nature of the disturbance. The purpose of any controller is not to eliminate variability but to move variability to a variable that is associated with lower cost. For feedback control this means transferring variability from the process variable to the manipulated variable, which is in most cases a flow rate.

For the analysis carried out in this work, SP, PV and OP are available and considered although in most cases only the relationship between process variables is investigated. The manipulated variable is only available if a so-called smart device is used measuring the MV.

Example: *A simple example of a tank given for instance by [126] with a controlled level is chosen to explain the derivation of process description and control scheme. Figure 2.4 (a) shows a tank with inflow $w_{in}(t)$ and outflow $w_{out}(t)$. The control objective is to keep the level in the tank constant despite any incoming disturbances through $w_{in}(t)$. To describe the process, the relationship between flow and volume in the tank can be*

expressed by the following equation:

$$\frac{dV}{dt} = w_{in}(t) - w_{out}(t) \tag{2.1}$$

where $V = A \cdot l$ is the volume in the tank; A the bottom surface and l the height or level. Using level l rather than volume V and applying the Laplace transform gives the following algebraic equation in the frequency domain:

$$AsL(s) = W_{\text{in}}(s) - W_{\text{out}}(s). \tag{2.2}$$

The transfer function is then as follows.

$$\frac{L(s)}{W_{in}(s) - W_{out}(s)} = \frac{1}{As} \tag{2.3}$$

The manipulated variable is the outflow $W_{out}(s)$ while $W_{in}(s)$ is considered as an outside disturbance. The control setup is designed to buffer this disturbance and keep the level constant. Instead, the disturbance is passed on to the control valve and thus the outflow W_{out}. Assuming ideal conditions for both the final control element and the sensor and applying a simple proportional controller results in the block diagram as shown in Figure 2.4 (b). If, through an outside disturbance in $W_{in}(s)$ or a set point change of L_{SP}, the level is larger than the set point, the controller error is smaller than zero which causes the outflow to decrease. If the level is smaller than the set point, the controller error is larger than zero so that the outflow increases. The resulting system is first order and therefore does not cause any oscillations. The time constant of the dynamic system is $\tau_D = \frac{A}{K_c}$. A small dynamic time constant τ_D ensuring fast tracking of process changes is achieved through a large controller gain K_c. The disturbance W_{in} however, is then passed on to a downstream unit through the outflow W_{out} while the level stays constant.

The type of a control loop refers to the process variable which is either a flow, level, temperature or pressure. Consistency controllers are also employed in some plants to regulate composition. Table 2.3 gives the percentage of flow, level, pressure and temperature loops in industrial plants. In a recent study by Paulonis and Cox [92] a large-scale controller performance system assessed 14,000 PID controllers in 40 plants at Eastman Chemical

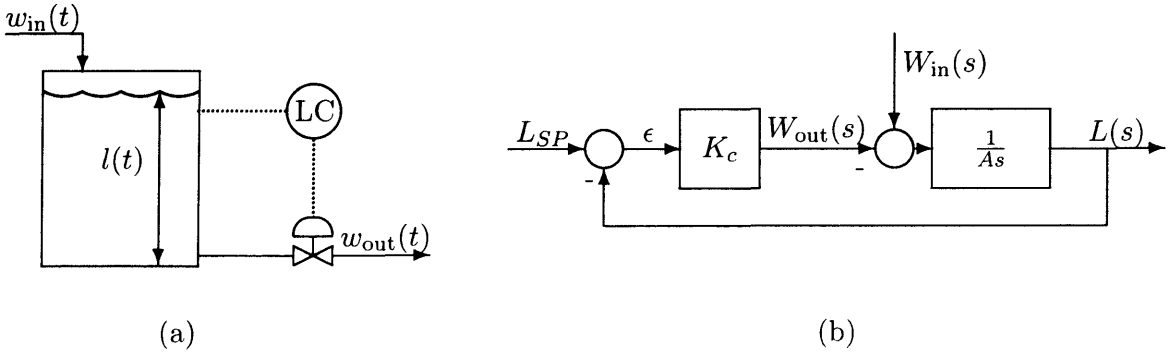


Figure 2.4: Process schematic and block diagram of level feedback control loop.

Control loop type	% as in [92]	% as in [21]
Flow	32%	39%
Level	18%	20%
Pressure	15%	20%
Temperature	20%	19%
Other	15%	2%

Table 2.2: Percentage of controller types averaged over a large number of industrial processes by Paulonis and Cox [92] and Desborough and Miller [21].

Company while Desborough and Miller [21] collected data from 11,600 controllers at eighteen sites in multiple process industries.

The role of the controller will be explained in the following sections. In general, controller performance is considered as good if variability is transferred from the controlled to the manipulated variable and thus the variability of the critical process variable reduced. The last part of this section gives an overview of the instrumentation used for control and monitoring purposes in a chemical plant.

2.3.1 PID control

The control algorithm for the feedback controller as shown in Figure 2.3 has three basic control modes: proportional (P), integral (I) and derivative (D) control. These modes are all combined in a PID controller that is implemented in most distributed control systems (DCS) in chemical plants. A proportional controller was used in the previous section for the example of a level controlled tank. The proportional control mode can be used to respond rapidly to large errors caused by disturbances. The controller gain K_c is adjusted to make the loop sensitive to deviations of the setpoint. Integral action can be added to remove any offsets or steady-state errors. This is achieved because the controller output depends on the integral of the error signal over time. The integral time constant τ_I controls the rate of integration. A small integral time removes the integrated error fast but also leads to decaying oscillations. A derivative control action counteracts integral and destabilising lags in the process or controller with a derivative time constant τ_D. The controller anticipates the future behaviour of the error signal by considering the rate of change. Derivative control is never used on its own but always in conjunction with proportional and integral parts. The transfer function of a PID controller is:

$$G_c(s) = K_c \left(1 + \frac{1}{\tau_I s} + \tau_D s\right). \tag{2.4}$$

For a detailed treatment of PID controllers see Åström and Hägglund, [159].

Tuning rules were established by Ziegler-Nichols tuning or Tyreus-Luyben tuning which are based on loop testing and are therefore invasive [116]. Luyben and Luyben [76] give

	Flow	Level (surge)	Level (other)	Pressure	Temperature
Type	PI	P	PI	PI	PID
K_c	small	scaled to level	scaled to level	depending on application	small
τ_I	fast (0.1min)	-	slow	depending on application	process time constant
τ_D	-	-	-	-	1/4 process time constant

Table 2.3: Tuning guidelines as rules of thumb by Luybens and Luybens [76].

rules of thumb for the tuning parameters of flow, level, pressure and temperature loops. These rules are extremely valuable when quickly checking the tuning parameters of an inspected loop. An overview of the guidelines is given in Table 2.3. For flow controllers, PI control is used with a low gain to limit noise due to flow turbulence. The value for the integral time is small for fast responses (around 0.1 minutes). An exception is the control of throttled reboilers, for example at the bottom of a distillation column, for which a slower response time is chosen. For level loops, the purpose of the control action has to be considered. Level controllers that provide for surge capacity need to react slowly and need not hold the level tightly at its setpoint. If the tank has a different purpose, the accurate level might be desirable and PI control can be applied. Temperature control loops are often slow and PID controllers can be used.

The use of PID controllers is widespread. Desborough and Miller [21] estimate that 97% of all control loops have PID controllers and give three reasons for its popularity: (i) the PID works well in the vast majority of applications; (ii) it is easy to understand and (iii) the PID algorithm is pre-programmed in every control system. Nevertheless, a large proportion of the control loops in chemical plants are open loop [21, 26, 92] also referred to as in 'manual'. This means that the operator has intervened to remove any automatic control action because the controller was not fulfilling its task.

2.3.2 Control Loop Instrumentation

The instrumentation of a control loop is shown in Figure 2.5 as it would be represented in a detailed P&ID. First, the sensing element, for example an orifice for flow or a thermocouple for temperature measurements, measures the process variable. The result is converted by the measurement transmitter into a signal readable by the control element. The control element then performs the control algorithm and sends the result to the signal transducer which converts it in a format readable by the final control element. The final control element is in most cases a control valve but can also be a pump or a heating element.

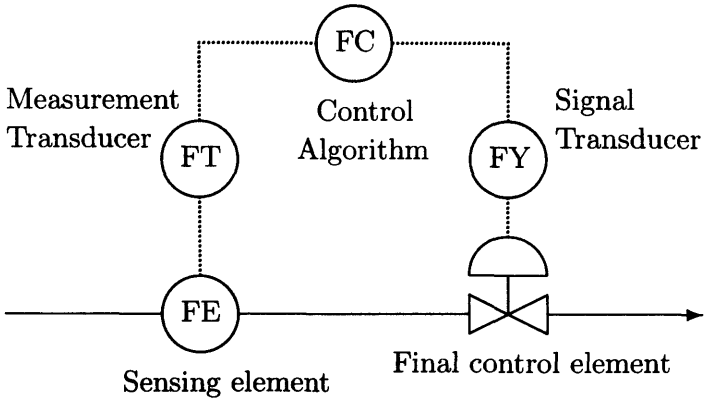


Figure 2.5: Instrumentation of flow control loop in P&ID representation.

The use of DCS in modern control instrumentation is widespread. The controller algorithm is in this case discrete instead of continuous. The controller therefore includes an analogue-to-digital converter (ADC) to process the signal received by the transmitter as well as a digital-to-analogue converter (DAC). Some electronic final control elements, such as an electronic pump, can process a digital signal and thus comprise the DAC. The sampling rate of the ADC, which results in term "DCS data" coined by process engineers and operators, is usually set to 0.1 seconds. Any time constants faster than 0.1 seconds can therefore not be implemented nor recorded and used for the fault diagnosis analysis. The data is logged and sent to the data storage system every five seconds.

Sensing Element: The main sensors used in chemical processes are temperature, pressure, flow rate and liquid level. For a detailed description of sensors see Lenk [71] or Lipták [75]. Seborg et al. [116] give a list of on-line measurement options most commonly used in process control. Knowing the type of sensor that is employed in a control loop can give insight into the process behaviour. Temperature sensors include filled systems, which are familiar from daily usage, and thermocouples. Filled systems use a liquid that expands with increased temperature. Thermocouples consist of a pair of wires of different metal, one maintained at a reference temperature, the other at the temperature to be measured, and both joined at one end. A temperature difference produces a voltage between the two metals also known as the thermoelectric effect. The time constants of a temperature sensor can be in the same range as the process time constants and thus can affect the control performance [76]. With electronic instrumentation, a strain gauge is often used to convert pressure into an elongation of resistance wires which changes a millivolt-level [116]. Pressure sensors are very unreliable and are often set in pairs when measuring a crucial pressure. Flow cannot be measured on its own easily but other effects are used to deduce the flow. One basic principle of sensing the rate of flow is to place an obstruction in the path of the fluid and measure the difference in pressure before and after the obstruction using a differential pressure sensor. The obstruction object is often an orifice plate that is a disk with a hole. Low flow rates can be measured through the heat loss from a heated element which varies with flow rate. Signals from flow measurements are usually noisy

Input	Output pneumatic	Output electrical
PV_{min}	3 psig	4 mA
PV_{max}	15 psig	20 mA

Table 2.4: Standardised ranges for pneumatic and electronic transmitters [116].

and often need to be filtered before sent to the controller. Liquid levels can be easily measured by following the position of a float that is lighter than the fluid, measuring the apparent weight of a heavy cylinder as it is buoyed more or less by the liquid or measuring the difference in static pressure between two fixed elevations, one in the vapour above the liquid and the other under the liquid surface [76].

Transmitters and Transducers: Transducers are the interfaces between the process and the control system and, in this role, measure process variables and convert them into a signal that can be interpreted by the controller. As a terminology convention, the combined setup of sensor and transmitter is referred to as transducer [116]. The output signal range of the transmitter is restricted to standardised ranges as listed in Table 2.4. These ranges are used in the majority of all process industries. The difference between the minimum and maximum value of the process variable, PV_{min} and PV_{max}, is called "span". The dynamic response of the transmitter is usually much faster than the process and the control valves. Thus, the transmitter can be considered as a simple gain [76].

Final Control Element: In chemical processes, the control valve is the most common final control element and in an average industrial process several thousand valves can be found [21]. This is because the variability of temperature, level and pressure can be transferred to the flow rate. Lenk [71] distinguishes between two types of control valves: shut-off type and throttling type. The opening of the shut-off valve is either completely open or closed acting as a switch or, in electrical terms, an ideal diode. The throttling valve, on the other hand, is adjustable to any opening between these two extremes acting as a tuner or transistor.

The most common type of throttling valve is the plug-and-seat valve in the form of a globe valve. The flow through the valve is controlled through a movable plug that is placed in the seat if closed and further away if open. Thus, the relative position between the plug and the seat determines the effective flow through the valve. A stem is attached to the plug and controls the opening or closing movement by means of a diaphragm that is moved by air pressure. The stem itself is controlled by an electrical or manual actuator. Other frequently used types of valve are butterfly, ball and angle valves.

The throttling valve has an inherent flow characteristic that describes the relationship between stem position and the maximum of flow through the valve [116]. The characteristic is grouped in two behaviours, linear trim and equal-percentage trim as shown in Figure 2.6. The term equal-percentage comes from the slope of the characteristic being a constant

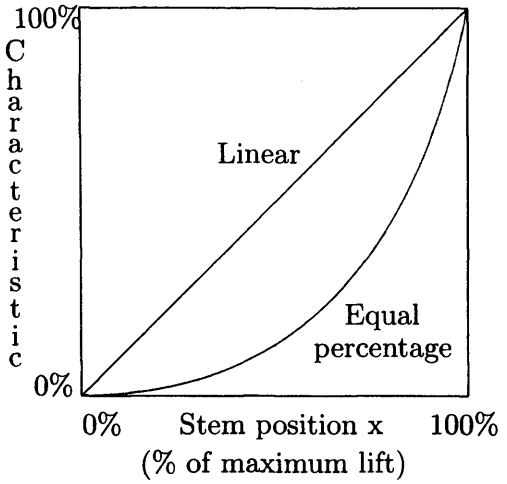


Figure 2.6: Valve characteristic as a function of stem position as in [76].

fraction of the flow [76]. Equal-percentage trim valves are more favourable if the valve is placed between process equipments. This is because the installed characteristic is a function of both the inherent characteristic and neighbouring pieces of equipment that act as resistances. As a result, installed pneumatic valves often have inherent second order dynamics contributing a gain and two time lags to the control loop [126]. The size of the valve chosen by the designer depends on the flow rate, specific gravity of the fluid and pressure drop over the valve.

To monitor the performance of the valve, a valve positioner can be attached to the valve. The positioner measures the actual stem position and compares the position to the value requested by the controller output. Additionally, an independent air source is used to drive the measured stem position.

2.4 Faults and Disturbances

Monitoring the performance of a process involves identifying and analysing unexpected events occurring within the process. These unexpected and mostly unwanted occurrences are called *faults*. An often used definition of a fault is given in the report for standard terminology in technical process monitoring by Isermann and Ballé [53]:

> "A fault is an unpermitted deviation of at least one characteristic property or variable of the system."

Patton et al. [90] define fault as follows:

> "A fault is an undesired change in the system that tends to degrade overall system performance, although it may not represent the failure of physical components."

The term fault is preferred to the term *failure* because a failure usually involves the breakdown of a complete system or plant. More precisely, a failure is a permanent interruption of a system's ability to perform a required function under specified operating conditions [53].

Measures of faults are: how often a fault occurs, the probability of faults occurring and the amount of available useful operating time [122]. The distinction between fault and *disturbance*, however, involves much debate. Gertler [34] notes that the difference between fault and disturbance is that faults are exceptional whereas disturbances are present most of the time. It is a matter of judgement as to what is exceptional and what is most of the time. Isermann and Ballé [53] define disturbance as an unknown (and uncontrolled) input acting on a system. The underlying cause of the exceptional event of a disturbance is referred to as the basic event or, more commonly, the root cause.

A number of industrial studies [7, 21, 26] investigate common root causes of the disturbance which will be reviewed in the following Section 2.4.1. Resulting disturbances often have common features in their data trend which will be looked at thereafter. The fault propagation mechanism that causes the disturbance to travel from one process variable to the next is the motivation of applying causality measures and will be reviewed in the last section.

2.4.1 Common Root Causes

The industrial study by Desborough and Miller [21] classifies the problems associated with control loops in three categories: valve problems, controller problems, process or constraint problems. From the 11,600 loops of a number of industries analysed they found the following split of problematic loops:

Process or constraint problems	50%
Controller tuning problems	40%
Valve problems	10%

A previous study limited to the pulp and paper industry [7] found 30% of the loops assessed having valve problems while a third of the loops cycled due to inappropriate loop tuning. Similar results were reported by Ender [26] (30% equipment problems, 15% tuning problems of all loops analysed). In the following paragraphs these main categories of root causes are discussed. A number of reasons for oscillations can be found in Horch [46] which include static friction, dead-zone, backlash, saturation and quantisation.

Process or Constraint Problems: Problems caused by the process configuration and constraints are as manifold as the equipment and design variations in the process.

Shunta [118] gives the following list of sources of instabilities that cause disturbances in the process:

- Reaction systems are typically nonlinear and often the cause of product variability;
- Separation systems (distillation, absorption, extraction, scrubbing, etc.) have several control loops that may interact, causing cycling. Loops are often cascaded which may amplify the effect of upsets;
- Heat exchangers are a common source of variability caused by physical constraints;
- Piping and recycles tie parts of the process together and cause instability.

Control loop performance measures such as the Harris index which will be described later on in Section 3.1.3 often fail when assessing process problems. The reasons for this are that not only a large number of process variables is affected but also the variable close to the root cause does not necessarily have the worst performance index. Unlike tuning and valve problems, process problems come in a large variety caused by various parts and configurations of the process. This makes it difficult to cluster or structure the disturbances into categories. Any knowledge of the fault propagation path is therefore extremely useful when investigating the root cause of the problem. An example of a process constraint problem is discussed in Section 9.3 using a case study from a plant at a BP Chemicals site.

Controller Tuning Problems: The vast majority of control loops are PID controllers, see Section 2.3.1, which can level out disturbances by introducing and adjusting controller gain and time constants for the integral and proportional part. The consequence of PID feedback control is that the controller can actually cause oscillatory responses. For good adjustments of the tuning parameters, an induced oscillation has a small amplitude and damps out quickly. However, for some settings of the control parameters the amplitude does not decrease with time and causes persistent cyclic oscillations. The adjustment of the control parameters can be made on site on the distributed control system. Luyben and Luyben [76] estimate that 80% of all loops are tuned experimentally by an operator and 75% of the time the operator can guess appropriate parameters by drawing parallels to past experience. However, in many cases the tuning is either not appropriate or changes are made to the process configurations, such as throughput or consistency adjustments. Thus, a large percentage of disturbances are due to tuning problems.

Valve Problems: Control valves are commodity products in the chemical process, they are numerous and often a cheaper model is chosen over a more reliable model to save on installation cost of the process. Until recently, most control valves installed had no positioner so that the performance of the valve was difficult to monitor[5]. Problems of the valves include oversizing and undersizing, excessive hysteresis, static friction and stick-slip

[5]For a description of control valves and positioner see Section 2.3.2.

behaviour [26]. Physical causes of the valve faults are listed by Patton [91] and include clogging, sedimentation, errosion, increase of brushing friction, leakage and evaporation. Stiction is caused as the valve movement does not follow the controller output, which is also the valve input, in a linear fashion. Instead, the valve does not move at all until a threshold is exceeded. If the pressure on the valve is high enough the valve moves too far and the controller must reverse the direction. The same effect will occur in the opposite direction when the valve should close. Small variations are enough to trigger this behaviour and the loop will fluctuate around the setpoint without any outside disturbance. For a detailed description of valve stiction see [18] by Choudhury et al. who in a recent paper model also derive a model for valve stiction [17]. The consequence of these problems is often a cyclic behaviour seen in the data trend of the following flow variable. An example of a valve with stiction is shown in the next section.

2.4.2 Data Characteristics of Disturbances

Most of the process variables in a continuous process are ideally constant. This applies equally to a level in a tank as to the temperature at a certain tray in a distillation column or a flow from one piece of equipment to the next. Most disturbances make the process variables deviate from the desired set point and thus affect the efficiency of the process or the quality of the product. The nature of the disturbance reflected in the data trend may have various appearances. Often, a cyclic, or periodic, behaviour can be observed but also irregular disturbances are common. In this section, sample data trends from industrial processes are displayed for different root causes to give a first notion of the time trend from which information will be extracted. Figure 2.7 shows time trends for some typical process disturbances. However, these trends are only exemplary and the disturbance can have very different features.

Process Problem: Plot (a) in Figure 2.7 is the data trend of a process disturbance originating from the upsets in the inert gas flow. The case study of this disturbance is investigated in Section 9.2. The prominent feature in the data trend are spikes occurring at irregular intervals. These spikes consist of a short period at an increased level which is followed by a sharp decrease before returning to the original value. During the interval from minute 800 to minute 1000 the spikes occur at an almost regular frequency.

Constraint Problem: A process variable that is constraint by a physical limit is shown in plot (b) of Figure 2.7. In this case, the control valve was always too close to the state of complete shutting off. Thus, the time trend of the flow seems to stick to a lower boundary. The high frequency of oscillation, a period of only a few seconds, indicates that the cycling is due to a mechanical equipment problem.

Tuning Problem: Oscillations due to tuning issues are very common and an example of a loop with poor tuning is shown in Figure 2.7 (c). The cycling is very regular and

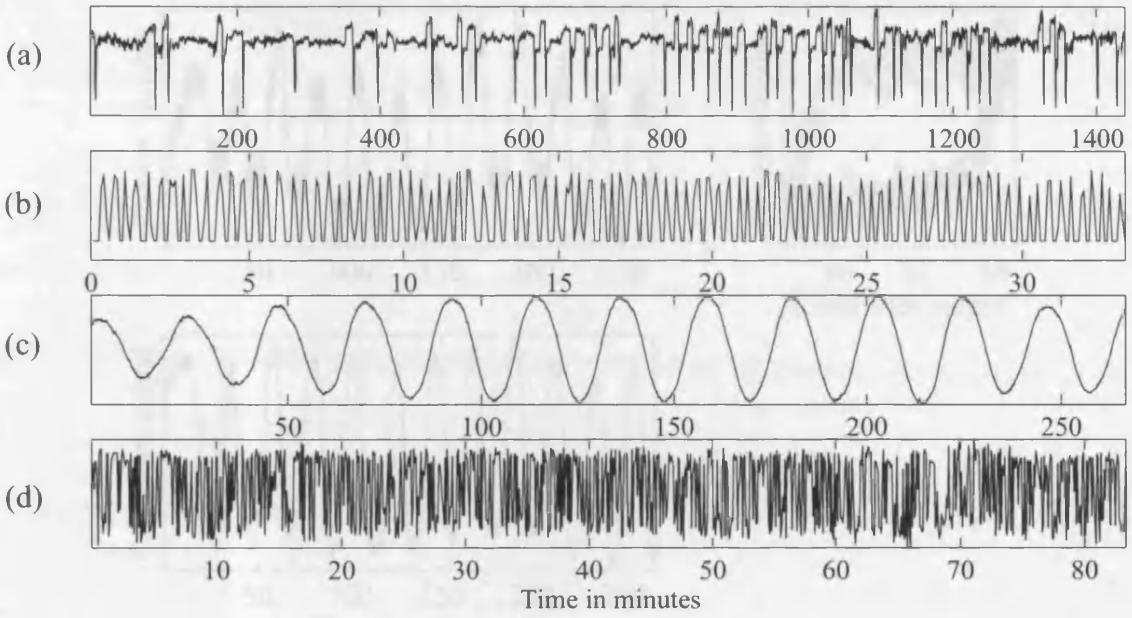


Figure 2.7: Time trends of disturbances by different root causes; (a) Irregular process disturbance via inert gas inflow, (b) Constraint problem (minimum value) resulting in high frequency oscillation, (c) Oscillation due to tuning problem, (d) random process noise.

almost completely sinusoidal. The magnitude of the oscillations changes with time but is very high at all times compared to the noise level. The oscillation period, in this case 20 minutes, can give insight to the tuning parameters.

Process and Measurement Noise: A common feature of any process is a certain amount of noise. Most of the time, outside process disturbances in the form of measurement or other noise is not significant enough to affect process performance. The noise depicted in Figure 2.7 (d) is uniformly distributed and has constant mean and variance. It was caused by a faulty sensor which upset the measurement.

Valve problem: A common root cause is stiction in control valves. A typical data trend of the process variable (PV) and output (OP) of a controller with valve stiction is shown in Figure 2.8. The PV value is sticking to a value for some while until it drops sharply. The controller output cycles continuously with the same oscillation period of approximately 24 minutes. The PV/OP plot clearly shows a hysteresis instead of a linear function. Since stiction, and in some cases slip stiction, is a common problem that affects directly the efficiency of a process, a large number of data-driven methods are available to detect and classify these control loops, see for example [18, 46, 102]. These examples use the PV/OP relationship of the loop with valve stiction either in the form of cross-correlation [46] or by extracting features of the PV/OP plots [18, 102]. The root cause analysis for control valves is actually not perfect because behaviour of PV and OP can in some cases actually be misleading. Looking at Figure 2.3 shows that the characteristic of the final control element, the valve, can be derived if the controller output and the manipulated variable (MV) are available. The availability of a measurement of the MV is rather the exception.

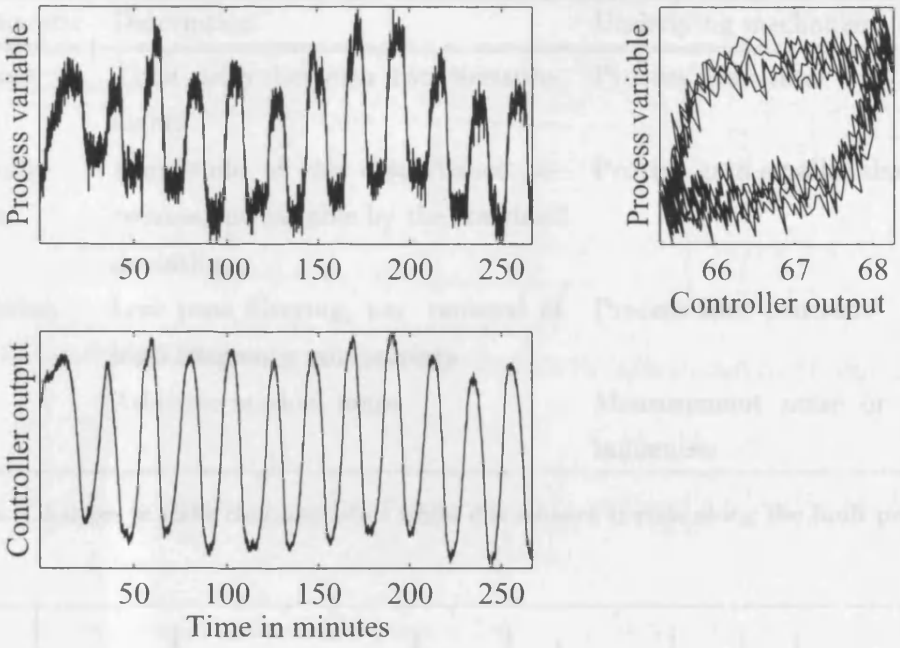


Figure 2.8: Fault propagation mechanism of magnitude decrease: Time trends of process variable and controller output for a loop with valve stiction.

This is the reason why the process variable is used instead. Replacing the MV by the PV only works well if the process has no relevant dynamics and no outside disturbances or sensor faults occur. Thus, it basically works for flow loops only.

2.4.3 Fault Propagation

In most cases, faults travel in the direction of the gas or fluid flow just as normal variations that occur in the process. However, the direction of propagation might either not be measurable under normal process conditions or can invert in some cases under abnormal situations. The knowledge of the propagation path improves root cause analysis significantly since the question of cause and effect can be answered and the disturbance traced back to its origins. In particular, challenges are posed through recycle streams in a process since a disturbance might affect all process variables in the recycle so that it becomes impossible to understand which variable started the disturbance.

Four common effects can be observed which change the features of the disturbance and allow conclusions of the direction of propagation. An overview of the changes to the signals due to fault propagation is given in Table 2.5. These observations are based upon experience from dealing with a number of process data sets and give a tendency of the behaviour of the disturbances as they travel from within the plant. The effects will be taken up later and their influence on the methods developed in this work will be investigated in Section 11.1.1.

Characteristic	Description	Underlying mechanism
Dead time	Time delay between two measurements	Process dead time
Magnitude decrease	Amplitude of the disturbance decreases, measurable by the standard deviation	Process gain smaller than one
Attenuation	Low pass filtering, i.e. removal of high frequency components	Process time constant
Noise	Additive random noise	Measurement noise or outside influences

Table 2.5: Changes in data characteristics while disturbance travels along the fault propagation path.

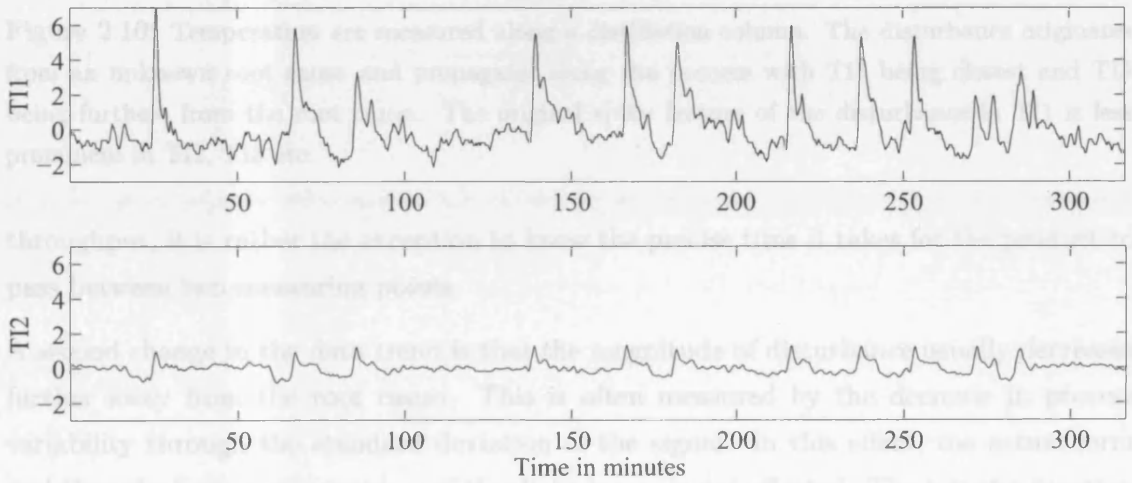


Figure 2.9: Fault propagation mechanism low pass filtering: Two temperature measurements with disturbance and different magnitude along a distillation column, TI1 at the top, TI2 at the bottom. The disturbance entered the column through the top inflow.

Firstly, the disturbance can be observed with the time delay that is equal to the dead time of the process between two observation points. An example for dead time is a composition or temperature change in a fluid pumped from a mixer to a stirred tank. The dead time injected by the connecting pipe will depend on the length of the pipe and the flow rate and can be large. Another example (see Shuta, 1993) is the exit temperature in a heat exchanger. Again, it will take an amount of time for the heated gas or liquid to reach the desired temperature. Dead time can also be inserted by measurement. A viscometer, for example, that measures viscosity draws a sample of fluid from the process through a tube. The dead time corresponds to the time it takes for a molecule to make its way down the tubing. A long length of tubing can add a significant amount of dead time. The knowledge of the dead time inherent in the process would simplify the search for the propagation path enormously. Because the dead time changes with changing process conditions, such the

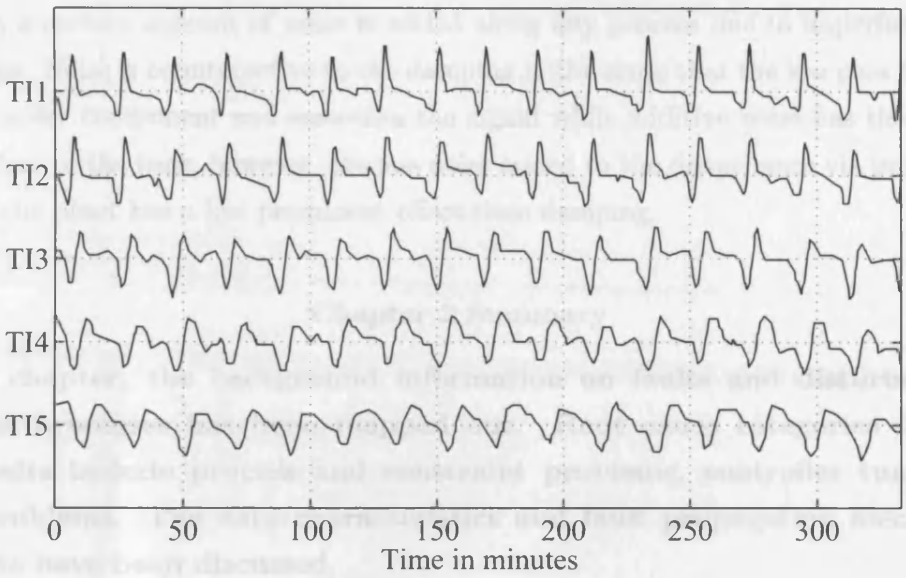


Figure 2.10: Temperature are measured along a distillation column. The disturbance originates from an unknown root cause and propagates along the process with TI1 being closest and TI5 being furthest from the root cause. The original spiky feature of the disturbance in TI1 is less prominent in TI2, TI3 etc.

throughput, it is rather the exception to know the precise time it takes for the product to pass between two measuring points.

A second change in the data trend is that the magnitude of disturbance usually decreases further away from the root cause. This is often measured by the decrease in process variability through the standard deviation of the signal. In this effect, the actual form and thus the frequency spectrum of the disturbance is not affected. The transfer function of the process, if known, would have a gain smaller than one which leads to the decrease. A number of statistical tools, such as principal component analysis (see Section 3.3.1) are based on this assumption of change in the data trend. Figure 2.9 shows an irregular disturbance in two temperature measurements at the top and bottom of a distillation column. The disturbance enters the column at the top. The standard deviation in the upper plot is 1.4 ^{0}C while the standard deviation of the signal in the lower plot is 0.3 ^{0}C.

Thirdly, the data characteristic often also appears attenuated or damped when distributing along the process flow. This is due to the time constant of most process transfer functions. Usually, the period of oscillation or the occurrence of abnormal peaks is shorter than the time constant of the system. Thus, the process acts as a low pass filter on the disturbance. The filtering of the high frequencies has the effect that potential harmonics inherent in a disturbance are suppressed while the main oscillation sustains. This feature is exploited in methods described in Section 6.2.3. Figure 2.10 shows an example of a nonlinear harmonic disturbance which appears smoother and more like a sinusoid the further it travels in the process. The decrease in magnitude and the damping are both aspects of the dynamic behaviour of the process.

Fourthly, a certain amount of noise is added along any process due to imperfect outside conditions. Noise is counteractive to the damping in the sense that the low pass behaviour removes noisy component and smoothes the signal while additive noise has the opposite effect. Most of the time, however, process noise added to the disturbance via propagation through the plant has a less prominent effect than damping.

Chapter 2 Summary

In this chapter, the background information on faults and disturbances in chemical processes has been mapped out. Root cause categories for common faults include process and constraint problems, controller tuning and valve problems. The data characteristics and fault propagation mechanisms for faults have been discussed.

Chapter 3

Process Monitoring

This chapter discusses process monitoring techniques for the chemical industry. The requirements for these techniques are defined and the respective literature is reviewed. The focus of this work is on data-driven methods for fault propagation and inferred qualitative models. The present status of industrial process monitoring systems is discussed.

Monitoring the performance and efficiency of critical process parameters is an essential component of process control. The reason for this is that processes conditions, such as throughput, outside temperature or consistency of the feed, change during operation. Also, equipment can show wear and tear and add unwanted characteristics to the process behaviour. The monitoring can be conducted on a localised level for each control loop and process variable or on a global level over a number of loops and variables. The area of control loop performance assessment (CLPA) or control loop performance monitoring (CLPM) has been studied in detail, for an overview see [98]. Most CLPA techniques also exploit the time trend of controller variable, output and set point but are supplementary to process monitoring techniques. The process monitoring methods that are investigated in this chapter usually do not consider the controller type or setup and investigate the historical data trends of process variables instead. Established statistical methods for process monitoring are summarised in [67]. The two approaches of control loop performance and process monitoring work hand in hand to achieve the objectives of process efficiency, product quality and safety.

Similar process monitoring methods are developed for and applied to a number of industries in which continuous processes can be commonly found. These industries include the production of chemicals, petrochemicals, pulp and paper, food and pharmaceutical products. The more critical production conditions are, the greater is the need for sophisticated control strategies and process monitoring techniques. In the pulp and paper industry the composition materials and the speed of the paper rolls are critical [7] and require careful control and observation. Pinder and Godfrey [95] describe the requirements for process

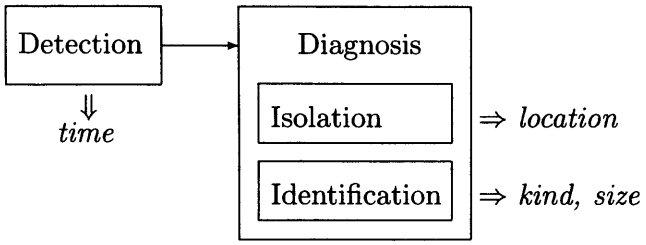


Figure 3.1: Stages of process monitoring system: fault detection and diagnosis give the time, location, kind and size of the disturbance.

monitoring techniques in the food industry while Luyben and Luyben [76] focus on the chemical and petrochemical industry.

This chapter is organised as follows. In the first section, an overview of the requirements and a list of the problems addressed by the monitoring methods are given. Benchmark criteria for the comparison of various techniques are discussed. Since process monitoring is usually compared to control performance monitoring, a short introduction to control performance monitoring techniques is included. In Section 3.2, classifications of monitoring techniques from the literature are compared and a classification suited for this work is introduced. The main groups are model-based and data-driven techniques. Since this work focuses on data-driven techniques, common techniques from the literature are described in Section 3.3 and 3.4.

3.1 Requirements for Process Monitoring

A process monitoring system is structured in several stages. These stages are fault detection, isolation, identification and diagnosis which are defined in the following. A number of benchmark criteria have been defined by Venkatasubramanian et al. [141] which are given and explained in Section 3.1.2. Another way of benchmarking is the comparison relative to control loop performance monitoring systems. A short overview of present monitoring indices is therefore given in Section 3.1.3. Control loop and process monitoring systems can also be combined to enhance the monitoring tool.

3.1.1 Stages of Monitoring

Process monitoring is often divided into the stages of fault detection, fault isolation, fault identification and fault diagnosis as suggested by Isermann and Ballé [53]. Figure 3.1 shows these stages. The first step is the detection of the disturbance, that is, the time when the disturbance affects the process. Fault isolation and identification follow the detection. Here, fault isolation involves locating the root cause of the disturbance while identification involves the characterisation of kind and size of the disturbance. The definition of fault detection is given by Isermann and Balleé as:

> "Fault detection is the determination of the faults present in a system and the time of the detection."

That is, the question "if" and "when" a fault has occurred is investigated by the detection. Patton et al. [90] point out that fault detection is more than making a binary decision whether the process is in normal operation and the presence of a disturbance. The challenge rather is to detect the fault at an early stage before it distributes through the plant. The reliable detection of small or early faults before an intervention is required are therefore objectives of a robust fault detection scheme.

The next step of fault diagnosis is defined Isermann and Ballé as follows:

> "Fault diagnosis is the determination of the kind, size, location of a fault which includes fault isolation and identification."

In this work, the focus is on fault isolation and retracing the fault propagation path of the disturbance. In most case studies investigated here, the disturbance has been detected and the number of processes variables to be analysed has been narrowed down. A straight forward approach to fault detection are upper and lower control limits which raise an alarm if a threshold value has been exceeded. Fault isolation and identification is sometimes tackled in a single step by arguing that the process variables with the largest disturbance characteristic is also the closest to the root cause. As an example, Thornhill [137] uses a nonlinearity index to trace the root cause.

3.1.2 Benchmark Criteria

Benchmark criteria are desired when comparing a number of process monitoring approaches. Venkatasubramanian et al. [141] provide a list of desirable characteristics that allow the choice of the best method for a given monitoring task. The list of properties and features of a fault monitoring system is given in Table 3.1. The list is very generic to allow a comparison of a variety of monitoring approaches. The first characteristic, for example, a fault detection and diagnosis facility is concerned with the early detection of faults and the diagnosis of the correct fault.

Not all benchmark criteria are equally important for all applications, for example, if no model of the process exist, characteristic 9. is very important and less important if a detailed model is already available. The focus of this work is on the feature of explanation facility. The data-driven causality methods developed in the next part find the root cause

Characteristic	Description
1. Fast detection & diagnosis	Early detection of faults. Trade-off: tolerable performance during normal operation, might be too sensitive to noise which results in frequent false alarms.
2. Isolability	Distinction between different faults. Trade-off: rejection of modelling uncertainties.
3. Robustness	Insensitiveness to noise and uncertainty. Trade-off: performance.
4. Novelty identification	Ability to decide if a detected fault is due to a known fault class or due to an unknown novel fault.
5. Adaptability	Adapting to process operating changes due to changing environmental conditions, e.g. product quantities, quality of raw materials or demands.
6. Explanation facility	Finding the fault origin and root cause as well as propagation to the current situation and identifying cause and effect relationships.
7. Multiple faults	Ability to identify more than one fault present at the same time.
8. Classification error	A priori estimate of classification error that occurred to confidence level.
9. Modelling requirements	Amount of modelling required for development of diagnostic system.
10. Storage requirements	Amount of memory storage necessary for operation
11. Computational effort	Amount of computation capacity required during operation.

Table 3.1: Desirable characteristics of process monitoring system after Venkatasubramanian et al. [141].

and argue the fault propagation path through cause-and-effect analysis. Thus, the question of fault origin and the propagation to the current situation is addressed. Several test can be carried out concerning the robustness, adaptability and isolability which will be carried out later on in this thesis.

3.1.3 Control Loop Performance Monitoring

Control loop performance monitoring (CLPM) exploits the well understood behaviour of traditional PID control loops. CLPM is usually conducted online and used as a non-invasive tool. A famous and established technique for control loop assessment is based on the diagnosis using minimum variance control. In 1967, Åström [158] introduced the minimum control variance (MVC) principle. MVC feedback control achieves minimum

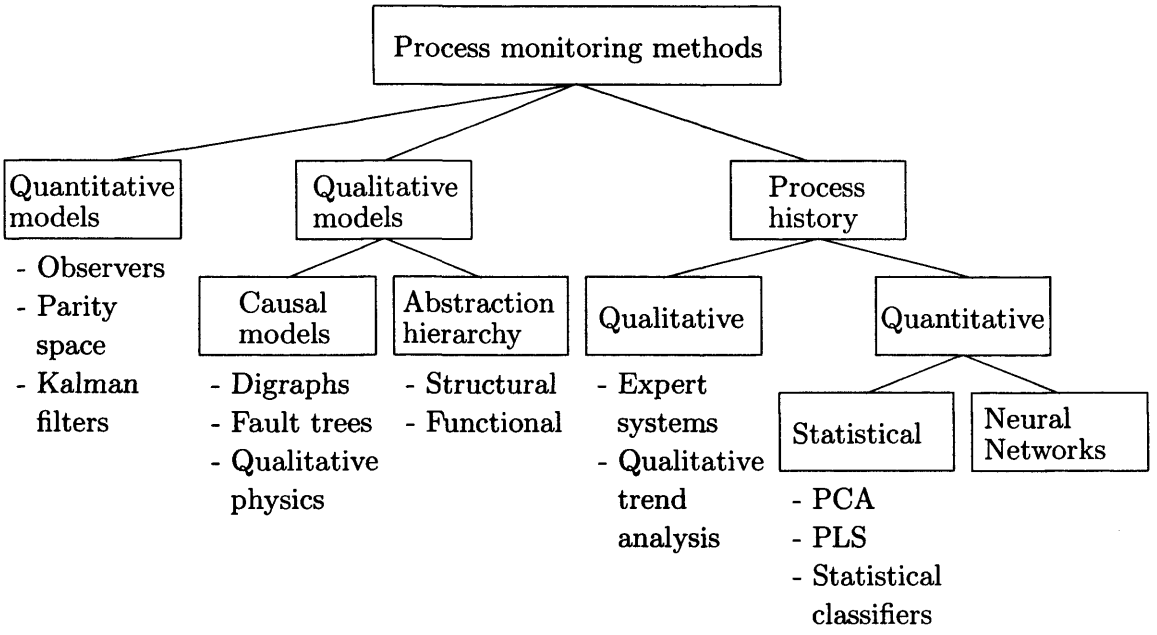


Figure 3.2: Process monitoring and fault diagnosis methods as organised by Venkatasubramanian et al. [141].

variance of the controller output from a process and disturbance model. The models are for example derived from the time trend through an autoregressive moving average (ARMA) model. Harris [42] proposed the use of MVC to assess the performance of a control law and thus initiated the developments of numerous loop performance assessment techniques. The resulting index is often referred to as the Harris index and is implemented in most commercial process monitoring systems, see Section 3.5. Qin [98] gives an overview of MVC and other monitoring methods. Hägglund [41], for instance, proposes an automated control loop performance monitor based on oscillation detection together with a diagnostic framework, exploiting the oscillations that can occur in malfunctioning loops. Thornhill and Hägglund [132] extend the diagnosis through spectral analysis and plotting controller set point versus the process variable. A linear quadratic Gaussian benchmark has been proposed by Huang and Shah [49] for models of PI and PID loops. More recent developments include an overall loop performance index (OLPI) by Xia and Howell [152] and a generalisation of the MVC principle by Harris [43].

CLPM can be combined with process monitoring by investigating both the process performance and the control performance separately and comparing or supplementing the results. A first approach undertaken by the joint project of ABB/University College London to develop a plant-wide disturbance analysis tool. The ABB/University College London tool is described later in Section 3.5.3. An evaluation of control performance monitoring algorithms from an industrial perspective was carried out by Desborough and Miller in 2001 [21].

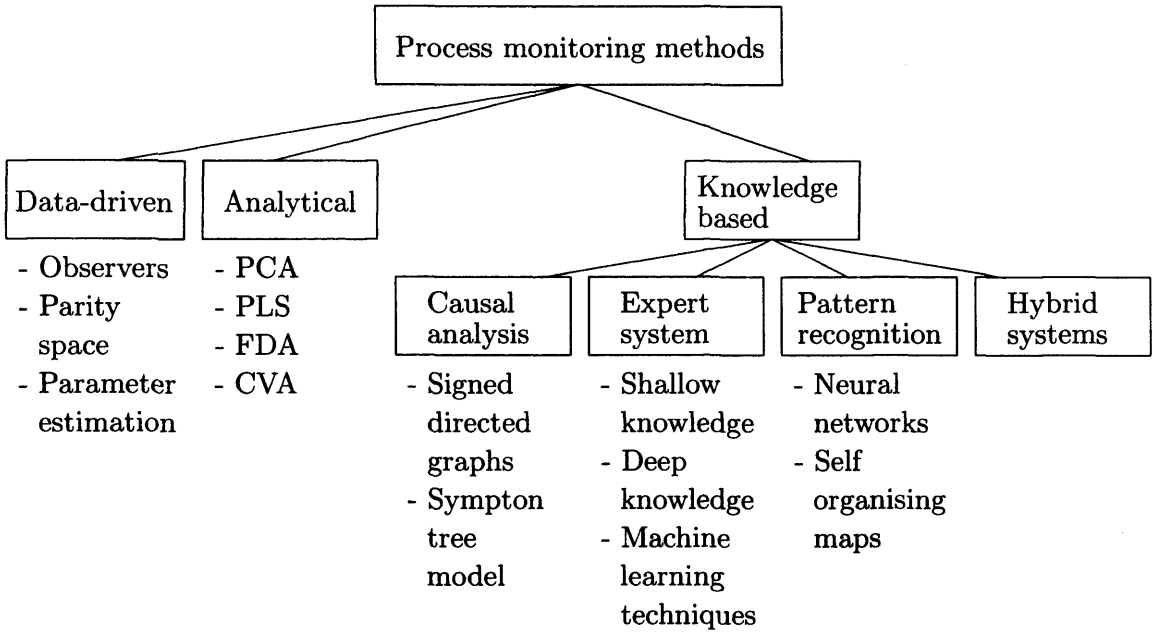


Figure 3.3: Process monitoring and fault diagnosis methods as organised by Chiang et al. [14].

3.2 Organisation of Process Monitoring Methods

Process monitoring methods can be classified in a number of ways. In his early textbook, Himmelblau [44] divides the techniques of fault diagnosis into two categories: estimation of variables or model parameters and pattern recognition methods. Estimation techniques require a model and are therefore also called model-based. Pattern recognition techniques on the other hand use process data comparing normal operation and disturbance data and are therefore also referred to as data-driven methods.

Venkatasubramanian et al. [141, 142, 143] follow the categorisation by distinguishing between the type of knowledge, which is either a priori or historical knowledge. Model-based methods require a priori knowledge to estimate model faults while data-driven methods extract the information from quantitative historical process data. The classification of process monitoring methods in [141, 142, 143] is shown in Figure 3.2. A further differentiation between quantitative and qualitative models is made. Quantitative models usually express the relationship between inputs and outputs in terms of mathematical functions while qualitative models, by contrast, express the relationship in terms of qualitative functions centered around different units in a process. Venkatasubramanian et al. also differentiate between qualitative and quantitative process history based methods. Quantitative approaches formulate the process monitoring problem as a data-driven problem, for example, to find common features in a number of time trends. Qualitative process history based approaches extract statements from the time trend that are represented by a qualitative model.

Chiang et al. [14] focus on data-driven methods which is reflected in their classification of process monitoring methods, shown in Figure 3.3. Here, the differentiation criterion is the

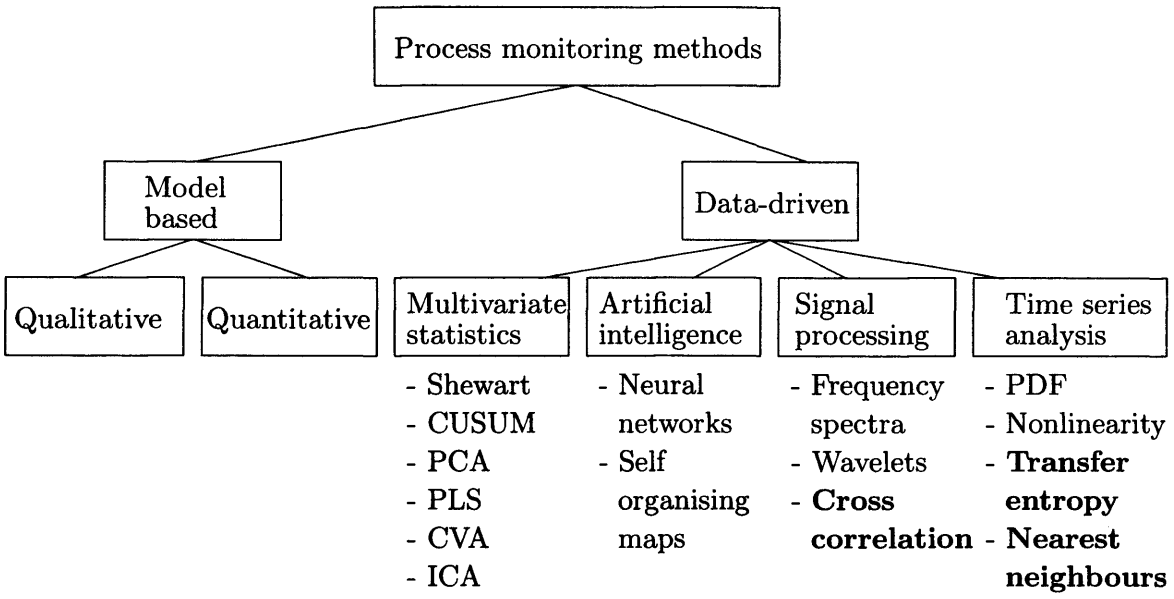


Figure 3.4: Process monitoring and fault diagnosis methods as organised in this work. Methods developed in this work are highlighted in bold font.

type of diagnostic search strategy rather than the kind of knowledge or whether the system is model-based. The first level of the organisation is therefore the differentiation amongst analytical, data-driven and knowledge based methods. Hybrid systems are introduced as a subclass of knowledge based techniques.

The focus of this work is on data-driven methods. The organisation of the review of process monitoring methods is therefore structured in more detail for the data-driven methods, as shown in Figure 3.4. The first category are statistical methods, often referred to as multivariate statistical process control (MSPC). A number of applications of artificial intelligence methods - neural networks and self organising maps - can be found in the literature. The third category is here called signal processing techniques, which could also be regarded as deterministic methods. An example is a cross-correlation algorithm which is developed in 5. Statistical methods that are usually not grouped under the term MSPC are here referred to as time series analysis methods. Transfer entropy and the nearest neighbours method developed in Chapters 6 and 7 for process monitoring purposes are examples of these time series methods. In the next section, established techniques listed here will be reviewed in more detail.

In most chemical processes, accurate dynamic models are normally not available although some qualitative guidelines might be known by the process engineer. In this case, qualitative model-based methods are of importance since a qualitative model is derived as part of the developed monitoring system in this work, see Section 4.3. Thus, Section 3.4 gives an overview of the development of digraphs and signed digraphs. No models were available for the case studies in this thesis.

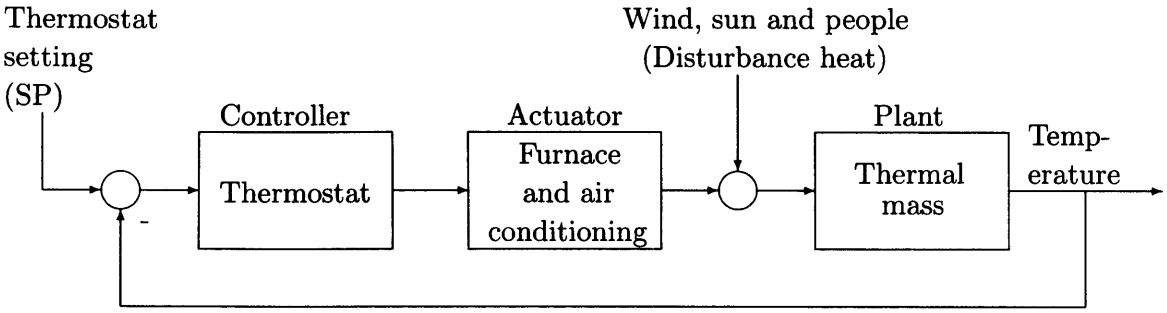


Figure 3.5: Example of model description of disturbance: Home heating and air conditioning control system by Ellis [25].

A number of reviews on quantitative model-based methods have been published by Venkatasubramanian et al. [141, 142], Frank [30] and Isermann [53]. Quantitative models incorporate the disturbance into the process model. As an example, Ellis [25] give the disturbance modelling of a home heating and air conditioning control system, as shown in Figure 3.5. Here, the disturbances enter the system after the actuator, the furnace and air conditioning, and before the plant, represented by the thermal mass. Controller, actuator and plant are described by transfer functions. If all these models are available, a quantitative model-based approach is preferable over a data-driven approach since all information about the process can be incorporated.

3.3 Data-Driven Methods

Data-driven methods are directly applied to process data. The measures extract features from high dimensional data providing information on the state of the process. An advantage of data-driven methods is the application to complex processes which are too large to allow derivation of a process model without excessive effort. The techniques are non invasive, can be applied to various processes and situations and work with data available from standard data acquisition tools. A disadvantage is that their proficiency is dependent on the quantity and quality of the process data used for the analysis. The structure of the following sections follows the classification of the process monitoring methods in Figure 3.4. Cross-correlation, transfer entropy and nearest neighbours, highlighted in bold font, are developed in this work and explained in the methods part.

3.3.1 Multivariate Statistics

In multivariate statistics, a number of measurements taken at the same time instance or batch are analysed. Objectives are compression or extraction of common data features, such as a spiky signal present in a number of measurements. For process control, the application of multivariate statistics is referred to as multivariate statistical process control

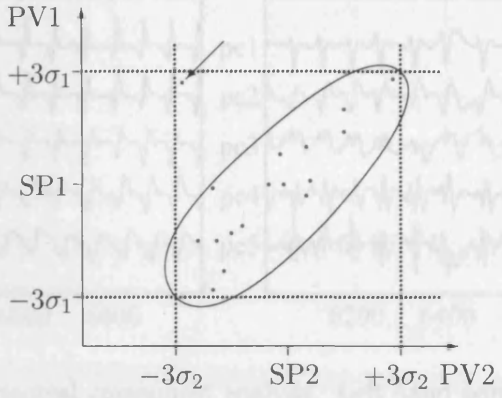


Figure 3.6: Example of multivariate Shewart chart.

and in particular for the application to chemical processes as chemometrics[1]. Overviews of MSPC are by MacGregor and Kourti [77] and Kourti and MacGregor [66, 67]. In 2002, a special edition of the IEEE Control Systems Magazine was dedicated to MSPC [96, 68]. Most frequently used MSPC methods are reviewed in the following paragraphs and include Shewart and CUSUM charts, principal component analysis (PCA), partial least squares (PLS), canonical variate analysis (CVA) and independent component analysis (ICA). In their textbook, Chiang et al. [14] apply PCA, PLS, CVA and Fisher discriminant analysis to simulation data from the Tennessee Eastman Process [23] and compare the efficiency of these methods for fault diagnosis. Multivariate statistical methods have been previously applied in a large variety of research fields to financial, medical, geographical or socio-demographical data. Textbooks on these methods are by Anderson [1] or Srivastava [124].

Shewart and CUSUM Charts: Shewart and cumulative sum (CUSUM) charts were developed as univariate statistical methods for process monitoring purposes. Extension to multivariate methodologies followed but since the emphasis is on the simultaneous investigation of several variables, both Shewart and CUSUM charts have been replaced by truly multivariate methods such as PCA and the like. In Shewart charts, key process variables are monitored and compared to a target value. If an upper or lower control limit is exceeded an alarm is raised. The limits are usually set to $\pm 3\sigma$, that is, the set point plus or minus three times the standard deviation under good operating conditions. An extension to multivariate Shewart charts is shown in Figure 3.6 where PV1 is plotted versus PV2 for the same time or sample instance. The control limit is here defined as the oval shape. Values, for example a point close to PV1 equals $+3\sigma_1$ and PV2 equals $-3\sigma_2$ as indicated by an arrow in Figure 3.6, are not permitted since they are outside the control limit. Thus, the multivariate case can give better performance than the univariate case, see [68]. CUSUM charts act on a similar principle as the Shewart charts by cumulated summing of the difference between a measurement over time and expected value. If the

[1]The term 'chemometrics' was coined by Bruce Kowalski, founder of the Centre for Process Analytic Chemistry in the 1980s[96]

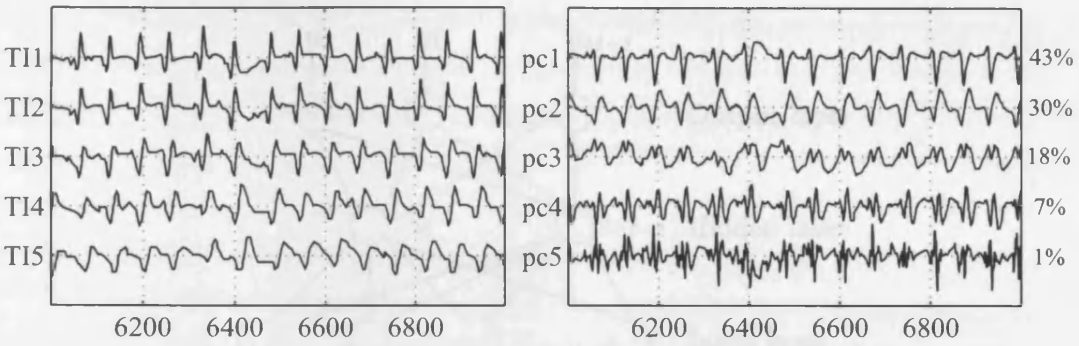


Figure 3.7: Example of principal component analysis. Left hand panel shows the original time trend while the right hand panel shows the five principal components pc1 to pc5. The percentages on the right give the percentage of variation contained in each principal component.

sum exceeds a threshold an alarm is raised. CUSUM was introduced by Page [88]. An overview of a multivariate extension of CUSUM charts is given in [104].

Principal Component Analysis (PCA): Principal component analysis is a frequently used multivariate statistics method in process monitoring. PCA was developed for data analysis and compression. The principal components are the characteristic vectors of the covariance matrix of the data set. The first principal component is the normalised linear combination of statistical variables with maximum variance. A statistical measure is derived by evaluating the power of the first principal components, applying Hotellings T^2 statistic or by comparing the residuals of each variables as the squared prediction error (SPE) [1]. PCA is illustrated in Figure 3.7 using the example introduced earlier in Figure 2.10. The first principal component, pc1, shows the spiky regular pattern which is prominent in all signals and contains 43% of the variation of the five time signals. PCA can thus be used for selecting and clustering process variables. PCA has been applied to process monitoring since the early 1990s for example by Kresta et al. [69] to overcome the problem of data overload arising from the large number of discrete measurements captured by plant sensors. Further investigations followed [151, 101]. The PCA algorithm has some shortcomings such as the inability to deal with time lags or large numbers of process variables. A number of variations of the original method, such as multi-block [148], consensus [99] or spectral [134] PCA, have been reported to address these problems. Most recent applications of PCA for process monitoring are by Choi et al. [17] and Yoon and MacGregor [154].

PLS, CVA and ICA: Partial least squares (PLS), canonical variate analysis (CVA) and independent component analysis (ICA) are linear dimensionality reduction techniques similar to PCA. PLS maximises a predictor or independent matrix and a predicted or dependent matrix for each component of the reduced space. PLS is often employed in statistical quality monitoring by selecting the dependent matrix to contain only product quality data while the independent matrix contains all other process variables. PLS was introduced around the same time as PCA in the early 1990s [69] and more recent appli-

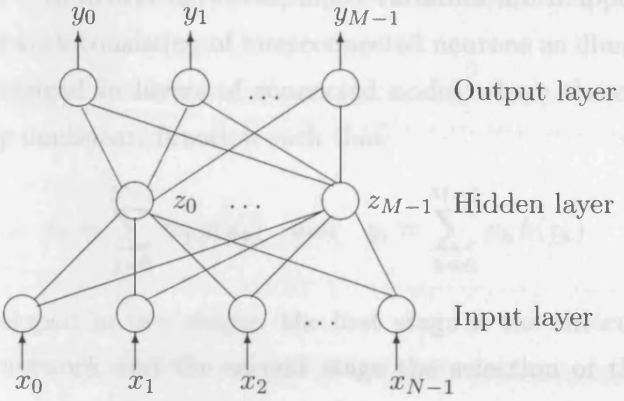


Figure 3.8: Two layer topology of a neural network: input, hidden and output layer.

cations to process monitoring are [40, 65]. CVA maximises a correlation measure between two sets of variables. It is related to PLS as it involves a singular value decomposition but the difference from PLS is that CVA maximises correlation whereas PLS maximises covariance. Scharper et al. [109] apply CVA to fault diagnosis and Russel et al. [105] compare CVA to PCA and found that CVA gave the best overall sensitivity and early detection of all faults. A combined index of CVA and PLS is applied to a continuous stirred tank co-polymerisation reactor by Simoglou et al. [120].

The basic assumption for ICA is that the new components are statistically independent, that is, the joint probability density function is factorisable. Li and Wang [74] demonstrate the use of ICA for process monitoring on a case study of two continuous stirred tank reactors, Kano et al. [56] investigate ICA for the purpose of fault diagnosis. Spectral ICA is proposed by Xia et al. [153] to analyse the frequency spectra rather than the time trends.

3.3.2 Artificial Intelligence

Artificial intelligence (AI) techniques are often grouped into the area of pattern recognition. Like a human brain, AI methods are trained with a number of examples. After the training period, the AI system is able to classify decisions with an input data set. An advantage of AI systems is the extraction of complex and potentially nonlinear input features. As a trade-off, large data sets for the training period are required and computational effort increases easily beyond practicability. Here, the application of two techniques of AI for process monitoring and fault diagnosis are reviewed. The two applications are neural networks and self organising (feature) maps. The use of neural networks in fault diagnosis is widespread and many examples have been described in the literature. Self organising maps are a more recent development, often more complex and less frequently employed for fault detection. A comparison of the two methods for chemical problems in monitoring and control is given by Zupan and Gasteiger [157].

Neural Networks: In neural networks, input variables are mapped onto output variables by a neural network consisting of interconnected neurons as illustrated in Figure 3.8. The neurons are arranged in layers of connected nodes where the connection represent weights or a, usually nonlinear, function such that

$$z_k = \sum_{j=0}^{N-1} w_j \, g(x_j) \quad \text{and} \quad y_i = \sum_{k=0}^{M-1} v_k \, h(z_k) \tag{3.1}$$

The network is developed in two stages: the first stage is the choice of the architecture or topology of the network and the second stage the selection of the learning strategy for developing strong connections. The most frequently used topology in fault diagnosis is a three layer neural network and the most common learning strategy is a feedback propagation algorithm. The neural network can be used for fault diagnosis by assigning the input neurons to process variables and the output neurons to fault indicators. A large number of data sets is needed to train the network. Sorsa et al. [123] used 3000 training cycles to train a continuous stirred tank reactor with 14 measurements and 10 faults. Becraft and Lee [4] study the hidden and input layer development during fault space training. More recent work includes the successful application of neural networks to an industrial reactor [84] and to the actuator of a process valve [61]. A general shortcoming of neural networks is the lack of novelty identification, as highlighted previously in Table 3.1.

Self Organising Maps: Self organising maps (SOM) were introduced by Teuvo Kohonen in the early 1980s are often also referred to Kohonen maps. The SOM consists of two layers: the input and the output layer. The high dimensional input data is projected on a two dimensional output grid[2]. The output "map" is a M by N array of reference vectors $\mathbf{r}_{m,n}$. The reference vectors are of the same dimension as the input vectors. At time $k = 0$, the map is initialised with random vectors for $\mathbf{r}_{m,n}[0]$ that have values within the range of the input vectors $\mathbf{v}[k]$. For the next time steps $k = 0...K$, the following two steps are repeated:

1. Find the reference vector $\mathbf{r}_{v,w}$ closest to input vector $\mathbf{v}[k]$ through minimising $\|\mathbf{v}[k] - \mathbf{r}_{m,n}[k]\|$.
2. Adjust the reference vectors in the proximity of $r_{v,w}$ the according to $\mathbf{r}_{m,n}[k+1] = \mathbf{r}_{m,n}[k] + k(m,n,\mathbf{r}_{v,w},d,k)a\left(1 - \frac{k}{K}\right)(\mathbf{v}[k] - \mathbf{r}_{m,n}[k])$ where k is the neighbourhood kernel function, d the kernel radius and a the learning rate.

A new class of variables is created in the form of a newly stored pattern in $\mathbf{r}_{m,n}$ whenever a pattern with a high distance is applied to the SOM. For fault detection, a SOM is trained with data from normal operating conditions. A fault can be detected by monitoring the distance between the observation and the pattern closest to the observation. Ignova et

[2]http://www.sbc.su.se/ maccallr/thesis/node137.html, March 2005.

al. [50] find the SOM to be a potentially useful tool for statistical quality monitoring by applying it to seed data of a fermentation process. Jamsä-Jounela et al. [54] apply SOMs for fault detection and diagnosis to a copper flash smelting process. SOMs in general require even more training data than neural networks but unlike neural networks, can deal with new fault categories.

3.3.3 Signal Processing

In this section, methods are grouped which are usually placed in the area of signal processing. Signal processing methods are used for the analysis of signals in many electrical engineering areas, such as communication, control and power electronics. Here, the focus is on the deterministic analysis tools of frequency spectra and wavelets which will be reviewed in the following paragraph.

Frequency Spectrum: One of most commonly used technique is the representation of a signal in the frequency domain via the Fourier transform. The Fourier transform splits a signal into sine waves with varying oscillation periods. Amplitude and phase of the oscillations present in the signal are plotted over the frequencies. The square of the amplitude spectrum is often also referred to as the power spectrum. The frequency axis is measured in seconds^{-1} =Hz or minutes^{-1}. Oscillating signals can best be observed in the frequency domain. The following list of useful applications of frequency analysis for fault diagnosis and process monitoring purposes has been made during the placement in the Advanced Controls Technology Group Eastman Chemical Company:

- The frequency spectrum can be used to focus on a smaller number of measurements which show a specific oscillation (fault detection);
- If oscillation are present in several variables then higher amplitudes may indicate that closeness to the root cause;
- Several oscillation periods in several signals can be detected and compared;
- Multiple oscillations in one signal are easy to detect;
- Harmonics at multiples of the main frequency may be caused by nonlinearities and tend to be close to the root cause;
- Various faults have different frequency ranges.

Caution has to be taken for some signals unsuited for frequency analysis. This is the case for the following situations.

- Frequency analysis is unsuited for irregular patterns;
- A drifting mean must be removed before generating the frequency spectrum;
- Some harmonics are due to the PI[a] compression algorithm, see Section 10.2.

[a]PI here refers to the PI data historian by OSIsoft, see Section 3.5.1, not the PI control.

The spectral analysis is carried out in conjunction with principal component and independent component analysis as described in previous Section 3.3.1. In the example in Section 4.2, Figure 4.4 shows the frequency spectrum of the time trend displayed in Figure 4.3. The main period of around 0.05 min^{-1} can be clearly observed in most process variables. Harmonics at multiples of this frequency are visible in some of the spectra.

Wavelets: The investigation of the power spectrum views the oscillations in a signal while disregarding all time information. Wavelet analysis combines both time and frequency analysis by decomposing the signal into elementary building blocks of wavelet transforms and plotting the blocks for a number different frequencies over time. An application of wavelet analysis to sensor fault detection and diagnosis is proposed by Zhang and Yan [156]. Different sensor faults, such as bias, cyclic, stuck, spike and erratic, show distinct features in one or more of the frequency ranges at different times. Tse et al. [139] apply wavelets to utilize the time localisation of vibrating signals of rolling element bearings.

3.3.4 Time Series Analysis

A number of methods investigate the statistical properties of a time series. These methods are here grouped under the term time series analysis to differentiate them from the methods traditionally referred to as multivariate statistics, see the previous Section 3.3.1. In the textbook by Kantz and Schreiber [57] methods are described dealing with the description of chaotic signals and properties such as determinism, predictability and nonlinearity. Only few methods from this area have been used for process monitoring and the two main contributions of this work, transfer entropy and the nearest neighbours method, originate from this area. The statistical method of probability density function is, however, widely used for fault diagnosis and is reviewed here. A method based on the same principles as nearest neighbours method is the nonlinearity detection by [57]. Thornhill et al. [136, 137] introduce this method for fault diagnosis, see also Section 6.2.3.

Probability Density Function: A full description and discussion of the probability density function (PDF) is deferred to Section 7.1 where it will be explored in detail. A random variable is fully described by its PDF and the PDF is therefore used for two purposes: data compression and analysis. Desforges et al. [22] use the PDF to detect abnormal and unexpected process conditions from measured response data. Wang and Lin

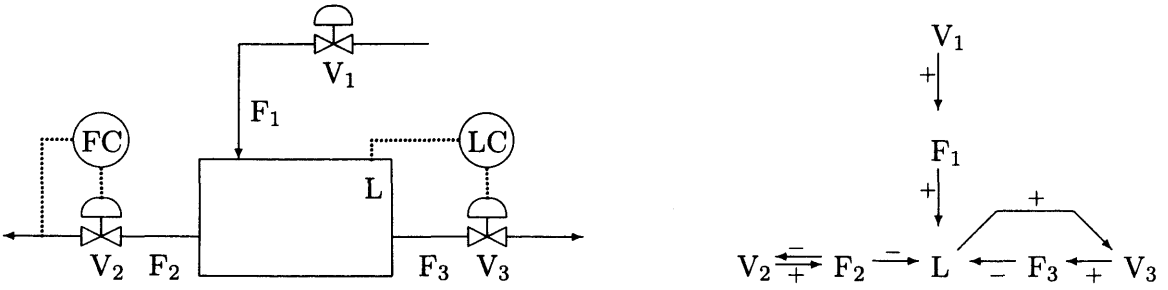


Figure 3.9: Signed digraph derived from expert knowledge: process schematic (left) and signed digraph model (right) of buffer tank process by Iri et al. [52].

[145] combine the analysis of probability density function with observer based techniques by applying the PDF to the system output signal which is then fed back to the input of the control system. Forbes et al. [29], more recently, use the shape the PDF of process states for control and monitoring purposes. Horch [46] uses the PDF to decide whether a loop fault is due to stiction or non-stiction of the control valve.

3.4 Qualitative Models

In this thesis, the data-driven methods described in the previous section are supplementary to the data-driven causality measure derived in this work. They are used to pre-process the data or to gain additional information about the process. In this section, causal qualitative models are reviewed that give the context for the causality methods and derivations developed in this work as they are representations of the cause and effect relationships in a process unit. The resulting models are digraphs or signed digraphs. A digraph is a graph with directed arcs between nodes where nodes represent process variables and the arcs represent the relationship between the process variables. Signed digraphs have additionally a sign attached to each arc where the sign indicates whether the dependent variable increases or decreases. All qualitative models require some knowledge about the process for the construction of a causal map or digraph. The process knowledge can either be a description of differential algebraic equations, expert knowledge of the process systems engineer or representations of the process schematic. These three approaches will be reviewed in the following sections. For an overview of qualitative models see the second part of the process monitoring assessment by Venkatasubramanian et al. [142].

3.4.1 Expert Knowledge

When first introducing signed digraphs for fault diagnosis and process monitoring, qualitative states were assigned to each unmeasured node in the digraph. Figure 3.9 shows the example of a buffer tank by Iri et al. [52] and the corresponding expert knowledge. The tank in Figure 3.9 has one inflow and two outflows. The inflow F_1 is controlled by

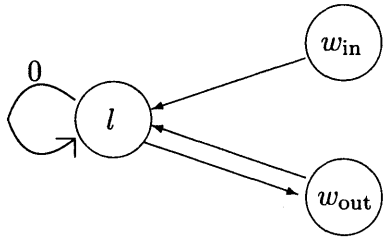


Figure 3.10: Digraph for the differential and algebraic system of a controlled level tank, see Section 2.3.

valve V_1, the first outflow F_2 by flow controlled valve V_2 and the second outflow F_3 by level control V_3. The causal map in the right hand panel of Figure 3.9 is derived through expert reasoning. For example, if the inflow valve V_1 opens, inflow F_1 increases. This is visualised in the causal map through an arrow pointing from V_1 to F_1. The plus sign next to the arrow means that the flow is increasing. On the other hand, if outflow F_2 increases, then level L decreases. The arrow pointing from F_2 to L is therefore labelled with a minus sign. Nam et al. [85] propose a scheme for constructing extended symptom-fault associations (ESFA) from the signed digraphs and thus identify the fault propagation path. This method has the same objective as the causality measures derived here. However, a construction of ESFAs is only possible if a signed digraph has been derived which is usually only available for less complex processes, such as a buffer tank or a continuous stirred tank reactor.

3.4.2 Differential and Algebraic Equations

A systematic framework for deriving digraphs from mathematical models is given by Maurya et al. [80]. The basis are firstly a differential equation, for example $\frac{\mathrm{d}x}{\mathrm{d}t} = f(y_1, y_2)$. The variable x on the left is called endogenous and the variables y_1, y_2 on the right side of the equation are called exogenous. Directed arcs are drawn from the exogenous to the endogenous variables since a constant value of y_1 or y_2 results in a change of x. Therefore, y_1 and y_2 influence x. If $\frac{\mathrm{d}x}{\mathrm{d}t}$ is not a function of x then a zero arc is drawn to node x. Algebraic equations, on the other hand, such as $y_2 = g(x)$, capture instantaneous behaviour and thus contain no causal information. However, strongly connected components can be identified by partitioning the variables into subsets and finding the dominant direction of causality. The developed digraph model can use propagation through the graph to predict the behaviour of the system. Signs can be assigned to the arcs to indicate the direction in which one variable will influence the other, that is, increase or decrease. The sign of the arc is equal to the partial derivative of the endogenous variable to the exogenous. The resulting graph is called signed digraph (SDG). Maurya et al. [81] deduce a SDG for fault diagnosis for the simple example of a continuous stirred tank reactor.

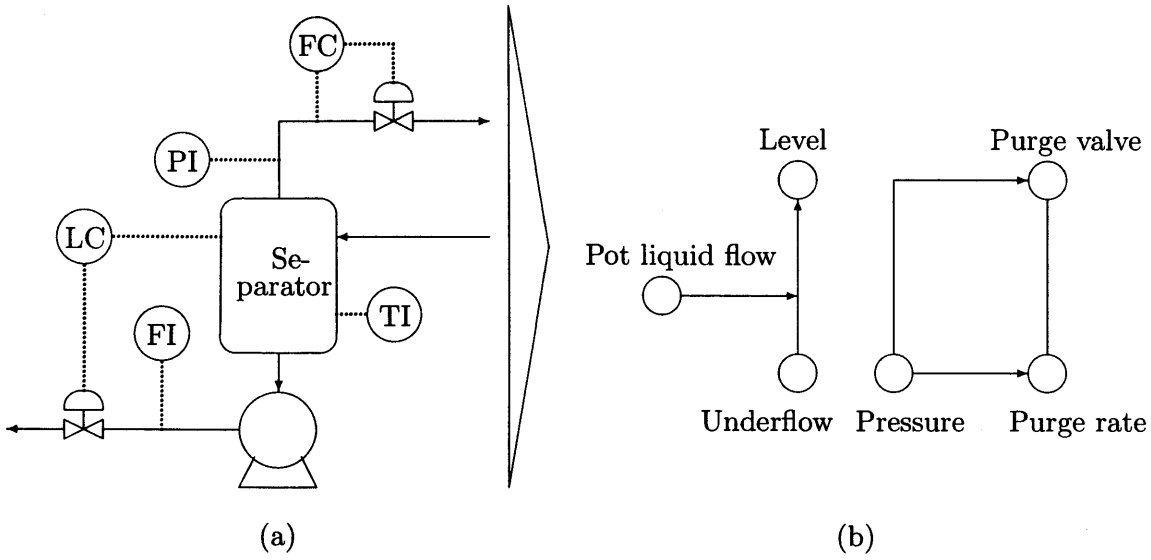


Figure 3.11: Causal map derived from process schematic: A part of the process schematic (a) of the Tennessee Eastman Process together with a digraph construction (b) by Chiang and Braatz [16].

Example: *For constructing the digraph from differential and algebraic equations consider the controlled level tank from Section 2.3. The system was described by the differential equation*

$$A\frac{dl}{dt} = w_{in}(t) - w_{out}(t) \tag{3.2}$$

with inflow $w_{in}(t)$ and outflow $w_{out}(t)$ and $l(t)$ the level of the tank. For this equation, two arcs are drawn from w_{in} and w_{out} to the level l. The feedback control law was set to be

$$w_{out}(t) = K_c l(t) \tag{3.3}$$

which gives an additional algebraic equation for the relationship of l, w_{in} and w_{out}. Since the level l controls the flow w_{out}, a further arc is drawn in the opposite direction from l to w_{out}. The digraph of the controlled level tank is shown in Figure 3.10. The arrow pointing from variable l to itself attached with a zero indicates that the derivative of l is used. Small examples like this are rarely useful for process analysis. However, if mathematical descriptions of large processes exist, a good overview of the behaviour can be achieved through digraph construction. ———————————————

3.4.3 Process Schematic

A digraph can be derived through expert knowledge by using the process schematic as described previously in Section 3.4.1. The process variables are indicated along the flow line and their consequential order depends on the direction of flow. Deriving the digraph from process schematics is often labour intensive and not always straight forward. Not all relationships between the variables can be investigated so that attention has to be paid to the ones in which the disturbance occurs. There is no differentiation between the case of

normal operation and disturbance. Chiang and Braatz [16] construct a digraph model of the Tennessee Eastman Process from knowledge of the process and the covariance matrix. An excerpt of that process model together with the process schematic is shown in Figure 3.11. Fault diagnosis is conducted by an entropy related measure of the Kullback-Leibler Information Distance (KLID) only after the digraph is derived. Maurya et al. [82] also deduce a signed digraph from the Tennessee Eastman Process using an initial response table for measured variables.

3.5 Commercial Data-Driven Monitoring Systems

Advantages of software based monitoring tools are their non-invasiveness, cost-effectiveness and their ability to therefore minimise downtime. All monitoring systems are usually online and use normal operating data received from the standard data source OPC[3]. Desborough and Miller [21] describe the use of a process monitoring system as follows.

> "The high-level goal of a process control monitoring system is to provide plant control engineers with enhanced capabilities to identify problems for many controllers while minimizing additional effort or expenses."

In this section, data acquisition systems are described that are state-of-the-art in process control. The ability to acquire a good data set within the time frame when the disturbance occurs depends largely on the data acquisition system. A number of control and process performance assessment tools by Honeywell, Matrikon and Expertune are available which will be summarised in the next section. The process measurement are captured and discretised by the DCS or supervisory control and data acquisition (SCADA) systems. The captured data is sent to the data historian, for example the PI system. The monitoring and assessment tools fetch the data from the historian data base and process it to gain information about the process.

3.5.1 Data Acquisition Systems

Process monitoring systems are embedded in data acquisition systems, that is, they receive data and send back control parameters. Data acquisition systems gather data from the hardware at the instrumentation in the plant, such as distributed control systems or programmable logic controllers (PLC). Often in this context the term SCADA is used for control of industrial plants. It is more widely used in power plants and water systems

[3]OPC: OLE for Process Control. OLE is short for Object Linking and Embedding and is a programming language attribute.

where large areas have to be covered. Some applications are reported for chemical plants [3]. Four major components are part of every SCADA system [131]:

- A master station where computers, peripherals and input/output systems are combined to enable the operators to monitor and control the process;
- Remote terminal units (RTU) which acquire all data from field devices, process the data and transmit it to the master station. RTUs also distributes control signals to the field devices;
- Field buses for communication between the master station and the RTUs with a defined bandwidth;
- Human machine interface (HMI) for the interaction of the operators with the master station.

Most SCADA system have facilities to record, present and store data. However, the quality of the HMI decides about the use of the system for process monitoring tasks. In the chemical industry, the most frequently used data historian system is the PIsystem^{TM}, a software product by OSIsoft^{TM} [4]. OSIsoft claims 5000 applications have been installed in large plants world wide. The focus of the PIsystem is on data acquisition but some monitoring facilities are incorporated. Its popularity is also due to a large number of interfaces to all major automation vendors such as ABB, Honeywell, Rockwell and Siemens. In the following paragraphs the main facilities of the PIsystem are listed together with their significance for process monitoring purposes. The data acquisition system is of particular importance when dealing with data quality issues such as compression, quantisation or sampling problems, described in Chapter 10.

PI data historian: The data historian acquires, processes, displays and stores process data. The system has real time facilities that allow process control, alarms and operations. The data is stored in a data historian (PI DataStorage) using the swinging doors compression[5]. From there, historical data can be retrieved by individual users of PI and exported as plain text or Microsoft Excel files.

PI server applications: A number of applications are implemented on the PI server to carry out monitoring tasks, such as alarms, batch events or statistical process control functions. PI alarm tracks and manages predefined alarm conditions which can be triggered by the duration of an event or deviation from norm. Alarm conditions are also stored in the PI data historian. A batch manager identifies processes stages and measures repeatability of batch processes. The events can be hierarchically grouped. The PI real-time statistical quality control stores test results and records control limits.

[4] http://www.osisoft.com; Information on the PI data historian by OSIsoft, March 2005.

[5] For the algorithm of the swinging door compression see Section 10.2.

PI module database: The PI module database manages equipment data such as the equipment specifications, manufacturer data and photographs. Process data can be stored and allocated to a piece of equipment. Additionally, all configuration changes are recorded. The data is structured into hierarchies and connectivity models to incorporate process knowledge. However, the process knowledge is not processed any further to automatically incorporate it into fault diagnosis.

3.5.2 Control Performance Assessment Tools

The publication of methods that assess the performance of control loops from the data trends, such as the Harris index [42], led to the development of a number of computer aided tools. These control loop performance assessment tools were developed by traditional engineering firms and new startups. John W. Cox, Eastman Chemical Company states: "As far as specific control performance monitoring vendor tools, we would view Matrikon ProcessDoctor, ExperTune PlantTriage, and Honeywell LoopScout as the three best current vendor tools.". The three tools are described in the following.

There are also drawbacks of these standardised and automated tools. In a publication in 1999, Eastman Chemical Company assessed the use of LoopScout [92] but chose to build their own loop assessment tool to for three primary reasons:

- Automated data collection at that time was limited to Honeywell control systems;
- Substantial amounts of process data would have to be sent to Honeywell, requiring complicating approvals;
- The cost to assess loops worldwide was prohibitively high given the emphasis on reducing business expenses.

The control performance assessment facilities of the tool developed by Paulonis et al. are described in [92].

ExperTune's PlantTriage: PlantTriage is a monitoring and diagnostic system that features a reporting scheme for economic performance criteria[6]. Up to 40 analytical measures are calculated per control loop to assess the performance of the loop. The assessment takes place at regular intervals for each unit of the plant. In particular, an oscillation detection, identification and diagnosis scheme scans all control loop. The analysis of the oscillation detection is based on the frequency spectrum and in a first step the periods and strengths are identified. The period of oscillation is measured as peaks in the spectrum and displayed in seconds. The strength is calculated from the percentage of energy of the identified peak compared to the total energy of the spectrum. The first three harmonics are considered in the analysis. The detected oscillations are classified into three categories:

[6] http://www.expertune.com/planttriage.html, PlantTriage software by ExperTune, March 2005.

- Load - the disturbance originates from load upset or from interaction with other loops,
- Tuning - aggressive tuning is the cause of the disturbance,
- Valve - the disturbance is caused by valve stiction or hysteresis.

A confidence level is established for all categories ranging from 0% to 100%. The higher the percentage, the more confidence can be placed in this diagnosis. This gives more flexibility in the interpretation since there is, for instance, no category allocated for root causes other than load, tuning or valve problems. A basic mechanism of fault propagation is suggested by ExperTune through grouping loops with the same oscillation period and checking if one of the loops has a high confidence level for valve problems. The argumentation is very rudimentary but practical and works for distributed disturbances caused by sticking valves. In addition to the oscillation analysis a control performance assessment is conducted for each control loop. Here, traditional indices such as the Harris index [42] are implemented. A further feature is the "Control System Availability" which is specifically designed to detect controllers in manual, controllers that are stuck at a limit or with frequent mode changes. A data historian and trending facility plots process variables that are received from the OPC historical data access. Expertune is currently extending its software to the application of batch processes.

Honeywell's Loop Scout: A frequently used control loop performance monitoring system is Honeywell's Loop Scout^{TM} [7] which is completely based on the internet, thus all the data remains with Honeywell. Pattern recognition techniques are applied to distinguish between tuning problems and valve problems. Characteristics of the loop are compared to pattern which are known to be caused, for example, by stick slip behaviour and which are stored in a data base. The performance classification is similar to the PlantTriage software, not exact but only a recommendation provided with a confidence level for the recommendation. Loop Scout^{TM} focuses on regulatory (SISO) rather than on multivariable (MIMO) controller assessment. The statistical tools employed by the software package include:

- Hysteresis characterisation - PV-OP plots and hysteresis metrics;
- Scatter plots and pattern recognition - for oscillation analysis and hysteresis identification;
- Minimum variance analysis - controller error analysis and computation of theoretical minimum variance;
- Spectral analysis - frequency plots that are linked with poor performance are identified and spectrum of PV is analysed.

[7] http://www.loopscout.com, March 2005; LoopScout process monitoring software by Honeywell; for a free trial version visit the website.

The assessment also provides metrics to detect certain states of the loop such as wind-up, broken cascades, control in manual, control mode off normal and recognition of patterns of known faults. A data historian is provided which is a shared facility with other Honeywell applications. The time series data is collected in RMPCT's MPT (multipoint trend format). For each controller it collects configuration information as well as 5000 samples of PV, SP and OP time series data. The default sample frequency values are 1 second data for flow loops, 5 second data for pressure loops and 30 second data for both temperature and level loops. A recent feature in Honeywell's Loopscout is the criticality index that weighs the loop performance index by a factor to decide the importance of the loop.

Matrikon's ProcessDoctor: The software package ProcessDoctor by Matrikon[8]. comprises different components for PID loops, model predictive control (MPC) and model identification. The package for PID loops identifies, prioritises and diagnoses control loop problems such as valve and tuning problems, oscillations and unwanted loop interactions.

3.5.3 Plant-Wide Disturbance Analysis Tool

At present, there are no known monitoring programs that use the data-driven methods developed for the investigation of plant-wide disturbance. A project between ABB and Imperial College / University College London Centre for Process Systems Engineering is under way to develop solutions for computer aided plant disturbance analysis [48]. The tool incorporates data-driven methods that draw information from the time trend of a recorded disturbance. The objectives are as listed in [48] as follows:

- Detection of the presence of one or more periodic oscillations indicated by a regular pattern in the measurements;
- Detection of non-periodic disturbances and plant upsets;
- Determination of the locations of the various oscillations/disturbances in the plant and their most likely root cause.

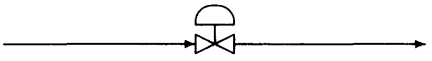
The detection of time trends with similar oscillation patterns is pursued with help of the threshold crossings of the autocovariance function as described in [136]. The tool uses spectral principal component analysis to cluster time trends with irregular disturbances and presents them in hierarchical classification trees [134]. Nonlinearity assessments of the time trends give conclusions towards the root cause as described in [137].

The methods developed in here will be incorporated in a plant-wide disturbances assessment tool as developed by the ABB / CPSE joint venture.

[8] http://www.matrikon.com/products/processdoc, March 2005; ProcessDoctor process monitoring software by Matrikon

Chapter 3 Summary

In this chapter, process monitoring techniques and systems were reviewed. Current industrial tools provide the facility of data-driven methods for automatic use. A literature search showed that no data-driven methods presently exists that retrace the fault propagation path and which can construct a qualitative model of the process using historical data.



Part II

Data-Driven Causality Methods

PART II - DATA-DRIVEN CAUSALITY METHODS

According to Desborough and Miller [21], the most common cause of disturbances in chemical plants are process problems such as oscillations caused by a recycle or interaction between control loops that compete for the same physical quantity. These disturbances usually affect a larger number of measurements and are therefore referred to as plant-wide disturbances. Unlike upsets caused by valve problems these disturbances can also not be classified or diagnosed easily. The root causes are manifold and the signature of the disturbance varies. For these reasons, identifying the fault propagation path assists in diagnosis of the root cause.

In the first chapter of Part II, an introduction to data-driven causality measures is given. Before describing the nature of causality measures and ways of representing measures for a number of variables, a reference case study is introduced. The causality problems and data trends of the reference case study are used throughout Part II of this thesis. Furthermore, the nature of causality measures is discussed and for graphical representation, digraph models of the system are constructed from the causality measure.

In Chapters 5, 6 and 7, three alternative ways of specifying the directionality of process variables from historical data are introduced and developed. These relatively recent statistical methods are available because of the development in the computational power of modern PCs. The simplest way is to measure the time lag between process variables using the cross-correlation function (CCF). The shortcoming of this method is that it requires a time lag in order to measure causality.

Two statistical methods presented here are the nearest neighbours and transfer entropy method. These have the key advantage of finding dependency also in case that no time lag exists between process variables. The reason for this is that the driver variable contains more process information than the second, response variable and thus a dependency can be measured. Thus, the driver variable can predict the response variable. The concepts of randomness, determinism, continuity and predictability are key to this work and will be explained in detail in Chapter 6.

Chapter 4

Introduction to Causality Analysis

This chapter provides an introduction to causality analysis by giving the prerequisites and properties for a root cause explanation facility. A reference case study for the development of the data-driven analysis methods is presented here. Graphical tools are investigated that allow the automatic visualisation of the causality analysis.

As a prerequisite of the causality analysis, the disturbance should already have been detected in a number of process measurements as done presently by the commercial controller performance tools. The objective of the causality analysis is to track the fault origin and to retrace the propagation path. The approach pursued here is to investigate the causal relationship between all combinations of the variables in which the disturbance was detected.

In the first section of this chapter, the causality measures are described and the expected descriptive form of the measure is formulated. In Section 4.2, a reference case study is introduced to which later on the developed causality measures will be applied. The reason for using a case study for the investigation is that simulated data is usually not rich enough to capture the stochastic structure of real plant data even when random noise is added. In Section 4.3, graphical representations are developed displaying the results retrieved from the causality analysis. The final graphical result is a causal map that shows the interaction between the process variables.

4.1 Nature of Causality Measures

Measures of causality draw information on dependency from time delays, functional attenuation and the noise content in the data. A measure of causality must exhibit an antisymmetric nature, that is

$$h_{X \to Y} = -h_{Y \to X} \tag{4.1}$$

where h is the causality measure and the plus or minus sign indicates whether X causes Y or vice versa. A positive value for $h_{X\rightarrow Y}$ is called driver or cause while the negative value is referred to as response or effect. The algorithms investigated in the literature do in fact not measure causality but dependence or predictability. The question of causality is addressed by asking the question:

> "Does X influence (predict) Y more (better) than Y influences (predicts) X?"

Dependency and predictability are indicated by positive values $H(X|Y)$ and $H(Y|X)$ which is also often referred to as coupling [86, 89]. If one of the two values is larger than zero one speaks of unidirectional coupling whereas both values $H(X|Y)$ and $H(Y|X)$ being larger than zero is termed bidirectional coupling. The bar here indicates conditional behaviour, that is X under the condition that Y is known.

Any quantity that measures the influence of variable X onto Y must be potentially asymmetrical to the inverted case of Y influencing X. In case of a symmetrical measure, which is the case for example for the linear correlation coefficient (see Section 5.1.1), correlation is measured instead of dependency and therefore no causality measure is achieved. The asymmetrical relationship is expressed by

$$H(X|Y) \neq H(Y|X). \tag{4.2}$$

After ensuring asymmetry of the measure, a quantity for causality is derived from the respective influences by comparing the quantity of X influencing Y with the quantity of Y influencing X. The comparison is achieved through taking differences.

$$\begin{aligned} h_{X\rightarrow Y} &= H(Y|X) - H(X|Y) \\ h_{Y\rightarrow X} &= H(X|Y) - H(Y|X) \end{aligned} \tag{4.3}$$

Here, h refers to the potentially bidirectional causality measure while H is a positive and therefore unidirectional measure. As a result, Equation 4.1 holds by definition. Furthermore, $h_{X\rightarrow X} = 0$ holds.

Multivariate Case: As a start, all relationships are represented in a matrix with the same order of the variables on both axis. The entries on the diagonal then represent the relationship between a variable with itself and are therefore zero, see the definition of the causality measure. The matrix is referred to as causality matrix in the following.

$$\mathbf{h} = \begin{bmatrix} 0 & h_{x_1\rightarrow x_2} & \dots & h_{x_1\rightarrow x_n} \\ h_{x_2\rightarrow x_1} & 0 & \dots & h_{x_2\rightarrow x_n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{x_n\rightarrow x_1} & h_{x_n\rightarrow x_2} & \dots & 0 \end{bmatrix} \tag{4.4}$$

The causality matrix is antisymmetrical, that is, $\mathbf{h} = -\mathbf{h}^T$ or for its elements the following equality holds: $h_{x_\mu \to x_\nu} = -h_{x_\nu \to x_\mu}$. The variables to be compared have to be pre-selected through a fault detection mechanism. Once the variables are selected, the causality matrix can be calculated automatically.

4.2 Reference Case Study

In this section a case study is introduced that has been chosen as a reference for developing causality methods. A strong emphasis of this work lies on the use of real process data. Due to the empirical nature of the methods there are no optimal parameters in the literature for the statistic methods of nearest neighbours and transfer entropy although some generic rules are given for simulated data such as the Rössler or Lorentz systems. An important contribution of this work is therefore to derive guidelines from application to real-life data from industrial processes. Criteria for selecting a good case study are given in Section 9.1.1. For the selected case study, a detailed process schematic exists locating all process variables. The measurements are complete and include setpoints and controller outputs. An onset of the disturbance is captured in the data. The process is well understood and a physical explanation is available. The root cause of the disturbance in the data had been identified prior to the analysis. But most importantly, two alternative hypotheses initially existed for the root cause that gave rise to the need of a causality measure to favour one or the other hypothesis. The causality measures defined in the next Chapters will be explained more visually by applying them to the reference case study. Examples in the following chapters and section that refer to this case study will be indicated in slanted font and the paragraph will be titled "Case Study".

The probably most famous reference case study of an industrial control problem is presently the Tennessee Eastman Problem (TEP) by Jim Downs and Ernie Vogel [23][1]. The TEP consists of a reactor and separator recycle arrangement and is fully described by a mathematical model. The simulation code features 21 common faults such as step changes or random variations of input parameters and stiction of control valves. The reason why the TEP was not chosen as a case study despite its fame is firstly, that simulated data is not rich enough to test data-driven methods, as mentioned earlier, and secondly, that the TEP has a process model and therefore is one case where model-based fault isolation would be more appropriate.

[1] More than 100 references citing the publication are listed in the Web of Knowledge, status February 2005.

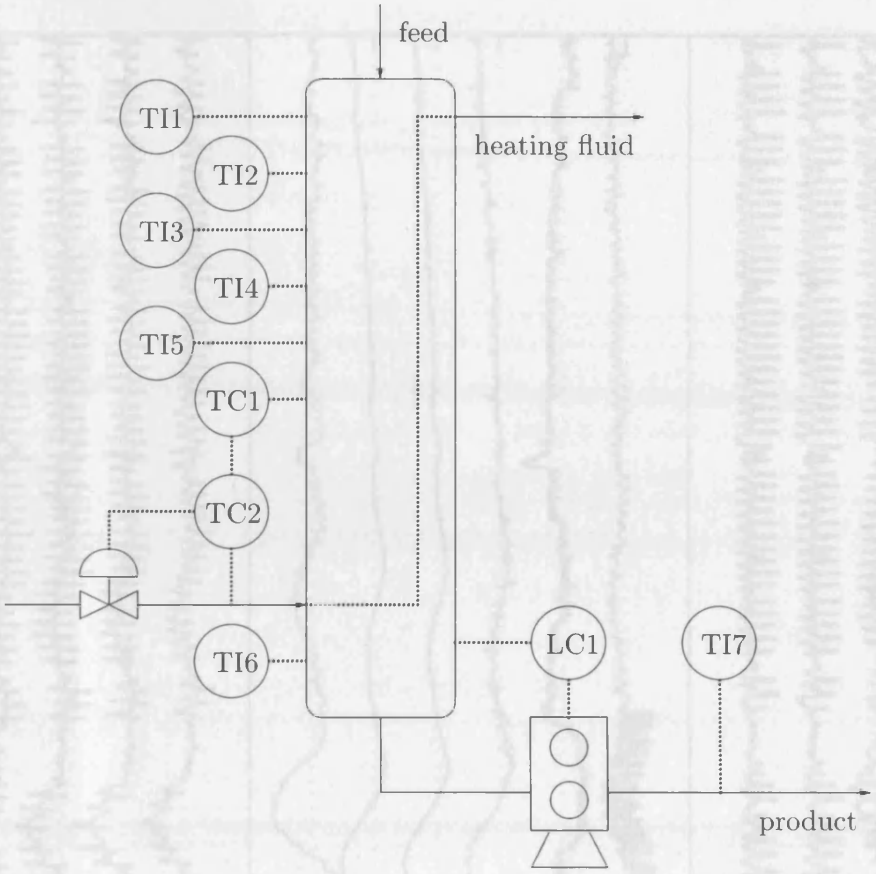


Figure 4.1: Process schematic of reference case study.

4.2.1 Process Description

The process investigated is a distillation and part of a larger plant. Figure 4.1 shows the process schematic. The column has a tray structure for distillation. A feed enters the top of the column and is separated into the desired product that exits the column at the bottom and a by-product that exits the column at the side draw not shown in Figure 4.1. A heating fluid is pumped through a piping system along the length of the column, without coming in contact with the product, and exits at the top. The heating fluid flow is controlled by the heating fluid temperature as the heating fluid is a shared facility with a varying temperature. Distillation is a well understood and most frequently employed process type and a column control setpoint as shown here is fairly common. The temperature in the column is controlled by a cascade loop for which the master controller (TC1) measures the temperature in the middle of the column and the slave controller (TC2) uses the temperature of the heating fluid. The flow out of the column is the manipulated variable for the bottom tray level (LC1) and is adjusted through a pump. Temperatures are measured along the upper part of the column (TI1 to TI5), at the bottom tray (TI6) and further downstream of the column (TI7).

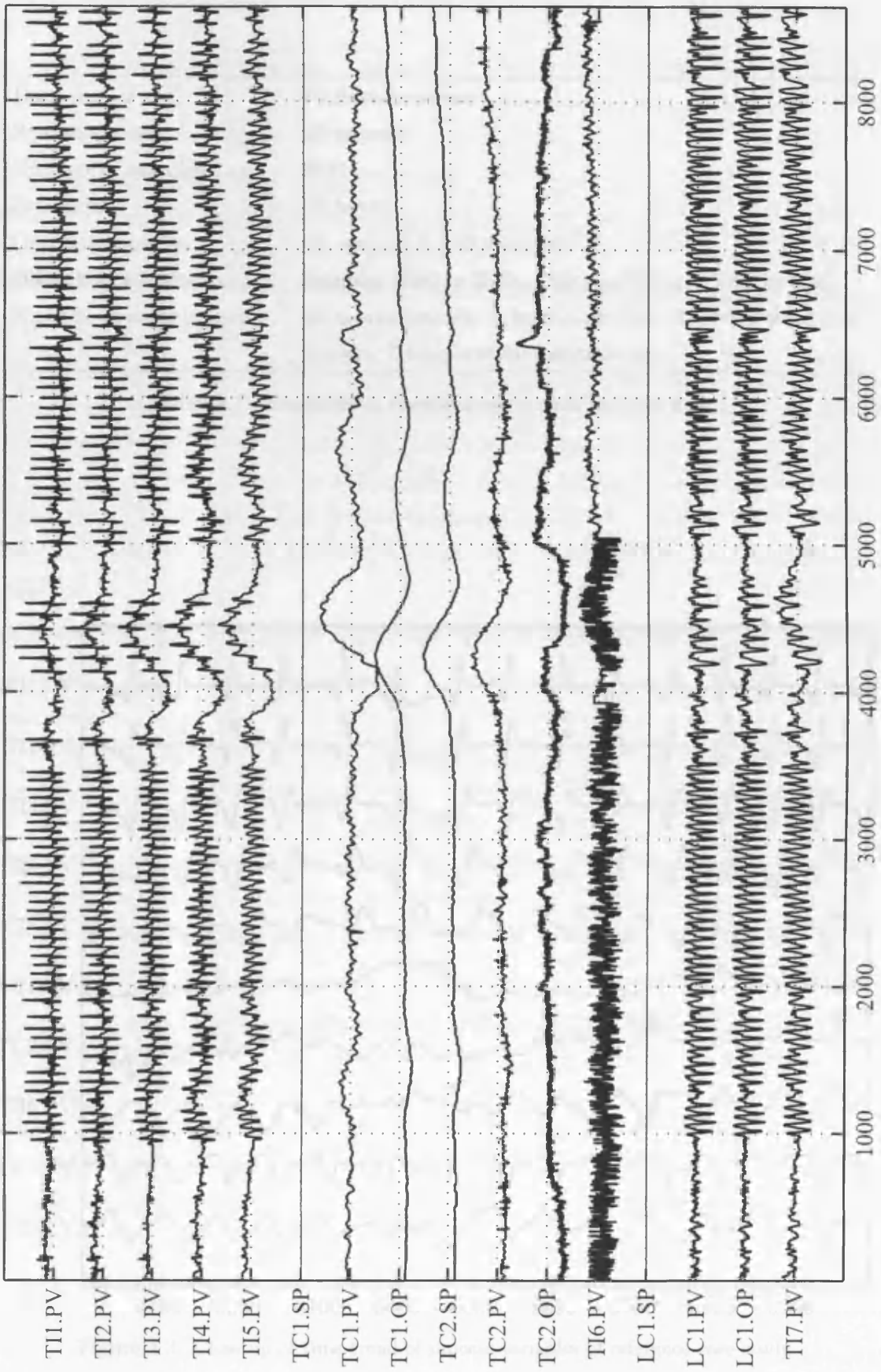


Figure 4.2: Time trend in samples for all measurements in reference case study.

Data source	PI data historian
Sampling rate	20 seconds
Number of samples	8641
Duration	48 hours
Oscillation period	61 samples ($\sim$ 20 minutes)
Disturbance intervals	Samples 1000 to 3500, 4200 to 4700 and 5000 to 8641
Number of measurements	10 measurements: 1 level controller, 2 temperature controllers, 7 temperature instruments

Table 4.1: Measurement specifications of reference case study.

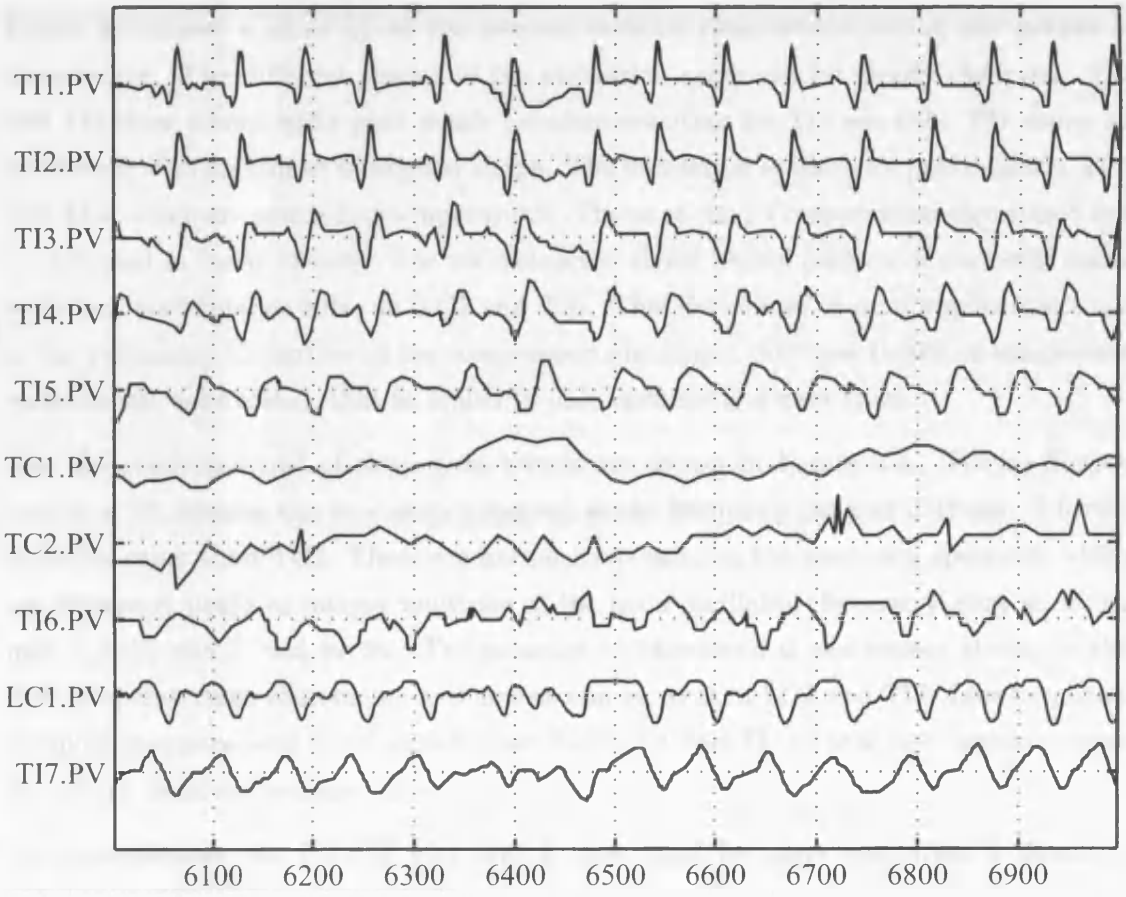


Figure 4.3: Close up of time trend of process variables of reference case study.

4.2.2 Process Disturbance

A periodic disturbance affected all variables in the process for most times with a few interruptions. The time trend of the process variables that was captured for analysis is shown in Figure 4.2. The disturbance most clearly affects TI1 to TI5, LC1 and TI7. A slow bump with high amplitude around sample 5000 disturbed the cascaded temperature controller TC1 and TC2. Since TC1 and TC2 are in cascade, the controller output of TC1 equals to the controller setpoint of TC2. No set point changes happened during the period of observation for TC1 and LC1. Temperature measurement TI6 was distorted until around sample 5000 with a high proportion of measurement noise.

The specifications of the data set captured for analysis are given in Table 4.1. The standard value of the sampling rate in the PI system is 20 seconds, that is, 3 samples per minute are captured. The period of the oscillation in Figure 4.2 is 61 samples and thus around 20 minutes. A problem with the mechanical equipment can therefore be excluded as the root cause since it usually results in an oscillation period in seconds or a few minutes. The data was captured for a period of two days which shows the persistency over a longer period.

Figure 4.3 shows a close up of the process variable time trends during the period of disturbance. The different shapes of the oscillation cycle can be clearly observed. TI1 and TI2 show a very spiky peak which becomes smoother for TI3 onwards. TI7 shows an oscillation with an almost triangular shape. The oscillation is the least prominent in TC1 and TC2 which are controlled temperatures. Traces of the PI compression algorithm[2] can be detected in many trends. The reconstructed signal shows periods of perfectly linear stretches, particulary visible in TC2 and TI6. This is because of inappropriate settings of the parameter CompDev of the compression algorithm. All time trends of the process variables are normalised, that is, scaled to unit variance and zero mean.

The frequency spectra[3] of these time trends are shown in Figure 4.4. The oscillation period of 20 minutes can be clearly observed at the frequency peak at 0.05 min^{-1} for all variables other than TC2. There are harmonics present in the frequency spectrum which are frequency peaks at integer multiples of the main oscillation frequency, that is, at 0.1 min^{-1}, 0.15 min^{-1} and so on. The presence of harmonics is particulary strong in the first temperature measurements and almost non existent in LC1 and TI7. Low frequency components are present in all signals apart from LC1 and TI7. These represent long term deviations from the average value.

For completeness, the PV/OP plot that is often used for valve evaluation is shown in

[2]For a description of the algorithm see Section 10.2.

[3]The frequency spectrum is often also referred to as "Power Spectrum" since it shows the power of a frequency bin compared to the total signal. Here, the term "Frequency spectrum" is preferred meaning the absolute value of the Fourier transform.

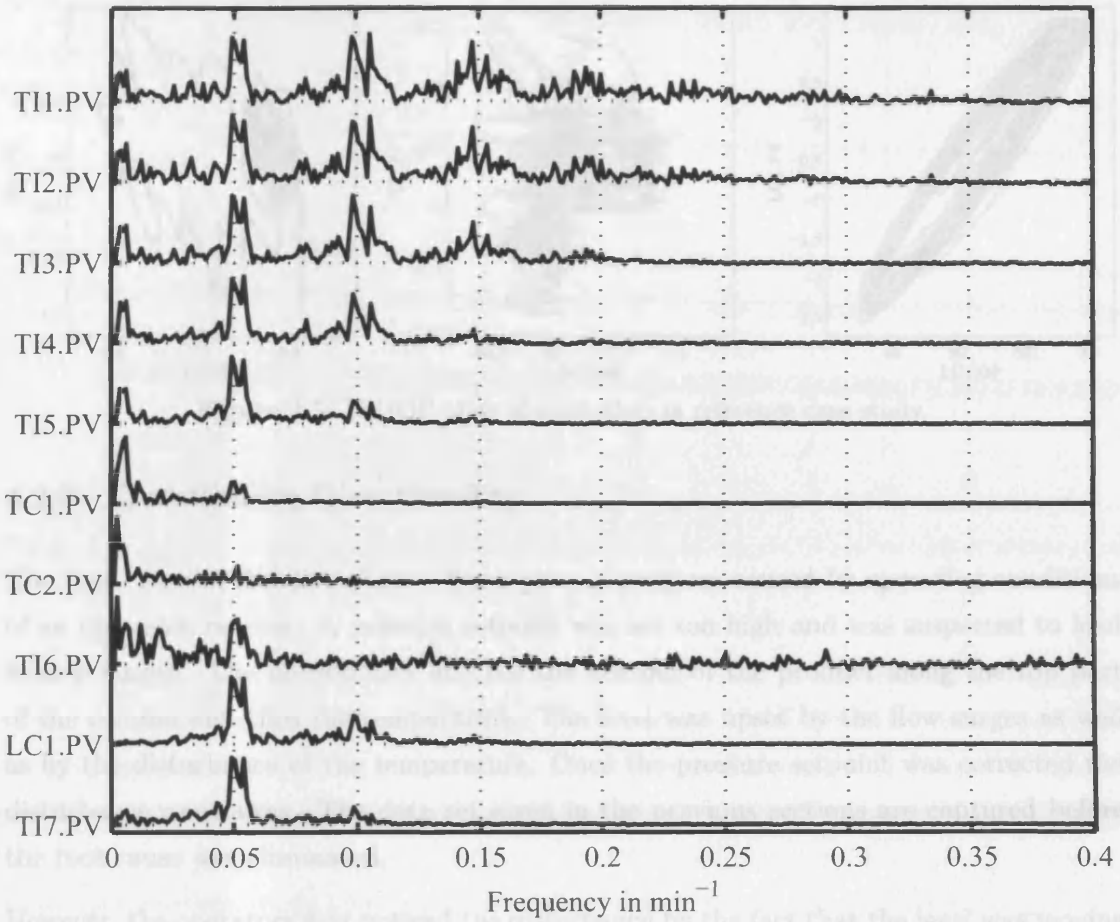


Figure 4.4: Frequency power spectra of process variables of reference case study.

Figure 4.5 for TC1, TC2 and LC1 in the interval between sample 1000 and 3500. The controller output of TC2 moves only in a small range between 46.5 and 48% while the output of LC1 ranges from 35 to 70% which is a rather large span. The process variable of LC1 also sticks to an upper limit which indicates also that LC1.PV is stuck at a value larger than the expected range. This can also be seen in the time trend in Figure 4.2. Physical explanations for the behaviour of LC1 is that either the level sensor is placed at a too low position or that the lowest tray is fitted with a weir that is overflowing when high value is reached. TC1 is a cascade controller and decides about the setpoint of TC2. Thus, the output is a temperature. A first glance assessment of the PV/OP plots does not come to a conclusion. No strong hysteresis can be observed in any of the valves.

Tag	TI1	TI2	TI3	TI4	TI5	TC1	TC2	TI6	**LC1**	TI7
$\frac{\sigma}{\mu}$[%]	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	**13.4**	0.2

Table 4.2: Standard deviation as percent of the mean of the process variables.

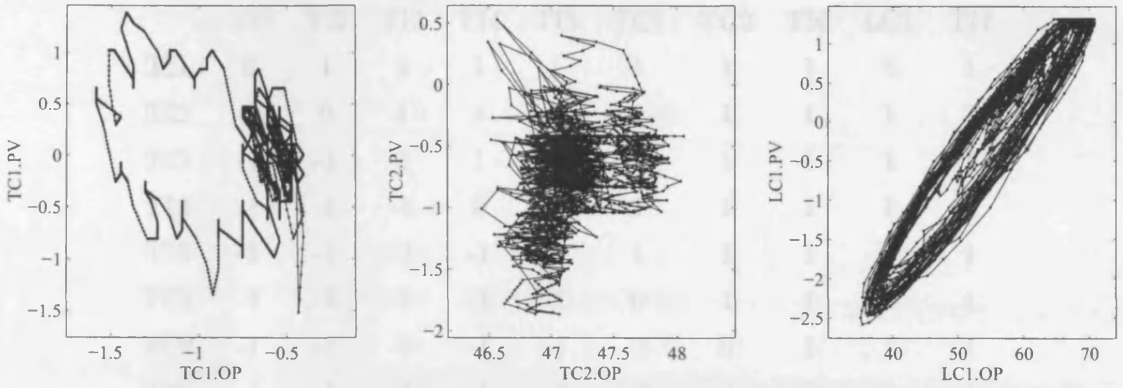


Figure 4.5: PV/OP plots of controllers in reference case study.

4.2.3 Question on Directionality

The root cause of the disturbance was a process problem caused by operating conditions of an upstream reactor. A pressure setpoint was set too high and was suspected to lead to flow surges. The uneven flow affected the heating of the product along the top part of the column and thus the temperature. The level was upset by the flow surges as well as by the disturbance of the temperature. Once the pressure setpoint was corrected the disturbance went away. The data set given in the previous sections are captured before the root cause was eliminated.

However, the operators first noticed the disturbance by the fact that the level was moving rather violently. Table 4.2 shows the standard deviation of the process variables as a percentage of the average value. All temperatures are actually oscillating with only 0.1 to 0.2% of the mean value while the variation of the level is 13.4%. A first conclusion therefore was that the level oscillates strongly and then affects the temperatures in the column and further downstream through its strong oscillation. A support of the theory would be that the temperatures TI6 and TI7 which are closest to LC1 vary more (0.2%) than the remaining temperatures in the column.

In this situation, a causality measure can decide whether LC1 is the root cause or if the disturbance enters the column from upstream. If LC1 influences TI6, TI6 influences TI5 etc, then LC1 would be favoured as the root cause. If TI1 influences TI2, TI2 influences TI3 and so on, this would surely be an indication that the disturbance propagates downstream in the plant rather than upstream originating from LC1.

4.3 Graphical Tools for Causal Information

This section outlines ways of representing and visualising dependencies between process variables. Data driven methods for fault detection and diagnosis evaluate a process variable or a group of process variables and usually result in a single index. Such a single index can be the power of the first principal component in principal component analysis, see

	TI1	TI2	TI3	TI4	TI5	TC1	TC2	TI6	LC1	TI7
TI1	0	1	1	1	1	1	1	1	1	1
TI2	-1	0	1	1	1	1	1	1	1	1
TI3	-1	-1	0	1	1	1	1	1	1	1
TI4	-1	-1	-1	0	1	1	1	1	1	1
TI5	-1	-1	-1	-1	0	1	1	1	1	1
TC1	-1	-1	-1	-1	-1	0	1	1	1	1
TC2	-1	-1	-1	-1	-1	-1	0	1	1	1
TI6	-1	-1	-1	-1	-1	-1	-1	0	1	1
LC1	-1	-1	-1	-1	-1	-1	-1	-1	0	1
TI7	-1	-1	-1	-1	-1	-1	-1	-1	-1	0

Table 4.3: Causality matrix of reference case study. Tags on the y-axis are the driver (cause) and tags on the x-axis are the response (effect).

MacGregor and Kourti [77], the power of a frequency peak or the oscillation regularity [136]. An index can be either displayed as a comparison of a number of process variables or for a single process variable over time. Classification techniques have been developed for the case of a number of indices per variable. For a textbook on pattern classification see Duda et al.[24]. As an example, Tan et al. [129] use hierarchical classification trees to map the first three component of the spectral principal component analysis onto a distance plot and thus enable a two dimensional image of the three dimensional data.

In causal analysis, one index is generated for every pairing of variables. The indices can well be displayed one by one for each combination of two time trends but to view the full information all relationships between the process variables should be compared. That is, for p process variables $p(p-1)$ relationships must be considered[4]. One way of representing the causal relationships is by visualising the causality matrix.

Case Study: *A theoretical causality matrix of the case study introduced in previous Section 4.2 is given in Table 4.3 for the case when a feed disturbance propagates. The matrix can be interpreted as the tags on the y-axis being the driver and the tags on the x-axis being the response. Here, the process variables or tags are already ordered in the sequence of their expected occurrence and along the process flow. The dependency will also follow along the process flow if the level is not the root cause. Thus, the elements above the main diagonal have the value of the ideal causality value '1' and the elements below the main diagonal the ideal causality value '-1'. For example, TI1 will then influence TI2 which is indicated by the value '1' in element (1,2) of the matrix. TI7 will influence no other variable and therefore no value '1' can be found in the last row. Furthermore, TI1 will also influence TI3 and then TI4 and so on. In this way, the upper half of the matrix is filled with '1' and the lower half with '-1'.*

[4]There is no such thing as the relationship between a process variable and itself.

	LC1	TC1	TC2	TI1	TI2	TI3	TI4	TI5	TI6	TI7
LC1	0	-1	-1	-1	-1	-1	-1	-1	-1	1
TC1	1	0	1	-1	-1	-1	-1	-1	1	1
TC2	1	-1	0	-1	-1	-1	-1	-1	1	1
TI1	1	1	1	0	1	1	1	1	1	1
TI2	1	1	1	-1	-1	1	1	1	1	1
TI3	1	1	1	-1	-1	0	1	1	1	1
TI4	1	1	1	-1	-1	-1	0	1	1	1
TI5	1	1	1	-1	-1	-1	-1	0	1	1
TI6	1	-1	-1	-1	-1	-1	-1	-1	0	1
TI7	-1	-1	-1	-1	-1	-1	-1	-1	-1	0

Table 4.4: Causality matrix of reference case study with tagnames in alphabetical order. Tags on the y-axis are the driver (cause) and tags on the x-axis are the response (effect).

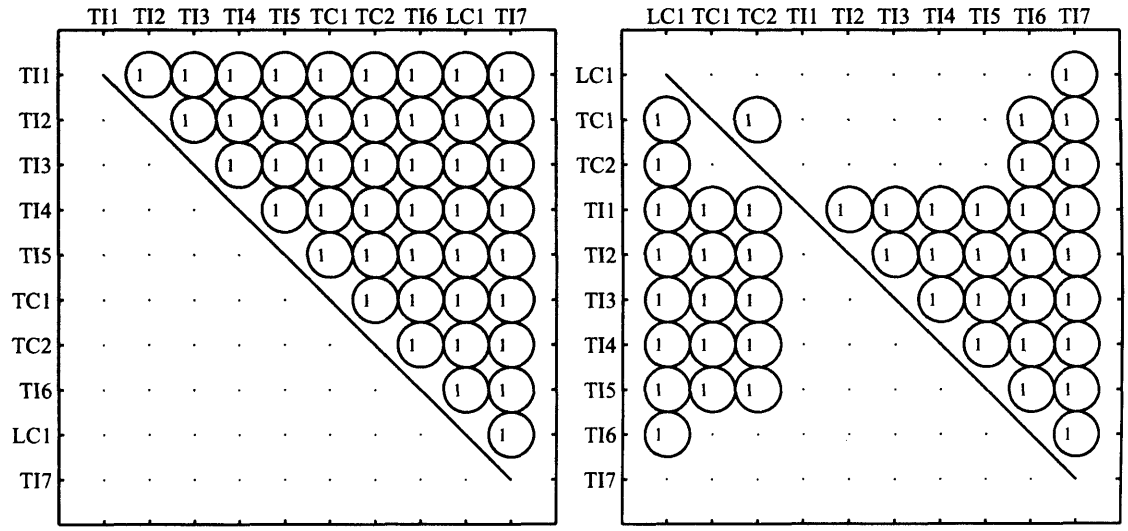


Figure 4.6: Bubble chart of causality matrix in Table 4.3 and of shuffled causality matrix of Table 4.4. Tags on the y-axis are the driver(cause) and tags on the x-axis are the response (effect).

In the following sections, a number of graphical representations are introduced. First, the causality matrix and their entries can be directly and conveniently represented by bubble charts. The directional information can then be displayed on a circular chart in a systematic way. Cause and effect for process monitoring and fault diagnosis is usually plotted in digraphs or causal maps which will be introduced in the last section.

4.3.1 Bubble Charts

Bubble charts are widely used in data presentation and are a feature of office software such as Microsoft Word or Excel[5]. Here, they are generated with Matlab. Bubble charts

[5]The Microsoft Excel description of a bubble chart reads: "Bubble. Compares sets of 3 values. Like a scatter chart with the third value displayed as the size of the bubble marker."

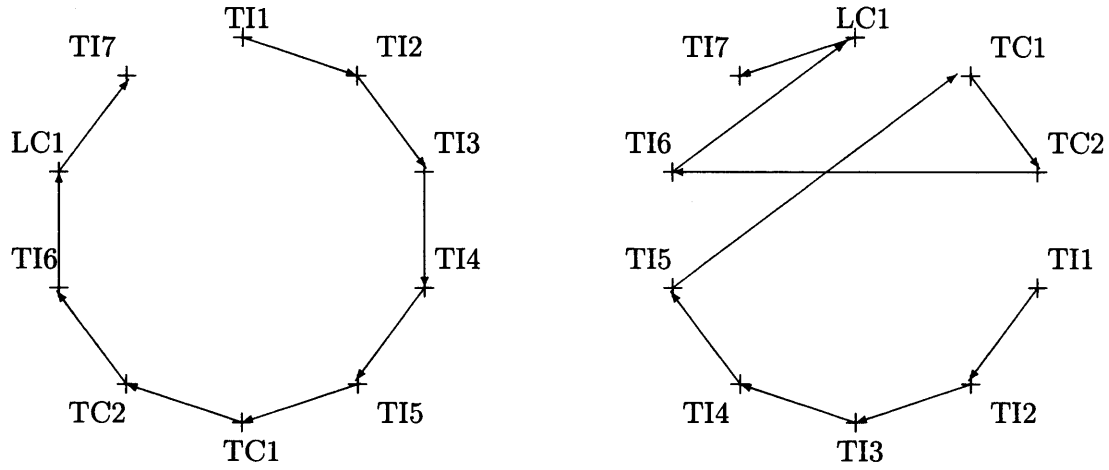


Figure 4.7: Examples of the circular chart. The left hand plot shows the chart in the original order while the right hand plot shows the case for the shuffled data.

are used to show matrix values that are a function of two inputs, for example, the cost of a company by business unit and product can be shown in one graph using the bubble chart[6]. The departments are listed on the y-axis while the products are listed on the x-axis. The size of the bubble at the intersection of a department and product reflects the cost that this department spends on a particular product.

In data-driven causality analysis, the bubble chart can be directly constructed from the causality matrix. The strength of interdependence between two variables is shown by the size of the bubble. Negative values are ignored but due to the anti-symmetry no information is lost.

Case Study: *The left hand plot of Figure 4.6 shows the corresponding bubble chart to the causality matrix in Table 4.3. The chart is rather simple due to the ideal assumptions (only values of ones and zeros are allowed.). The value of the causality measure is shown in the centre of the bubble. However, the bubble chart can be read at a first glance while the causality matrix has to be studied thoroughly for entries unequal to zero. The benefit of bubble charts becomes more obvious if the tags are not arranged in the order of process flow. Table 4.4 shows the causality matrix with the tags shuffled into alphabetical order[7]. It cannot be seen from a first glance which variable influences the other. Thus, an automatic algorithm would be of use to rearrange the variables like shown in the left hand plot of Figure 4.6 where the dependencies can be seen at first glance. An algorithm that achieves this task is introduced in Section 4.4.*

[6]The idea of using bubble chart came whilst being on an internship with The Boston Consulting Group (BCG) in February 2004, taking two months leave of absence of the Ph.D. course. BCG like other management consultancies uses bubble charts extensively in presentations clients.

[7]When capturing data sets the tags are often in alphabetical order due to a sorting mechanism of the data acquisition system.

4.3.2 Circular Directional Charts

An alternative display is a circular chart in which the directional information that is contained in the causality matrix can be visualised. Arranging the data in this way goes back to Florence Nightingale who was also a pioneer statistician and developed graphical displays of descriptive statistics[8]. For this chart, only the elements in the row above the main diagonal are considered. All p variables are arranged in a circle separated by $360/p$ degrees. The sign of the values of the causality matrix that are above a confidence level are translated into vectors pointing from the driver to the response variable. The left hand plot of Figure 4.7 shows the circular directional chart constructed from the causality matrix in Table 4.3. The fault propagation path can be easily followed from TI1 to TI2 to TI3 and so on. The strength of the dependency can be indicated by either attaching the values to the vectors or by changing the line width of the vectors. The former option is rather impractical in general as shown on the right hand graph of Figure 4.7. Here, the tags are again sorted alphabetically corresponding to the causality matrix Table 4.4. The fault propagation path can be followed but is less obvious. Circular directional charts as well as bubble charts can be generated in an automated way. The circular charts are not used in this work since they are more difficult to read than the causal map which will be introduced in the next section. The causal map is used for all case studies.

4.3.3 Digraphs and Causal Maps

When turning a causality measure into a graphical representation the desired output is a representation that is meaningful to the process engineer. That is, the graph should say: the flow out of a column affects the flow in the following tank that then subsequentially upsets the level. The root cause can be identified by finding the first variable in this chain of causal arguments. A graphical representation of cause and effect are digraphs which are also referred to as causal maps. A digraph is a graph with directed arcs between nodes. The nodes represent process variables and the arcs the relationship between the process variables. There can be several arcs leading to or from one node.

Digraphs can be constructed in a number of ways. If a mathematical model of the process exists in the form of algebraic and differential equations then these can be turned into digraphs as described in Section 3.4. A second method is to extract expert knowledge to construct a causal map from other sources describing the process such as the process schematic (see Section 2.2) or knowledge that the process engineer has. The approach followed in this work is to derive the causal information, and especially the causal information in case of a disturbance, from historical process data. The cause and effect relationships gained from the process data can be used even when no model exists and

[8]For information on the statistics of Florence Nightingale visit the Florence Nightingale Museum, 2 Lambeth Palace Road, SE1 London, UK, or at www.florence-nightingale.co.uk

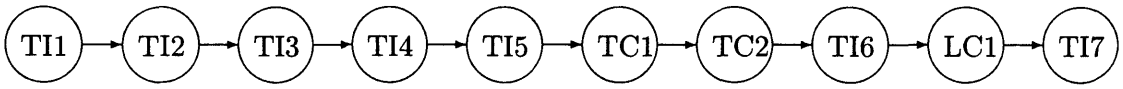


Figure 4.8: Causal map of reference case study described by causality matrices in Tables 4.3 and 4.4.

provide a valuable supplement to process knowledge that is either ambiguous or deviates from normal in case of a disturbance. The construction of digraphs from the causality matrix will be explained using the example in the previous Section 4.2. An overview of how digraphs can be retrieved from expert knowledge and differential equations is given in Section 3.4.

Case Study: *The signed digraph that can be retrieved from the causality matrix in Table 4.3 is shown in Figure 4.8. The chart is neither surprising nor very sophisticated since the causality matrix follows an idealised and simplified structure. However, it is important to remark that the digraph derived from the causality matrix is identical for both arrangements of the tags, that is, both the right hand and left hand panels of Figure 4.6 give the same digraph.*

4.4 Automatic Generation of Causal Maps

This section focuses on bubble plots and shows how a causal map can be automatically generated. For industrial applications the relationship between several measurements is considered, say p measurements x_ρ with $\rho = 1 \ldots p$, and a representation of the results in a causal map or digraph is desired. The causality measure of Equation 4.3 is computed for all combinations of variables x_ρ, that is, $p(p-1)/2$ relationships. The results can be denoted in a causality matrix as follows.

$$\mathbf{h} = \begin{bmatrix} 0 & h_{x_1 \to x_2} & \ldots & h_{x_1 \to x_p} \\ -h_{x_1 \to x_2} & 0 & \ldots & h_{x_2 \to x_p} \\ \vdots & \vdots & \ddots & \vdots \\ -h_{x_1 \to x_p} & -h_{x_2 \to x_p} & \ldots & 0 \end{bmatrix} \tag{4.5}$$

The rows represent cause variables while the columns represent the effect variables. The task is to generate a causal map that shows the relationships between all x_ρ and x_η in an automated way. Due to symmetry, negative causality measures are ignored and set to zero without loosing any information. All relationships with a causality measure larger than a defined threshold are accepted and treated equally, such that $\Delta = \Theta(\mathbf{h})$ where $\Theta(\cdot)$ is the step function. Matrix $\Delta \in \mathrm{R}^{p,p}$ has a maximum of $p(p-1)/2$ unity entries while the remaining elements are zero.

Using the step function Θ removes information about the size of values $h_{x_\rho \to x_\eta}$. As the decision of direction of propagation is binary, that is, does A influence B or vice versa, the

size of the value is less important if the significance level lies above the threshold. In case of conflict between two options of fault propagation paths as described in the following algorithm the option with the higher entropy $h_{x_\rho \to x_\eta}$ value is considered as the more likely option.

The automatic generation of a causal map from the causality matrix Δ is accomplished in two steps. First, the order of measurements x_ρ is rearranged to $\bar{\Delta}$ to bring a maximum number of entries above the main diagonal. In the second step, the causal map is constructed from $\bar{\Delta}$. The automatic construction of a causal map is shown here for a model example and later on in the paper for the test case study.

4.4.1 Algorithm to Rearrange the Order of Variables

The new order of variables that maximizes the number of non-zero entries above the main diagonal is found by the following procedure.

► **Step 1:** Initial sorting. The measurements x_ρ are sorted by the number of non-zero entries in the ρth row to create a new causality matrix Δ' whose first row has the least number of zero entries. The new indices are ρ'. If there is a tie then the number of non-zero entries in the ρ'th column is used as a tie break. In case of a further tie the order remains unchanged.

► **Step 2:** Establishing rules. The entries above the main diagonal of Δ' are translated into inequality rules to ensure that the element remains above the main diagonal when further sorting. For example, the element in the second row and third column of Δ' in Table 4.5 gives the inequality $x_3 < x_5$.

► **Step 3:** Shifting elements above diagonal. If a non-zero element is below the diagonal in the ρ'th row and η'th column then x_ρ and x_η can be exchange if no previously established rule is violated as a consequence of the exchange. If a rule is violated, alternative positions with row index smaller than ρ' for x_η or larger than η' for x_ρ can be explored in the same way. In case that no position change without a rule violation can be found, the option with the highest value of $h_{x_\rho \to x_\eta}$, that is, the summed value for all rules in question, is chosen. A new rule is established for the exchanged variables.

► **Repeat** Step 3 for all elements below the main diagonal. The resulting matrix is the optimized causality matrix $\bar{\Delta}$, the order of the rows is denoted by $\bar{\rho}$.

The algorithm is illustrated using a model example in Table 4.5. The top table shows the example causality matrix Δ for five process variables $x_1 \ldots x_5$. In the initial sorting step, the variables are rearranged with x_2 and x_3 both with two entries in each row and x_5 one entry. x_4 is excluded from further analysis and the modified causality matrix is shown in the middle table of Table 4.5. Rules for the four elements above the main diagonal are established in Step 2, that is, $x_2 < x_5$, $x_2 < x_1$, $x_3 < x_5$ and $x_5 < x_1$. The element below

$\Delta =$

	x_1	x_2	x_3	x_4	x_5
x_1	0	0	0	0	0
x_2	1	0	0	0	1
x_3	0	1	0	0	1
x_4	0	0	0	0	0
x_5	1	0	0	0	0

$\Delta' =$

	x_2	x_3	x_5	x_1
x_2	0	0	1	1
x_3	1	0	1	0
x_5	0	0	0	1
x_1	0	0	0	0

$\bar{\Delta} =$

	x_3	x_2	x_5	x_1
x_3	0	1	1	0
x_2	0	0	1	1
x_5	0	0	0	1
x_1	0	0	0	0

Table 4.5: Example for reordering of variables x_r for automated causal map generation.

the diagonal for x_3 and x_2 is dealt with by Step 3. x_3 and x_2 can be exchanged since none of the established rules is violated. Thus, the modified causality $\bar{\Delta}$ results as shown in the bottom table of Table 4.5.

4.4.2 Algorithm to Construct the Causal Map

Once the modified causality matrix $\bar{\Delta}$ is retrieved the construction of a causal map is straight forward using the following steps.

► **Step 1:** Initializing layout. To generate a basic structure $x_{\bar{\rho}}$ are placed in a row according to the new order $\bar{\rho}$. Variables that were excluded from analysis in Step 1 can be omitted since they have no causal connection with all the other variables.

► **Step 2:** Inserting arcs. The relationships for all non-zero entries in causality matrix $\bar{\Delta}$ can be represented by arcs pointing from a cause variable in the detected row to an effect variable in the corresponding column. These arcs are used to represents the information flow.

► **Step 3:** Removing shortcuts. In some situations, if a variable A causes a variable B and B in turn causes C, a further dependency between A and C can be detected. This dependency is only secondary and is not of primary interest. It can be easily ignored by removing all shortcut arcs, that is, all arcs from variable x_ρ to variable x_η for which two relationships via a third variable exist.

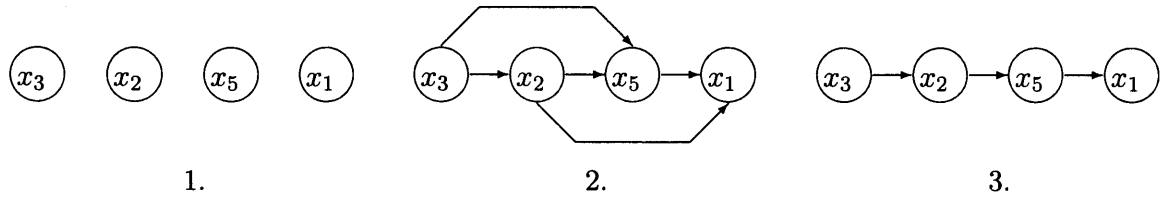


Figure 4.9: Construction steps of causal map for examples in Table 4.5. Step 1: initializing layout; Step 2: inserting arcs; Step 3: removing shortcuts.

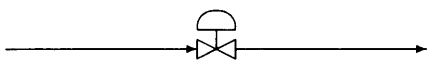
The variable on the left of the causal map is then the suspected root cause. The procedure is illustrated by the model example in Figure 4.9. In the left hand plot representing Step 1., all variables are laid out in the new row order $\bar{\rho}$. In Step 2., arcs for all relationships are drawn representing non-zero entries of causality matrix $\bar{\Delta}$ in Table 4.5. In the right hand plot of Figure 4.9, the third step is illustrated in which all shortcuts are removed. For example, the shortcut from x_3 to x_5 is removed since a way from x_3 to x_5 via x_2 exists. The resulting causal map represents the process dependency detected with the transfer entropy measure. x_3 is suspected to be the root cause of the investigated disturbance.

The main advantage of the algorithm is that it focuses on the order of occurrence of events which is desired for fault propagation and root cause analysis. The most important question of which variable comes first and is followed by which other variable is addressed. The situation of branches, that is, A influences B and C, can also be represented by the automatically generated causal map. Variables are set in the order A,B,C or A,C,B with arcs drawn from A to B and A to C.

Chapter 4 Summary

In this chapter, the nature of causality methods was described. The following items which will be used throughout the thesis were introduced:

- **A reference case study of an Eastman Chemical Company process for the development of the causality methods;**
- **A graphical representation, the bubble plot, that displays the pxp causality matrix;**
- **An algorithm for the automatic generation of causal maps that summarises the relationships between p variables from the causality matrix.**



Chapter 5

Cross-Correlation Function

In this chapter, a method based on cross-correlation is proposed to argue causal relationships from the presence of a time delay between two measurements. The concept of linear correlation is introduced and an algorithm proposed. The algorithm detects time delay and decides about its statistical relevance. The reference case study is used to test the method.

Correlation is a characteristic that establishes whether two variables are correlated in the most simplest functional sense, that is, a linear correlation. While being restricted to linear systems, correlation has the important benefit of being most tolerant to noise [86]. Cross-correlation is symmetric with respect to two variables x and y and thus cannot distinguish a symmetrical interaction from an asymmetric one [150]. The cross-correlation function can be twisted to an asymmetrical measure building on an observation made in Section 2.4.3, that is, when the fault or disturbance affecting a number of process variables is observed with a dead time or time lag in the variables. Correlation functions are suited to detect these time lags which can be used to retrace the fault propagation path.

In this chapter, the linear correlations measures of correlation coefficient, autocorrelation function and cross-correlation are explained and their use in fault diagnosis is reviewed. In Section 5.2, an algorithm using the cross-correlation as a causality measure is proposed. Guidelines for when to use the measure, its merits and limitations together with a significance level for its reliability are given. The method is illustrated using the example from Section 4.2. Correlation functions have found a number of applications in the area of fault diagnosis and monitoring. Applications that have been proved to be successful in the past will be introduced together with the definitions of the functions.

5.1 Introduction to Linear Correlation Measures

In this section, popular linear correlation tools are reviewed that describe the correlation either between two time series or between two time points of one series, namely the linear correlation coefficient, the autocorrelation function and the cross-correlation function. These methods assume that the investigated underlying mechanism is linear and fail if a nonlinear functionality is analysed. However, some of these tools, such as the autocorrelation function, are basic analysis particulary well suited for oscillatory signals. The application of the correlation tools for process monitoring purposes is also reviewed in this section.

5.1.1 Linear Correlation Coefficient

A measure for linear dependency of two discrete series is the linear empirical correlation coefficient which is also often called Pearson's correlation coefficient. The correlation coefficient between two time series x and y with N samples is defined as follows [10]:

$$r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}} \tag{5.1}$$

where $\overline{(\cdot)}$ indicates the linear mean:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{5.2}$$

The denominator comprises the standard deviation of the series x and y and therefore normalises the coefficient to values between $+1$ and -1 such that

$$-1 \leqq r_{xy} \leqq +1. \tag{5.3}$$

The correlation coefficient has a maximum value of ± 1 if the values x_i and y_i lie on a linear line. In this case the two series are called linearly dependent or correlated. Independency results in $r_{xy} = 0$.

The computation of r_{xy} as defined in 5.1 is only an estimate of the correlation coefficient. Due to the finite sample length, the coefficient estimate will only be approximately zero for two uncorrelated sequences. The variance of r_{xy} for two uncorrelated sequences can be estimated as

$$\sigma_r^2 = \frac{1}{N} \tag{5.4}$$

where N again is the number of samples in the sequences x and y and larger than 400-500 [97]. The derivation of the σ_r^2 estimate can be found in Appendix A.1. The restriction to linear dependencies becomes obvious. If x and y are related by any other dependency, lets say $y_i = x_i^2$, the correlation coefficient will not reflect the dependency. Instead, the correlation coefficient will be close to zero if x is normally distributed, that is, the two

sequences will appear uncorrelated. Also, if a time lag is inserted between x and y, for example $y_i = x_{i-1}$, the correlation coefficient will not detect any dependency.

5.1.2 Autocorrelation Function (ACF)

The autocorrelation function (ACF) describes the similarity of two time points of a signal separated by a time lag κ. The similarity of two separated time points is particulary useful when analysing oscillatory signals where time points separated by the oscillation period have a strong similarity or dependency. The ACF was originally designed as a deterministic description function of random sequences [35] but also reveals useful properties for deterministic signals, such as sinusoidal or other periodic signals. For discrete and finite time series the ACF is defined as follows:

$$\bar{\phi}_{xx}[\kappa] = \sum_{i=-N}^{+N} x_i \, x_{i-\kappa} \tag{5.5}$$

where κ is the discrete time delay. The bar indicates that this is only a preliminary estimate of the ACF. Normalisation has to be carried out. There are two ways of normalising the ACF, namely biased and unbiased normalisation. The biased normalisation divides the ACF by the number of samples,

$$\phi_{xx}[\kappa] = \frac{1}{N} \sum_{i=-N}^{+N} x_i \cdot x_{i-\kappa} \tag{5.6}$$

while the unbiased normalisation takes into account that for large values of κ only small numbers of samples are available to compute the estimates. Thus, the unbiased normalised ACF is written as

$$\phi_{xx}[\kappa] = \frac{1}{N - |\kappa|} \sum_{i=-N}^{+N} x_i \cdot x_{i-\kappa}. \tag{5.7}$$

In the following, the biased ACF is used since the unbiased ACF can have unwanted outliers in estimations using small samples.

One of the effects of the ACF when applying it to a periodic function is that it filters out any noise components and additionally smoothes the original signal in a fashion similar to a low pass filter. A number of characteristics follow from the definition of the ACF (see Girod et al. [35]), in particular, the symmetrical behaviour of the ACF, that is:

$$\phi_{xx}[\kappa] = \phi_{xx}[-\kappa]. \tag{5.8}$$

Transforming the definition of the ACF results in the largest value of the ACF which is for $\kappa = 0$, that is,

$$\varphi_{xx}[0] = \sigma_x^2 + \mu_x^2 \geqq \phi_{xx}[\kappa] \tag{5.9}$$

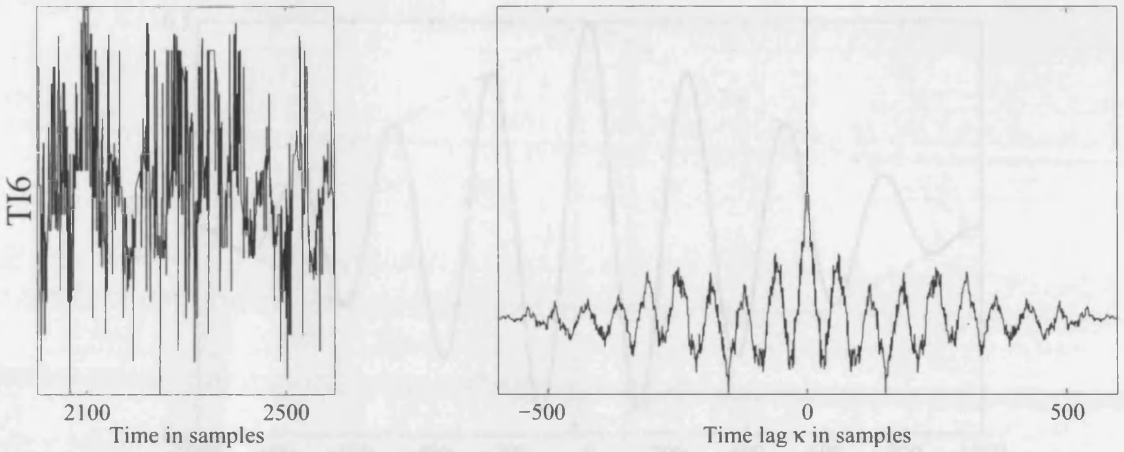


Figure 5.1: Time trend and corresponding autocorrelation function of TI6 from reference case study.

where σ_x^2 is the variance of x and μ_x is the expected value $\mathrm{E}\{x\} = \mu_x$ that can be approximated for stationary signals by the mean from Equation 5.2. Equation 5.9 implies that a large mean or variance will result in a high value for $\kappa = 0$. It is therefore recommended to subtract the mean before calculating the ACF.

Case Study: *The left hand plot of figure 5.1 shows the time trend of TI6 from the example in Section 4.2. The signal appears noisy and without any structure. Calculating the ACF as shown in the right hand plot of the same figure eliminates noisy components and reveals a periodic structure inherent in the signal. The high peak for $\kappa = 0$ reflects the variance that is both due to the oscillation and to the noise component. The linear mean was subtracted prior to the analysis.*

Wiener-Kinchin Theorem: The ACF is related to the power spectrum by the Wiener-Kinchin Theorem (Girod [35]): The Fourier transform of the autocorrelation function equals the power spectrum.

$$|X[2\pi f_n]| = \sum_{\kappa=-N}^{N} \phi_{xx}[\kappa] e^{-j\kappa 2\pi f_n} \tag{5.10}$$

where f_n is the discrete frequency. The Wiener-Kinchin theorem is particular important as it links the ACF to frequency analysis.

The ACF is often used for oscillation analysis since it compares the similarity of two points in time and also has some low pass filtering characteristics. As shown by the Wiener-Kinchin theorem it can therefore be used instead of frequency analysis which is often less intuitive and familiar to process engineers. Most process monitoring tools described in Section 3.5 have the facility of plotting the ACF overlayed in one graph. The period of oscillation can be compared by establishing the location of the maxima and taking the difference of the time parameter κ:

$$k_{osc} = \kappa_{max}^{(1)}, \tag{5.11}$$

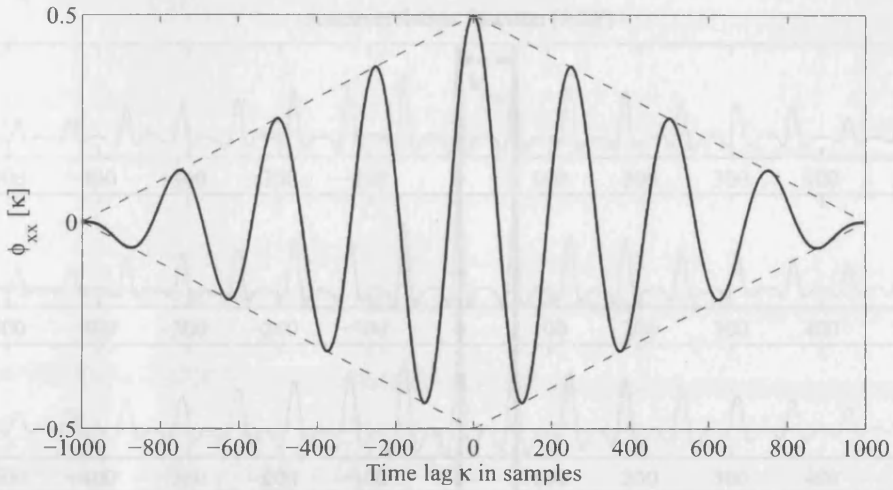


Figure 5.2: Autocorrelation function of sine wave with length of four oscillation cycles.

where $\kappa_{max}^{(1)}$ is the time lag of the first maximum after the absolute maximum at $\kappa = 0$.

The continuous ACF of a sine wave described by function $x(t) = \sin t$ defined over an interval T_1 can be expressed by the following equation:

$$\phi_{xx}(\tau) = \begin{cases} \frac{1}{2T_1}[(T_1 - \tau)\cos\tau - \sin(2T_1 - \tau) + \sin\tau] & 0 \leq \tau \leq T_1 \\ \phi_{xx}(-\tau) & -T_1 \leq \tau \leq 0 \\ 0 & \tau \text{ otherwise} \end{cases} \tag{5.12}$$

where T_1 is the duration of the sine wave and τ is the continuous time lag used similar to the discrete time lag κ. The derivation of this formula is shown in Appendix A.2. The discretised function of $\phi_{xx}(\tau)$ for four cycles is shown in Figure 5.2. The function is bounded by straight lines with a slope of $\pm\frac{1}{2N}$. This is due to the factor $\frac{(T_1-\tau)}{2T_1}$ acting on $\cos\tau$ and overshadowing the other contributions. Equation 5.12 is the exact ACF of a sine wave that is nonzero within an interval T_1. It resembles the unbiased ACF.

Case Study: *The autocorrelation functions of the first five signals in the example from Section 4.2 are shown in Figure 5.3. For displaying the ACF, the time trend was restricted to samples 2000 to 2600 because this was the time frame in which the oscillation was purest. The oscillation characteristic in the ACF is similar to the time trend in Figure 4.2 but smoothed. The largest values of the ACF can be observed at $\kappa = 0$. The two vertical bars indicate the maxima of the first oscillation for which the period of oscillation, k_{osc} can be measured. In this example, k_{osc} is 61 samples which, at a sampling rate of 20 seconds is equivalent to 20 minutes. Plotting the ACFs of the signals on top of each other enables an easy comparison of the oscillation period. In this case, the oscillation periods of TI1 to TI5 are identical. The shape of the oscillation, however, is not, which can be easily seen from the ACF better than from the frequency spectrum.* ———————

Further processing is applied by Thornhill et al. [136] to deduce the period of as well as the regularity of oscillation from the ACF in case of multiple oscillations present in the signal. The measure assesses the zero crossings of the ACF.

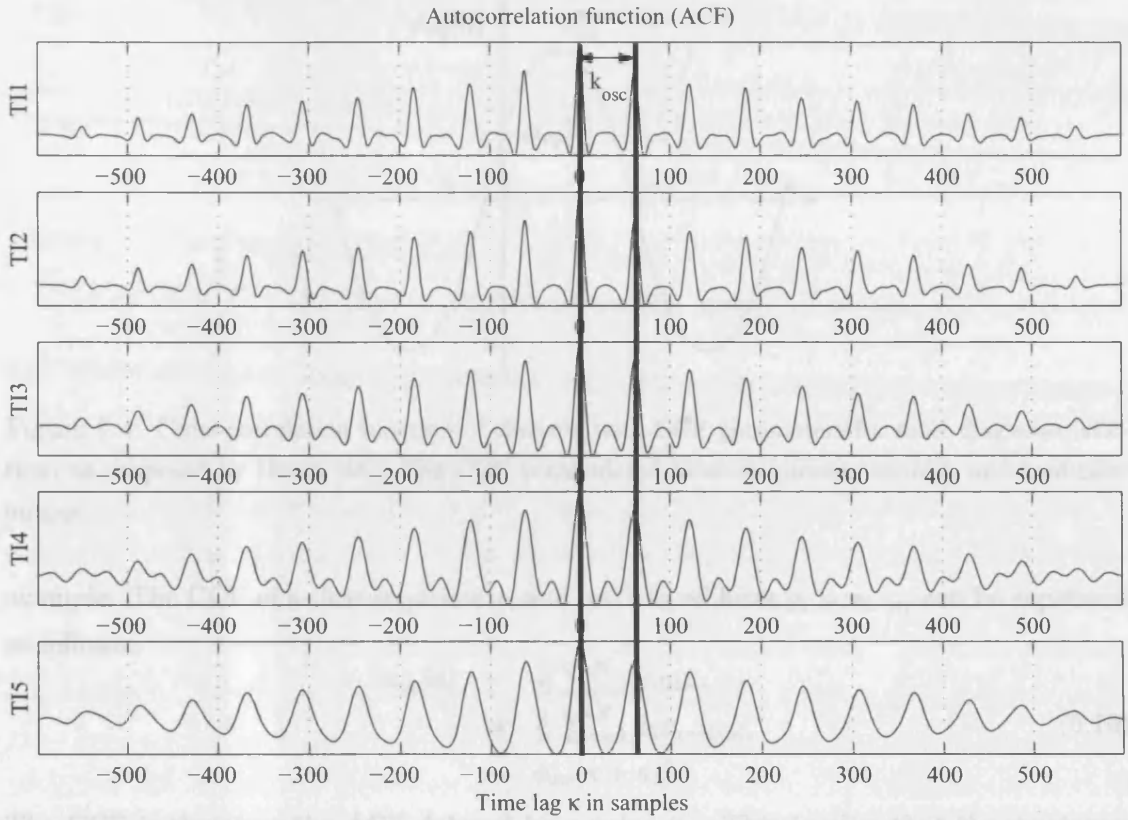


Figure 5.3: Autocorrelation function of time trends from reference case study.

5.1.3 Cross-correlation function (CCF)

In a fashion parallel to the ACF, the cross-correlation function (CCF) measures similarities of two time series delayed in time by a varying discrete time parameter κ.

$$\phi_{xy}[\kappa] = \frac{1}{N} \sum_{i=-N}^{+N} x_i \, y_{i-\kappa} \tag{5.13}$$

For zero mean signals, calculating the CCF is equivalent to the linear correlation coefficient if one of the signals is delayed by κ. The correlation coefficient is additionally standardised to unit variance, see Equation 5.1. Important characteristics of the CCF is that it is not symmetrical or commutative. However, the following relationship holds:

$$\phi_{xy}[\kappa] = \phi_{yx}[-\kappa]; \tag{5.14}$$

and two signals are called uncorrelated if the following equation holds:

$$\phi_{xy}[\kappa] = \mu_x \mu_y \tag{5.15}$$

for all values of κ. If x and y are two signals with the same oscillation frequency, then the CCF will reflect the oscillation but will also show the time lag with which the two oscillations occur. To show the measurement of the time trend consider the following

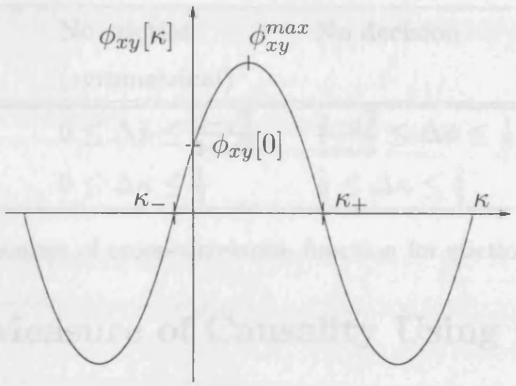


Figure 5.4: Cross-correlation function of sinusoid with CCF parameters for fault diagnosis (stiction) as proposed by Horch [45]. The CCF is calculated between process variable and controller output.

example. The CCF of a time sequence x_i and its delayed form $y_i = x_{i-\kappa_1}$ can be expressed as follows:

$$\begin{aligned} \phi_{xy}[\kappa] &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i y_{i-\kappa} \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i x_{i-\kappa-\kappa_1} \\ &= \phi_{xx}[\kappa+\kappa_1]. \end{aligned} \tag{5.16}$$

The CCF is therefore the ACF delayed by a value κ_1. This implies that the maximum value that was previously the value of the CCF at $\kappa = 0$ is now the value at $\kappa = -\kappa_1$. Finding the position of the maximum value is therefore a measure of the time lag. In other words, the time lag is retrieved for the delay κ at which the similarity between the two sequences, and thus the cross-correlation, is at its maximum. This is true for the case of a simple delay between two linearly correlated signals. The continuous cross-correlation function of a sine wave $x(t)$ and its delayed version $x(t-\tau_1)$ is derived in Appendix A.3 and results in:

$$\phi_{xy}(\tau) = \begin{cases} \frac{1}{2T_p}\left[(T_p-\tau)\cos(\tau+\tau_1) - \frac{1}{2}\sin(2T_p-\tau-\tau_1) + \frac{1}{2}\sin(\tau-\tau_1)\right] & 0 \le \tau \le T_p \\ \frac{1}{2T_p}\left[(T_p+\tau)\cos(\tau+\tau_1) - \frac{1}{2}\sin(2T_p+\tau-\tau_1) - \frac{1}{2}\sin(\tau+\tau_1)\right] & -T_p \le \tau \le 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{5.17}$$

CCF for Detection of Valve Stiction: Horch [45] observed that stiction in a control valve introduces a time delay of a quarter period, or $T_p/4$, between the process variable and the controller output of the affected loop whereas other oscillations result in an insignificant time delay between process variable and output. For time trends of a process variable and output of a loop with stiction see Figure 2.8. The parameters that are used to calculate the measure for stiction as developed in [45] are shown in Figure 5.4. Here, κ_+ is the zero crossing for positive lags, κ_- the zero crossing for negative lags and ϕ_{xy}^{max} the maximum value. The measures $\Delta\phi$ and $\Delta\kappa$ together with their threshold values for stiction and non stiction are given in Table 5.1. $\Delta\kappa$ and $\Delta\phi$ are both measures of the symmetry; if they are zero the CCF is symmetric.

Definition	No stiction (symmetrical)	No decision	Stiction (asymmetrical)
$\Delta\phi = \frac{\lvert\phi_{xy}[0]-\phi_{xy}^{max}\rvert}{\lvert\phi_{xy}[0]+\phi_{xy}^{max}\rvert}$	$0 \leq \Delta\phi \leq \frac{2-\sqrt{3}}{2+\sqrt{3}}$	$\frac{2-\sqrt{3}}{2+\sqrt{3}} \leq \Delta\phi \leq \frac{1}{3}$	$\frac{1}{3} \leq \Delta\phi \leq 1$
$\Delta\kappa = \frac{\lvert\kappa_- - \kappa_+\rvert}{\kappa_- + \kappa_+}$	$0 \leq \Delta\kappa \leq \frac{1}{3}$	$\frac{1}{3} \leq \Delta\kappa \leq \frac{2}{3}$	$\frac{2}{3} \leq \Delta\kappa \leq 1$

Table 5.1: Symmetry measures of cross-correlation function for stiction detection by Horch [45].

5.2 A Simple Measure of Causality Using CCF

In this section, an algorithm for detecting cause and effect between two process variables is introduced using the cross-correlation function. The underlying principle of the method is that when the disturbance propagates through the plant the disturbance can often be observed at a number of process variables with a time lag, see Section 2.4.3. The knowledge of the exact time lag hints towards the root cause by arguing that the variable closer to the root cause will show the disturbance before a variable further away. The CCF measures the similarity between signals at different time points and is therefore ideal to measure the time lag if a number of conditions are fulfilled. These limitations are given in the following after first introducing the algorithm and establishing guidelines and a significance level for the method.

5.2.1 Proposed Algorithm

In the proposed algorithm, a time lag between two variables is found by looking at the maximum and minimum value of the cross-correlation function of the two variables. If the two sequences are similar enough but delayed by a time delay κ_1 then Equation 5.16 can be used that is, $\phi_{xy}[\kappa] = \phi_{xx}[\kappa + \kappa_1]$. The maximum value ϕ_{xy} without a delay will be close to 0 while a time lag κ_1 will shift the maximum value to $\kappa^{\max} = \kappa_1$. Thus, the time delay is retrieved by finding the maximum value $\phi^{\max}$ and is then the corresponding time lag $\kappa^{\max}$. It is also possible that the cross-correlation has a minimum that is larger than the maximum. In this case, the time index κ of the minimum is the time delay because one of the time series was inverted. For the algorithm, the minimum and maximum as well as the corresponding time indices are noted:

$$\begin{aligned} \phi^{\max} &= \max\{\phi_{xy}[\kappa]\}; \quad \text{with } \phi^{\max} > 0 \\ \phi^{\min} &= \min\{\phi_{xy}[\kappa]\}; \quad \text{with } \phi^{\min} < 0 \\ \kappa^{\max} &= \{\kappa | \phi_{xy}[\kappa] = \phi^{\max}\} \\ \kappa^{\min} &= \{\kappa | \phi_{xy}[\kappa] = \phi^{\min}\} \end{aligned} \tag{5.18}$$

The maximum of the cross-correlation function will always be positive while the minimum will always be negative [35].

The detected time delay depends then on the ratio of $\phi^{\max}$ to $\phi^{\min}$. If $\phi^{\max}$ is larger than $\phi^{\min}$ then $\kappa^{\max}$ is the detected time delay. The detected time delay is temporarily referred

to as λ. The time delay has to pass further tests to be detected.

$$\lambda = \begin{cases} \kappa^{\max}, & \phi^{\max} + \phi^{\min} \geq 0 \\ \kappa^{\min}, & \phi^{\max} + \phi^{\min} < 0 \end{cases} \tag{5.19}$$

Also, a certain similarity between must exist between the two time series. The linear correlation coefficient is therefore calculated for the time series where one is delayed by $\kappa^{\max}$. The difference between taking the linear correlation coefficient and the CCF value for $\kappa^{\max}$ is that the correlation coefficient can take the reduced number of samples, $N - \kappa^{\max}$ instead of N, into account. Furthermore, the correlation coefficient will have extremum values of ± 1 for identical time series, unlike the CCF whose value for identical time series will depend on the nature of the series. The linear correlation coefficient is then defined by

$$r^{\max} = r_{xy_\lambda} \quad \text{where } y_\lambda[k] = y[k + \lambda] \tag{5.20}$$

In the case that the two signals x and y are of oscillatory nature, a further test has to be carried out. The reason is that if the signals are time-reversal periodic[1] with an oscillation period T_{osc}, then both λ and $\lambda - T_{osc}/2$ can be viewed as the time delay. If λ is positive and detects a maximum, then the first time delay λ detected by the maximum would mean that x influences y while the second time delay $\lambda - T_{osc}/2$ detected by the minimum would mean that y influences x. This is clearly contradictory. A further test therefore involves the ratio of the minimum and maximum. A comparative measure is given by

$$\psi = \frac{|\phi^{\max} + \phi^{\min}|}{\frac{1}{2}(\phi^{\max} + |\phi^{\min}|)} \tag{5.21}$$

If $\phi^{\max}$ and $\phi^{\min}$ are similar, then ψ is close to zero; if the difference between the two is large then ψ is significantly larger than zero. The largest value that ψ can adopt is if either $\phi^{\max}$ or $|\phi^{\min}|$ equals to one and the other to zero. In this case, $\psi = 2$. Thus, $0 \leq \psi \leq 2$.

The proposed algorithm can be summarised by the following steps:

1. Compute the cross-correlation function $\phi_{xy}[\kappa]$ of the two variables x and y;
2. Find the minimum and maximum $\phi^{\max}$ and $\phi^{\min}$ as well as the corresponding time lags $\kappa^{\max}$ and $\kappa^{\min}$;
3. Set the detected time lag to $\lambda = \kappa^{\max}$ if $\phi^{\max} + \phi^{\min} \geq 0$ else to $\lambda = \kappa^{\min}$;
4. Calculate the correlation index $r^{\max} = r_{xy^\lambda}$ using $y_i^\lambda = y_{i+\lambda}$ and check if $r^{\max} > r_{\text{thresh}}$;
5. Calculate oscillation index $\psi = \frac{|\phi^{\max}+\phi^{\min}|}{\frac{1}{2}(\phi^{\max}+|\phi^{\min}|)}$ and check if $\psi > \psi_{\text{thresh}}$

The values for thresholds r_{thresh} and ψ_{thresh} will be described in the following section.

[1] Time reversal periodicity means that the oscillation is symmetrical with respect to time.

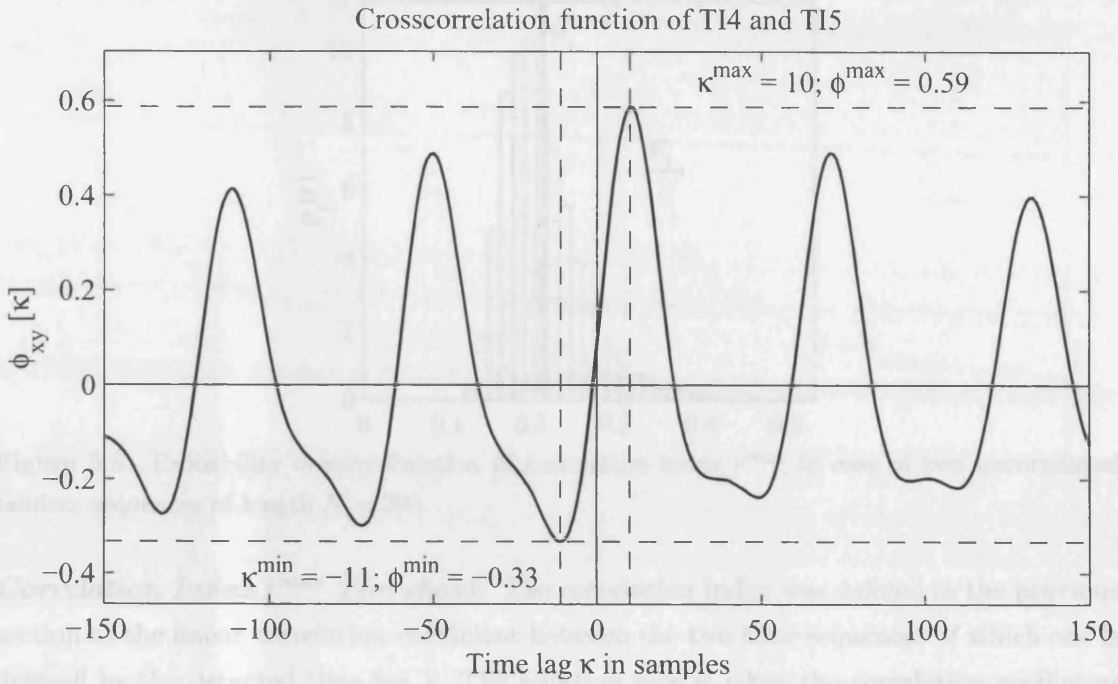


Figure 5.5: Time lag of maximum value of CCF for TI4 and TI5 from reference case study. A time lag of 10 samples was detected between TI4 and TI5.

Case Study: *Figure 5.5 shows a detail of the CCF of the two variables TI4 and TI5 from the reference case study in Section 4.2. The focus shows the CCF around the zero time lag* $\kappa = 0$. *The position of the time lag of the absolute maximum and minimum are indicated by dashed lines. The minima occurs at* $\kappa^{min} = -11$ *and has a value of* $\phi^{min} = -0.33$. *This means that if the minima would be the time difference between TI4 and TI5, then TI5 would happen 11 samples before TI4. The maxima, on the other hand occurs at* $\kappa^{max} = +10$ *and has a value of* $\phi^{max} = 0.59$. *If this would be the time difference then TI4 would happen 10 samples before TI5. Because these are two contradictory results the coefficient* ψ *is calculated to check if the maximum is significantly larger than the minimum:* $\psi = \frac{|0.59-0.33|}{\frac{1}{2}(0.59+|-0.33|)} = 0.57$. *A significance level of* ψ *to check if this is large enough will be given in the next section. The correlation index* r^{max} *for the time lag* $\lambda = +10$ *has a value of 0.99.*

5.2.2 Significance Level and Thresholds

In this section, the values of the thresholds r_{thresh} and ψ_{thresh} that are required for the causality algorithm described in the previous section are derived. If no threshold is set for $r^{\max}$ then a time delay is detected for all time trends since all cross-correlation function have a maximum or minimum value at some point. It is therefore necessary to introduce the lower boundaries for r and ψ. Threshold r_{thresh} can also be regarded as a significance level of the detected time delay λ if the criteria of $\psi > \psi_{\text{thresh}}$ is fulfilled.

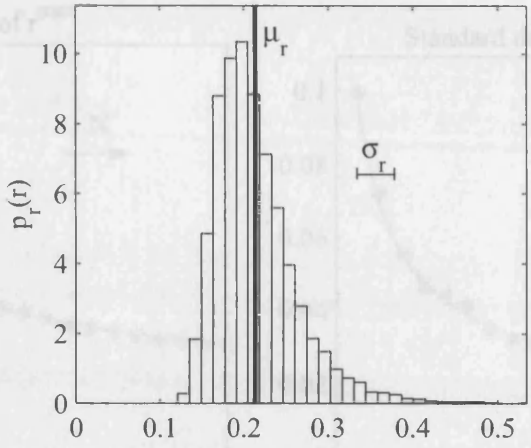


Figure 5.6: Probability density function of correlation index $r^{\max}$ in case of two uncorrelated random sequences of length $N = 200$.

Correlation Index r^{max} Threshold: The correlation index was defined in the previous section as the linear correlation coefficient between the two time sequences of which one is delayed by the detected time lag λ. The question now is when the correlation coefficient $r^{\max}$ is significant. Equation 5.4 in Section 5.1.1 gives the variance of the linear correlation coefficient of two uncorrelated signals. The variance is inversely proportional to the number of samples of the sequences from which it is estimated. Estimating the variance of $r^{\max}$ in this way assumes that the expected value of $r^{\max}$ is zero. This however is not the case since $r^{\max}$ is chosen at the maximum value of the cross-correlation function.

Deriving the actual variance of $r^{\max}$ analytically is a complex task since the maximum value has to be considered. Instead, it will be derived empirically here. Figure 5.6 shows the probability density function (PDF) of the correlation index $r^{\max}$ for two uncorrelated random sequences of length $N = 200$. The index was computed for 5000 uncorrelated sequences and the PDF was estimated using the histogram method[2]. It can be clearly seen that the mean, indicated by a vertical bar in Figure 5.6 is not zero but lies around 0.2. The standard deviation of $r^{\max}$ for uncorrelated data sequences of length $N = 200$ is approximately 0.04.

If $r^{\max}$ now lies outside the main distribution as given in Figure 5.6, say above 0.4, it can be said with a small remaining uncertainty that the two sequences are correlated. For this reason, a three sigma test is used such that

$$r_{\text{thresh}} = \mu_{r^{\max}} + 3\,\sigma_{r^{\max}}, \tag{5.22}$$

where $\mu_{r^{\max}}$ is the mean and $\sigma_{r^{\max}}$ is the standard deviation of the distribution. The PDF shown in Figure 5.6 is only valid for the case of $N = 200$.

To derive a functional relationship between mean, variance and the number of samples N investigated, the dependency is empirically derived. Therefore, the correlation index

[2]The concept of the probability density function and the estimation using histograms is described in more detail in Chapter 7

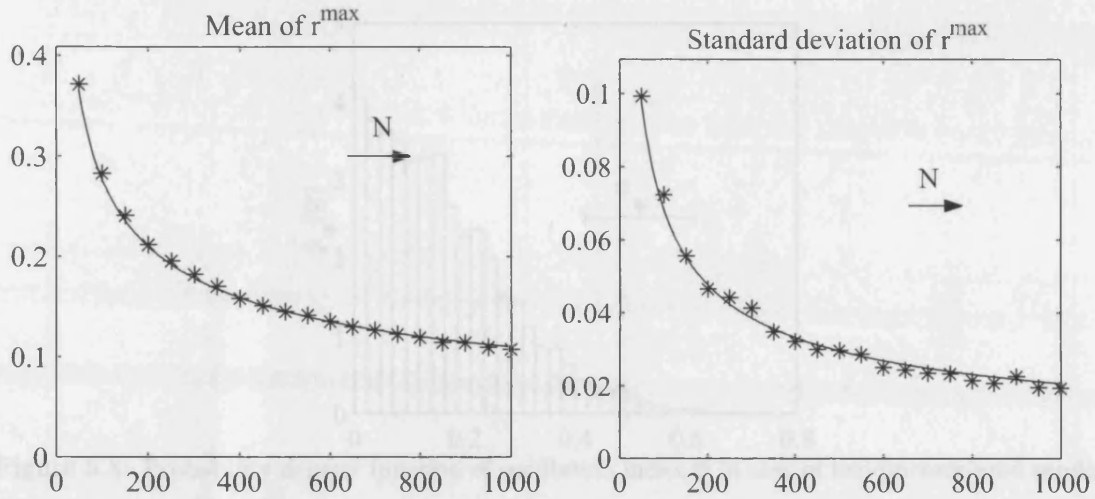


Figure 5.7: Mean (left panel) and standard deviation (right panel) of correlation index as a function of sample length N. The solid line indicates the approximation function as given in 5.23 and 5.24 while the stars give the experimentally measured values.

$r^{\max}$ is computed 1000 times for a number of values of N. Mean and standard deviation are estimated from this sample set. Figure 5.7 shows the resulting functional dependency. Both mean and standard deviation decrease monotonically with increasing number of samples. Curves can be fitted to the sample values. Since the expected standard deviation was a polynomial function of N, the same is assumed here. The following function can be fitted to match the observed dependency between the mean of $r^{\max}$ and the number of samples N:

$$\mu_{r^{\max}}(N) = 1.75\, N^{-0.4} \tag{5.23}$$

while for the standard deviation of $r^{\max}$ the following equation can be fitted.

$$\sigma_{r^{\max}}(N) = 0.75\, N^{-0.52} \tag{5.24}$$

Thus, the threshold r_{thresh} results by substituting the fitted functions 5.23 and 5.24 in Equation 5.22 such that

$$r_{\text{thresh}}(N) = 1.75\, N^{-0.4} + 2.25\, N^{-0.52}. \tag{5.25}$$

The resulting value threshold for the threshold will be applied to the case study later on in this section.

Oscillation Index ψ Threshold: The same approach as for the correlation index is pursued for the oscillation index ψ. Again, the expected value and the variance of ψ are difficult to derive analytically and are therefore deduced empirically. The oscillation index is first computed for 5000 random sequences of sample length $N = 200$. Figure 5.8 shows the resulting probability density function (PDF) that has been estimated using histograms. The shape of the PDF is different from the shape of the PDF of the correlation index in Figure 5.6. However, if the oscillation index is above a certain threshold, say 0.6, it is unlikely that it originates from a random sequence. For calculating the standard deviation

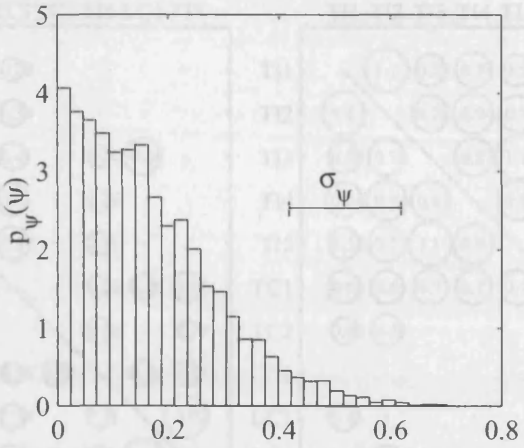


Figure 5.8: Probability density function of oscillation index ψ in case of two uncorrelated random sequences.

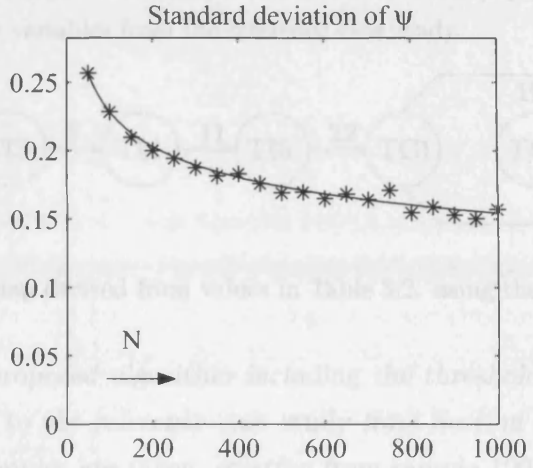


Figure 5.9: Standard deviation of oscillation index as a function of sample length N. The solid line indicates the approximation function as given in 5.27 while the stars give the experimentally measured values.

of the distribution, the PDF is mirrored on the y-axis. The standard deviation is, in the case of 200 samples, around 0.2, as indicated in Figure 5.8. The threshold value for ψ_{thresh} is then again defined as a three sigma test such that

$$\psi_{\text{thresh}} = 3\,\sigma_\psi. \tag{5.26}$$

For the derivation of the functional dependency between the standard deviation of ψ and the number of samples, the oscillation index ψ is computed 1000 times for different values of sample length N. The standard deviation from the sample set is shown in Figure 5.9. Again, the standard deviation decreases with increasing sample length.

The following curve can be fitted to the empirical values and is also indicated in Figure 5.9:

$$\sigma_\psi(N) = 0.5\,N^{-0.17}. \tag{5.27}$$

The threshold value of the oscillation index is then defined as

$$\psi_{\text{thresh}}(N) = 1.5\,N^{-0.17}. \tag{5.28}$$

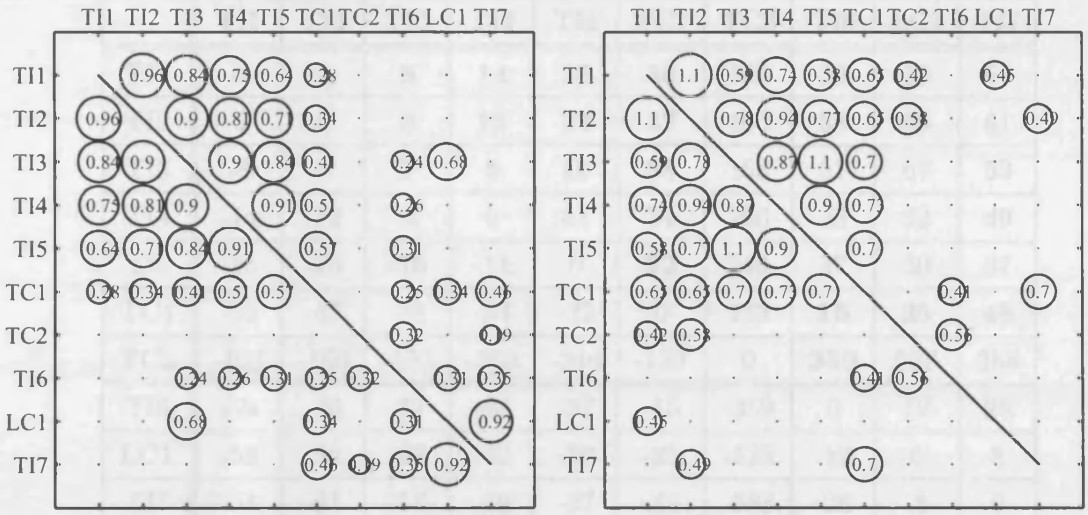


Figure 5.10: Correlation index $r^{\max}$ (left) and oscillation index ψ (right) above the thresholds r_{thresh} and ψ_{thresh} for the variables from the reference case study.

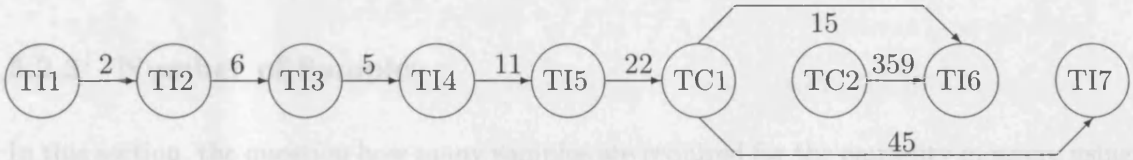


Figure 5.11: Causal map derived from values in Table 5.2. using the automated algorithm.

Case Study: *The proposed algorithm including the thresholds values derived in this section is now applied to the reference case study from Section 4.2. For the analysis, a total length of 2500 samples are taken, starting from sample 1001 to sample 3500 during which the oscillation is clearly present. This interval is selected since the disturbance is uninterrupted during that time period. Correlation index $r^{\max}$ and oscillation index ψ that exceed the thresholds are shown for all combinations of the variables in Figure 5.10. A cluster for the variables TI1 to TI5 and TC1 can be observed in both panels. The corresponding detected time delays λ are given in Table 5.2. It can be seen that all positive time delays lie above the main diagonal. In addition, the time between events can be seen from the detected time delay. For example, if TI1 occurs first, then after 2 samples TI2, after 8 samples TI3 and so on. Thus, the cause and effect relationship is identified correctly as expected from Figure 4.6 and argued in Section 4.2. However, only the values of λ for which the correlation and oscillation indices exceed both thresholds r_{thresh} and ψ_{thresh} are considered. These are highlighted in bold in Table 5.2. A causal map is derived using the algorithm from Section 4.4. The resulting causal map is shown in Figure 5.11.*

	TI1	TI2	TI3	TI4	TI5	TC1	TC2	TI6	LC1	TI7
TI1	0	**2**	**8**	**14**	**25**	**50**	163	39	58	64
TI2	-2	0	**6**	**12**	**23**	**47**	160	38	56	61
TI3	-8	-6	0	**5**	**16**	**38**	153	51	67	53
TI4	-14	-12	-5	0	**11**	**34**	266	48	42	49
TI5	-25	-23	-16	-11	0	**22**	246	37	30	37
TC1	-50	-47	-38	-34	-22	0	153	**15**	35	**45**
TC2	-163	-160	-153	-266	-246	-153	0	**359**	528	388
TI6	-39	-38	-51	-48	-37	-15	-359	0	19	28
LC1	-58	-56	-67	-42	-30	-25	-528	-19	0	8
TI7	-64	-61	-53	-49	-37	-45	-388	-28	-8	0

Table 5.2: Detected time delay results of the cross-correlation function algorithm of time trends from reference case study.

5.2.3 Number of Samples

In this section, the question how many samples are required for the causality measure using the cross-correlation is addressed. For all signals, the resolution of time lags depends on the sampling rate. If, for instance, only zero time lags are detected it means the sampling rate is not high enough because the actual time lag cannot be captured. For an oscillatory signal, the required number of samples N_{min} should depend on the number of samples per cycle as well as the number of cycles investigated.

Here, the number of samples will be investigated for a signal from the reference case study investigated earlier in Section 5.2.1. The cross-correlation algorithm is applied to time trends TI4 and TI5. The number of samples used for the analysis is varied from 2 to 1500 by looking only at subsets of the original time trends. Figure 5.12 shows the detected time delay λ, the correlation index r^{max} and the oscillation index ψ as functions of N. In the upper plot, it can be seen that time delay λ is correctly detected for a subset of around $N = 50$ samples or larger. The value fluctuates between 9 and 11 samples for a while which might both be correct since the actual time delay might be a value between 9 and 11. The correlation index r^{max} reaches a value close to one after approximately 40 samples. The oscillation index ψ, however, requires a much larger number of samples, around 300, to be larger than the threshold value ψ_{thresh}. It requires observation time to establish whether the detected time delay accredited to the oscillation or to the actual dead time between the two signals.

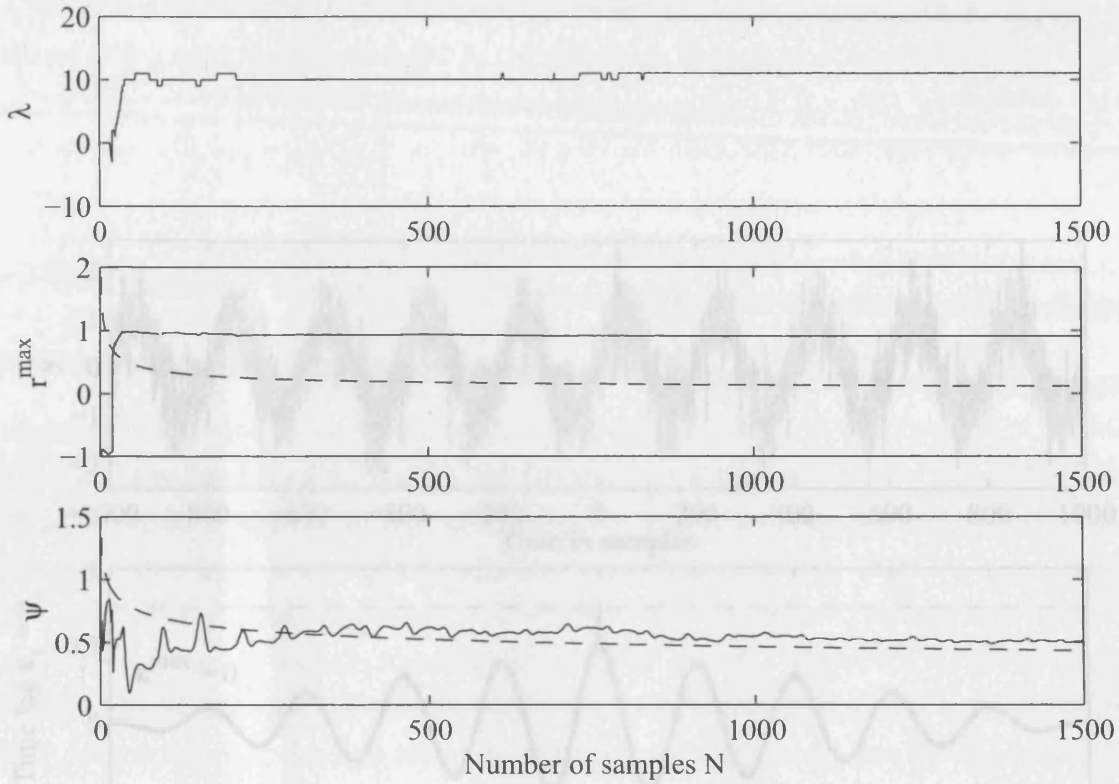


Figure 5.12: Detected time delay λ, correlation index $r^{\max}$ and oscillation index ψ as a function of the number of samples N for the time trends TI4 and TI5 from the reference case study. Solid line: measured indices, dotted line: threshold values r_{thresh} and ψ_{thresh}.

5.2.4 Merits and Limitations

The CCF algorithm for time lag detection has a number of important benefits. Firstly, it is easy to understand and to implement. The CCF is a common feature of process monitoring tools and therefore readily available. There are no parameters which, if changed only slightly, have a large impact on the results. The guidelines are equally straight forward. Good results can be achieved for oscillatory signals as well as irregular disturbances and the case study of Section 4.2 gave unambiguous indications towards the root cause.

There are also a number of drawbacks that have to be considered when applying the algorithm. Firstly, the requirement for any result is the presence of a time lag between the two investigated variables. The time lag or dead time is only one consequence of fault propagation, as explained in Section 2.4.3. All other effects, attenuation, magnitude decrease or the adding of noise are not considered with the CCF method. Secondly, the CCF is a linear tool that cannot deal with nonlinear systems[3]. Nonlinear in this context means that the second signal is derived from the first signal through a nonlinearity. As an example, the time delay between a random noise function and the square of the identical noise function delayed by a lag cannot be measured.

[3]For discussions of nonlinearity see Section 6.2.3.

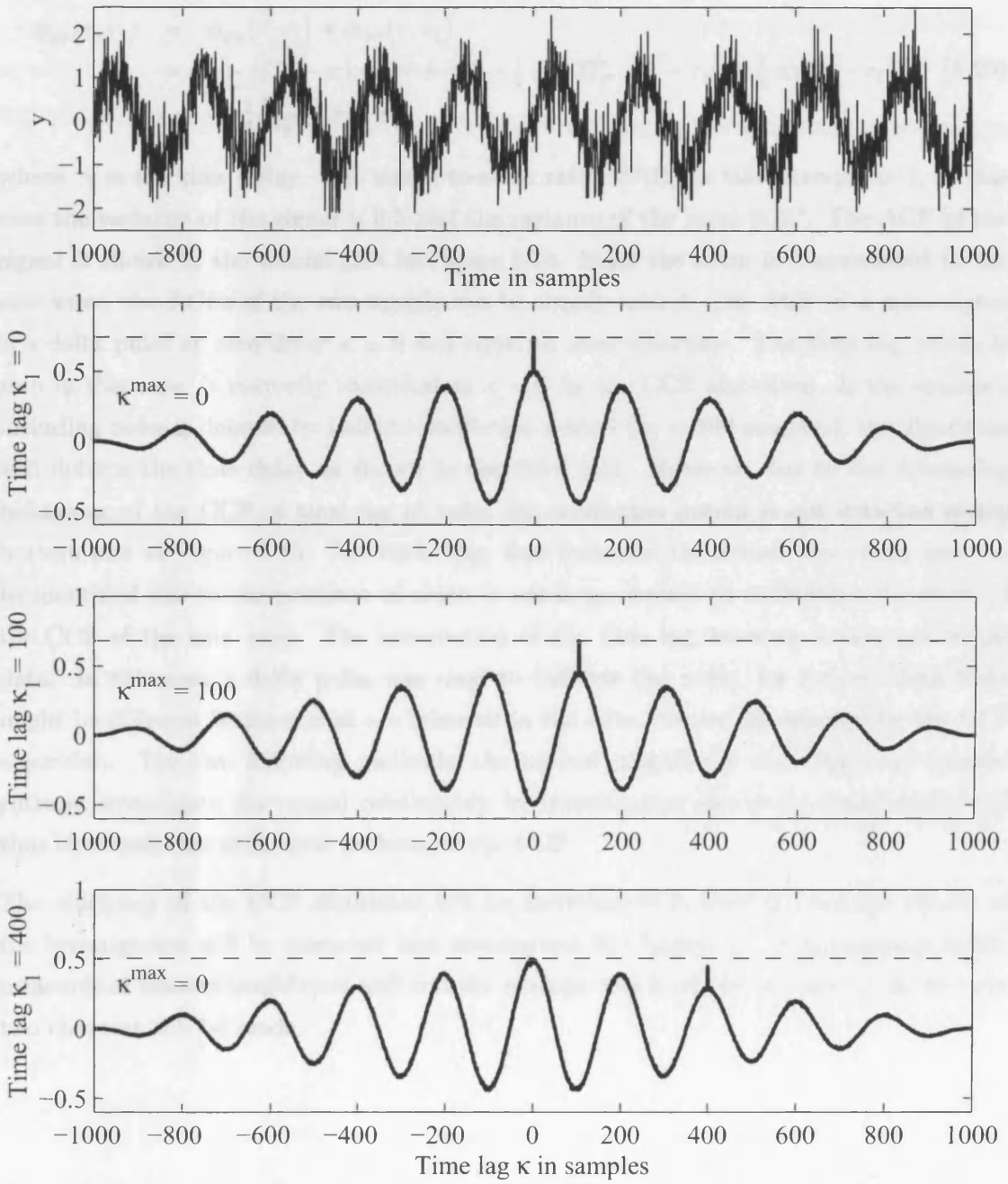


Figure 5.13: Sine wave with different time lags κ_1, samples per oscillation period: 200, number of total samples: 1000. Signal-to-noise-ratio: 2.

A structural problem of the CCF algorithm caused by the finiteness of the time series is shown in Figure 5.13 and explained in the following. Here, a sine wave with added noise is investigated. The CCF of the combined function $y(t) = x(t) + n(t)$ with $x(t) = \sin t$ and $n(t) \sim N(0, \sigma_n)$ where N is a Gaussian noise function with zero mean and a variance of σ_n:

$$\begin{aligned} \phi_{yy}(\tau, \tau_1) &= \phi_{xx}(\tau, \tau_1) + \phi_{nn}(\tau, \tau_1) \\ &= \frac{1}{2T_p}\left[(T_p - \tau)\cos(\tau + \tau_1) - \frac{1}{2}\sin(2T_p - \tau - \tau_1) + \frac{1}{2}\sin(\tau - \tau_1)\right] \\ &\quad + \sigma_n^2 \frac{T_p - \tau}{T_p}\delta(\tau - \tau_1) \end{aligned} \tag{5.29}$$

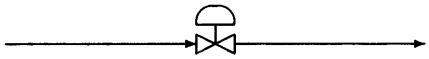
where τ_1 is the time delay. The signal-to-noise ratio (SNR) in this example is 2, in this case the variance of the signal is 0.5 and the variance of the noise 0.25[4]. The ACF of the signal is shown in the second plot in Figure 5.13. Since the noise is uncorrelated to the sine wave, the ACFs of the two signals can be simply added. The ACF of a noise signal is a delta pulse at zero delay $\kappa = 0$ and equal to zero otherwise. The time lag, which is zero in this case, is correctly identified at $\kappa = 0$ by the CCF algorithm. If the sequence including noise is delayed by half the oscillation period ($\kappa_1 = 100$ samples), the algorithm still detects the time delay as shown in the third plot. However, due to the decreasing behaviour of the CCF, a time lag of twice the oscillation period is not detected in the bottom plot of Figure 5.13. The little blip, that indicates the actual time delay and can be identified due to the presence of noise, is not large enough to outbalance the decay of the CCF of the sine wave. The information of the time lag, however is inherent in the data. In this case a delta pulse was used to indicate the noise, for process data there might be different features that are inherent in the data but are disregarded by the CCF algorithm. The two following methods, the nearest neighbours algorithm and transfer entropy, investigate the causal relationship by investigating also exceptional events and thus overcome this structural problem of the CCF.

The efficiency of the CCF algorithm will be investigated in Part III and the results of the investigation will be discussed and summarised in Chapter 11. A comparison to the measures of nearest neighbours and transfer entropy which will be introduced in the next two chapters will be made.

[4]The SNR is defined as SNR $= \frac{\sigma_s^2}{\sigma_n^2}$ where σ_s^2 is the variance of the signal and σ_n^2 the noise variance.

Chapter 5 Summary

In this chapter, an algorithm is proposed that uses the cross-correlation function to detect a time delay and thus argues cause-and-effect. For validation of the time delay, two indices, a correlation and an oscillation index are developed. Threshold values for these indices were derived from empirical observation of two uncorrelated signals. The proposed algorithm could be successfully applied to the reference case study.



Chapter 6

Nearest Neighbours

This chapter develops the concept of nearest neighbours for causal analysis between two process variables using historical process data. The basic concepts of phase space representation and embedded vectors which underly the nearest neighbours approach are discussed. The properties of predictability and nonlinearity of the time trend and nearest neighbours measures for these are introduced. The application of these properties for fault diagnosis is given. A centrepiece of this chapter is a new method which uses nearest neighbours and which has been developed for cause-and-effect analysis in the context of fault diagnosis.

Various methods that use the nearest neighbours principle to argue cause and effect have been introduced during the last decade. In the first section, the definition of delay coordinate maps and embedded vectors are given and the background terminology is discussed. Embedded vectors are the foundations of the methods for predictability and nonlinearity. In Section 6.2, measures for both predictability and nonlinearity are introduced and their application to historical process data is shown. The use of measures for nonlinearity for fault diagnosis is summarised and explained using the case study from Section 4.2.

A contribution of this work is the development of a generic structure that forms the basic steps of the nearest neighbours method. In Section 6.3, the established methods are grouped in the framework of variations in the basic algorithm. Furthermore, the effect of these variations is tested on process data and a best method is established on the basis of the results from the application to process data. A significance level that give guidance whether to accept or reject the causality measure is proposed. The developed causality measure is applied to the reference case study.

6.1 Phase Space Representations of Dynamic Systems

When receiving data from the process, a sample time trend is the only information available while the underlying dynamic system that generated the time trend is unknown. Measurements are easy to obtain while the construction of a model of the system might be very complex. This section shows that useful information about the underlying nature of the process can be derived by constructing trajectories from the time trend. These representations describe the dynamics for both linear or nonlinear systems. The foundation for a number of properties that describe the underlying system from a sample time trend is Taken's theorem proposed by Takens in 1981 [128]. It loosely says that the trajectory formed from the sample time series is related to the actual phase space trajectory of the underlying system. Here, phase space trajectory refers to the movement of the variables of a dynamical system that is described for example by a transfer function. The trajectories give an estimate of a variable in relationship to its previous status. Taken's theorem thus says that once properties are established for the time trend they are also valid for the dynamical system.

Phase space representations of trajectories are delay coordinate maps and embedded vectors. Embedded vectors are trajectories formed from the time series by stacking the time series in groups in sequential order and are used extensively to describe the system structure. The concept of delay coordinate maps and embedded vectors will be explained in more detail in the next sections. Applications that follow from the reconstruction of the phase space through delay coordinate maps and embedding are measures for continuity, determinism and predictability. If the predictability is taken a step further to deal with two variables at the same time, a measure of causality can be derived. This is achieved by asking the question whether a variable A can predict a variable B more than B predicts A. Sauer [108] gives a description of the same principle for one variable:

> "We identify the present state of the system which is producing the time series and search the past history for similar states. By studying the evolution of the observable following the similar states information about the future can be inferred."

The principle is, for all derived measures, the method not only of embedded vectors, but more specifically the method of finding the nearest neighbours of the embedded vectors. The term "nearest neighbours" is used throughout this chapter to indicate that methods for prediction and embedding are considered that at some stage find the nearest neighbours. The method of transfer entropy, introduced in Chapter 7, also uses the same prediction mechanism but pursuing an alternative approach. Most of the literature

that discussed the methods of nearest neighbours, is grouped under the terms nonlinear prediction, embedding, chaos and forecasting. Background literature on embedding and surrogate data can be found in Kantz and Schreiber [57], also in [112] and in [114] by Schreiber and Schmitz. Part III in Ott et al. [87] as well as the proceedings of the Santa Fe competition edited by Weigend and Gershenfeld [147] comprise a number of interesting articles on the prediction of discrete dynamic systems.

6.1.1 Delay Coordinate Maps

The motivation for using delay coordinate maps is that time evolution of a process can be defined in a phase space [57]. Delay coordinate maps are therefore also referred to as phase portraits. In delay coordinate maps all available measurements of a time series x_i are plotted against one or more previous values $x_{i-\kappa}$ to represent the phase state. The adjustment of the parameter κ is crucial for an appropriate construction of the trajectory. The optimal value of κ depends on the nature of the time series. For example, a sinusoidal sequence with N_p samples per cycle can be best displayed in the delay coordinate map if κ is set to $N_p/4$, because the resulting map is a circle. If κ is set to a too small value the map will be close to a linear function. This effect is called "false nearest neighbours" and used by Kennel et al. [62] to find the optimal delay κ as explained briefly later on in this section.

The idea of constructing maps from experimental time series was first introduced in the late seventies [87] by, for example, Glass and Mackey [36]. If the data is periodic, the delay map expresses the significant features of the nature of the oscillation. Visual inspection shows these features such as symmetry or resemblance to squares and circles which can be used for diagnosis as described in the following application to the reference case study from Section 4.2.

Case Study: *Three time trends of the case study from Section 4.2 are chosen to visualise the use of delay coordinate maps. The upper left plot, TI1, is a very nonlinear time sequence (the nonlinearity of this particular sequence will be investigated in Section 6.2.3) with spikes and episodes when the signal is constant. This is reflected in the delay coordinate map to the right by a corner consisting of a horizontal and vertical stretch of line. The middle left plot, TI7, looks far more like a sinusoidal time sequence. The delay map is not a perfect circle, which it would be for a sinusoidal oscillation, but very close to it with some indent on the upper left corner. The random signal, TI6, in the lower plot gives at a first glance no indication of regularity. However, calculating the ACF of the time trend of TI6, as shown in Figure 5.1, revealed an underlying oscillation. This shows that the delay coordinate map is not robust to a low signal to noise ratio. The delay of the embedding was chosen differently for the two time trends, a delay of 5 samples was used for TI1, 12 samples for TI7 and 1 sample for TI6.*

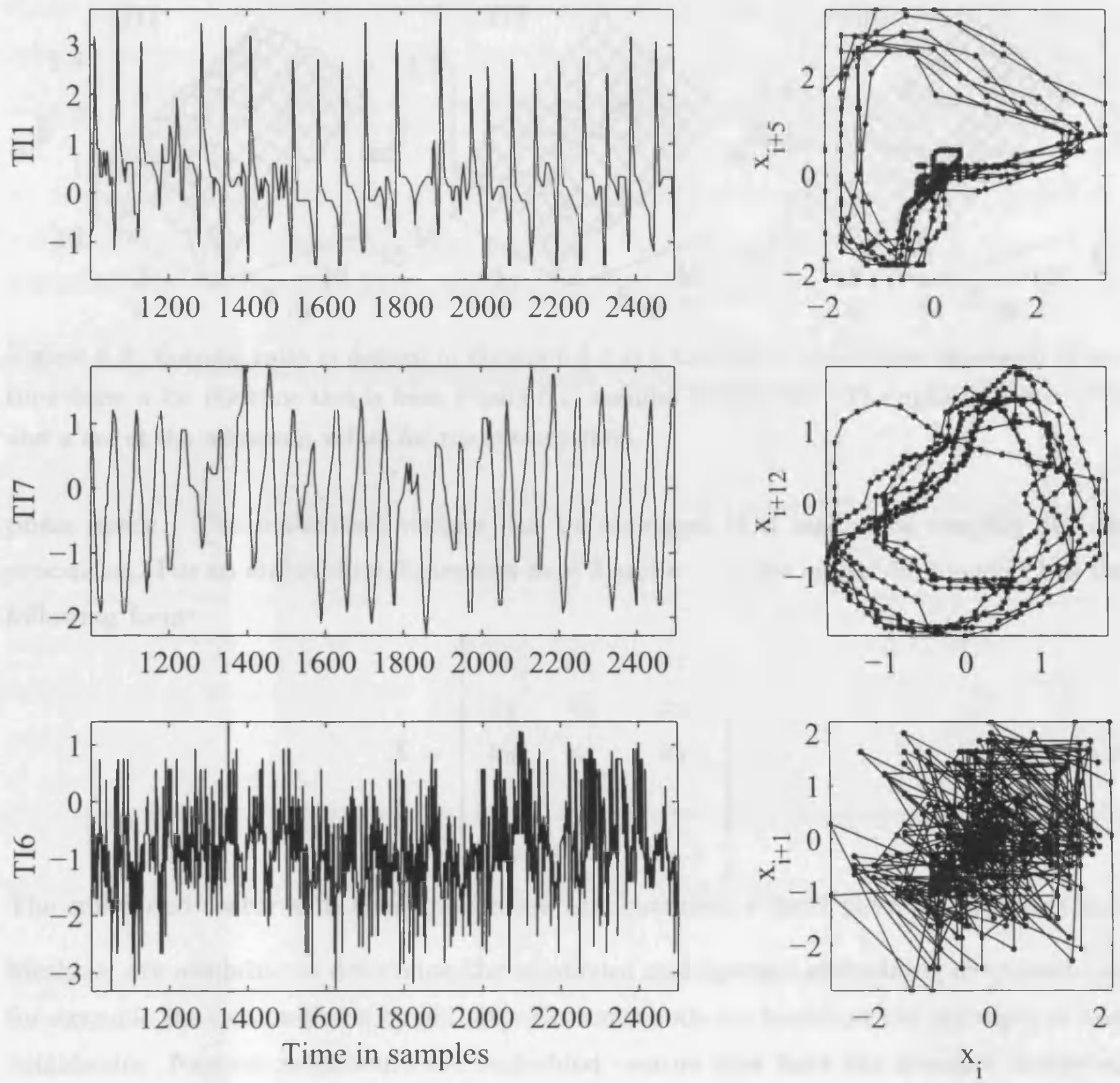


Figure 6.1: The delay coordinate maps for time trends from reference case study show almost a square form for the nonlinear signal in the upper plots and a circular, or linear, structure for the nearly sinusoidal signal in the middle plots. No structure can be observed for the random appearing time series, lower plots.

6.1.2 Embedded Vectors

In the previous paragraph, delay coordinate maps were introduced as a way of describing the dynamic system. A more generic, high dimensional and therefore non-graphical representation of the time trend are embedded vectors. Embedded vectors[1] are constructed by arranging the N available measurements $x_1 \dots x_N$ in $N^* = N - (m-1)\kappa$ m-dimensional vectors

$$\mathbf{x}_i = [x_i, x_{i-\kappa} \dots x_{i-(m-1)\kappa}] \text{ for } i = (m-1)\kappa + 1 \dots N. \tag{6.1}$$

Two critical parameters are embedding dimension m and time lag κ. The embedding dimension is the critical parameter that defines the quality of the representation of the

[1] For mathematical formulation of the embedding methods see article "Embedology" by Sauer et al. [107]

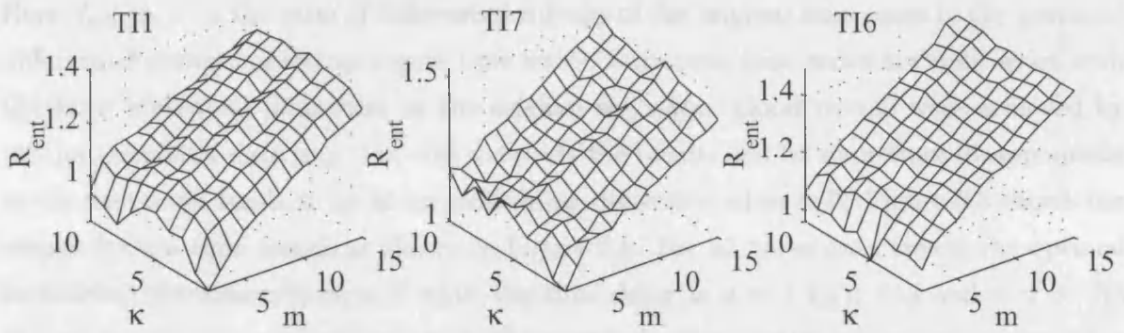


Figure 6.2: Entropy ratio as defined in Equation 6.3 as a function of embedding dimension m and time delay κ for the time trends from Figure 6.1, samples 2000:6:3200. The optimal values of m and κ are at the minimum values for the entropy ratio.

phase space. The embedded vectors can be arranged in a matrix to simplify further processing. For an embedding dimension $m = 3$ and $\kappa = 1$, the embedding matrix has the following form:

$$\mathbf{X} = \begin{bmatrix} x_3 & x_2 & x_1 \\ x_4 & x_3 & x_2 \\ x_5 & x_4 & x_3 \\ \vdots & \vdots & \vdots \\ x_N & x_{N-1} & x_{N-2} \end{bmatrix}. \tag{6.2}$$

The embedded vector $\mathbf{x}_i$ is then the ith row and therefore a short piece of the time trend.

Methods are available to determine the minimum and optimal embedding dimension, see for example the two methods by [62, 59]. Both methods are based on the principle of false neighbours. Nearest neighbours are embedded vectors that have the smallest Euclidean distance because of repetitive dynamics. False nearest neighbours occur if an embedding dimension is too small to unfold the delay coordinate map, not all points that lie close to one another will be neighbours because of the dynamics. The false nearest neighbours criterion of the optimal value for m is based on the increase of the distance of two nearest neighbours if the embedding dimension is increased by one to $m + 1$. If the increase is large, then it is assumed that they were false nearest neighbours.

Time lag κ is in most cases set to one since adjusting the sampling rate is equivalent to adjusting κ. A combined optimisation scheme of embedding dimension and time lag is presented by Gautama et al. [32]. They argue that the set of optimal parameters for m and κ yields a phase space representation which best reflects the dynamics of the underlying signal production system. Therefore it is expected that this representation has a minimal value for a quantity called differential entropy. Differential entropy is a measure of uncertainty of a signal which is estimated using the probability density function of the signal. The entropy ratio to be minimised is given by:

$$R_{\text{ent}}(m, \kappa) = I_{\text{ent}}(m, \kappa)(1 + \frac{m \ln N^*}{N^*}). \tag{6.3}$$

Here, $I_{\text{ent}}(m, \kappa)$ is the ratio of differential entropy of the original time series to the averaged differential entropy of the surrogate time series. Surrogate time series are time series with the same statistical properties as the original sequence. Good results were achieved by [32] for simulated data sets. For real data sets the results can be sometimes disappointing as the minimum tends to lie at an embedding dimension of $m = 2$. Figure 6.3 shows the results for the time trends as shown in Figure 6.1. For all three data trends the optimal embedding dimension is $m = 2$ while the time delay is $\kappa = 1$ to $\kappa = 3$ and $\kappa = 5$. As the results are not very encouraging, the approach of empirical optimisation through a number of data sets is pursued in Section 6.3.4.

6.2 Nearest Neighbours for Predictability and Nonlinearity

This section discusses measures for predictability and nonlinearity based on the concept of nearest neighbours. The terms continuity, determinism, randomness and uncertainty are defined first in both their literal and mathematical meaning. In particular, measures for predictability and nonlinearity are defined and discussed. The application of these methods to fault diagnosis purposes is given alongside the previously introduced examples.

6.2.1 Continuity, Determinism and Predictability

Characteristics of both the dynamical system and the time trend it produces include continuity, determinism, predictability and uncertainty. These properties, as well as statistics to quantify them, have been studied by the following authors [60, 93, 106]. The following paragraph argues that continuity, determinism, predictability and uncertainty are closely related and even identical in some aspects in a mathematical sense. These characteristics can be used for fault diagnosis as they give insight into the nature of the time trend of a disturbance.

Before establishing a measure for continuity, Kaplan [60] defines it as follows:

> "To test for the *continuity* of the underlying map, we want to answer a question like, 'If two points x_i and x_j are close together, are their images x_{i+1} and x_{j+1} also close together?'"

If the question has a positive answer than the time series is considered continuous. In this context, the difference between continuity and determinism is that the term deterministic applies to time dependent functions, that is, $x = f(t)$ while continuity applies to invertible functions that map one variable onto another, such that $y = f(x)$. Pecora et al. [93] give in this context the following description of a deterministic system:

> "A basic property [...] of *deterministic systems* is that of continuity forward in time. That is, points very close in phase space should map forward to points still close in phase space."

The counterpart of determinism is randomness. Pecora's view conflicts at a first glance with the traditional definition of continuity. This is because continuity is usually understood in the context of continuous functions. Here, however, the time series is available as a discrete function with distinct points x_i. Continuity is not necessarily visible in the time trend of discrete functions but rather in the phase portraits that reflect the functional relationships. A further property that is often given to describe time series is predictability. A definition that links determinism closely to predictability is given by Salvino et al. [106]:

> "*Predictability* indicates to what extent the past can be used to determine the future."

Based on this definition determinism and predictability will be treated as equal properties in this chapter. If predictability is treated as equivalent to certainty, then uncertainty is defined as the counterpart to predictability, such a definition makes sense in the non-scientific use of the two words. If a time series is not predictable it is unpredictable or uncertain.

The development of statistical methods for predicting a time series through knowledge of historical data of the time series was boosted in 1991 when the Santa Fe Institute called for a competition to predict the future trend of time sequences [147]. The aim of this competition was to provide a structure and compare quantitative results by analysing six time trends from disciplines such as physics, physiology, economics, astrophysics and music. The winning prediction method was by Sauer [108] and based on delay coordinate embedding and finding the nearest neighbours. In the following years research effort was focused on developing further predictability methods and providing a mathematical framework for these methods. The following predictability test by Kaplan is a direct result of the competition. The Kaplan method has been chosen because it gave robust results.

6.2.2 A Test for Predictability

A test for continuity and predictability of the underlying dynamic system proposed by Kaplan in 1994 is the delta-epsilon method [60]. It is based on the principle of finding the nearest neighbours in a sample time series. The distance between all pairs of sample

points x_i and x_j, $\delta_{i,j} = |x_i - x_j|$, are compared to the distance between their images x_{i+1} and x_{j+1}, $\epsilon_{i,j} = |x_{i+1} - x_{j+1}|$. The ϵ-values for which the corresponding $\delta_{i,j}$ lie in the proximity of a radius ϱ are averaged by:

$$\epsilon(\varrho) = \overline{\epsilon}_{i,j}, \text{ for all } i,j \text{ such that: } \varrho \leq \delta_{i,j} \leq \varrho + \Delta\varrho. \tag{6.4}$$

Here, $\Delta\varrho$ denotes the width of the "bins" used for averaging and the overline indicates averaging. The purpose of the calculation is to check if the images are close together when the sample points are close together. Predictability is argued through this comparison. An upper limit for ϵ is given by the a linear approximation if x_i and x_j are close together and if ϱ is small:

$$\epsilon(\varrho) \leq \overline{\lambda} \cdot (\varrho + \Delta\varrho), \tag{6.5}$$

where $\overline{\lambda}$ is the averaged absolute value of the slope of a linear approximation of the distance between the images of x_i and x_j on an ϵ versus ϱ graph. The reason for Equation 6.5 is that small values of δ imply small values of ϵ while large values imply large values if there is a functional relationship between a value and its image. When plotting all ϵ versus all δ values for the combinations of i,j, the points lie under a curve that is piecewise linear according to $\overline{\lambda}$, in case of continuity. These scatter plots are difficult to measure by a singular number so rather the plot of ϵ over ϱ is used for a measure. If ϵ approaches zero for ϱ approaching zero then the underlying nature of the time series can be considered as continuous. The delta-epsilon method was developed for simulated data such as the tent map[2] and is not very robust for noisy signals.

The continuity measure of the delta-epsilon method can be made more robust by applying embedded vectors[3]. For this purpose, an array of embedded vectors is constructed as defined in Equation 6.1. The embedded vectors are arranged in a matrix as defined in Equation 6.2. The embedded vector $\mathbf{x}_i$ is then of length m and a small part of the time trend. The future image x_{i+1} continues to be a single value. The parameter δ is given by the distance between two embedded vectors $\delta_{i,j} = ||\mathbf{x}_i - \mathbf{x}_j||$. The distances ϵ are $|x_{i+1} - x_{j+1}|$ as before and form now the images of the embedded vector. The nearest neighbours of each embedded vector are found by minimising δ; the result is now more robust towards noise. This method is only one in a number of methods for evaluating predictability. Previous approaches that predict time series include a forecasting technique by Farmer and Sidorowich [27] and by Casdagli [12]. The delta-epsilon method, though, has the advantage of being simple and well established at the same time. Cao and Mees [11] use a one step ahead prediction which compares the predictability of a future value rather than the current value.

[2]The tent map a time series that appears at a first glance similar to white noise but is based on the following simple equation: $x_{i+1} = 2x_i$ if $x_i < 0.5$ and $x_{i+1} = 2(1 - x_i)$ if $x_i > 0.5$.

[3]A Matlab implementation of the delta-epsilon method using embedded vectors can be found on the website of D.T.Kaplan http://www.macalester.edu/~kaplan/Software/, July 2004.

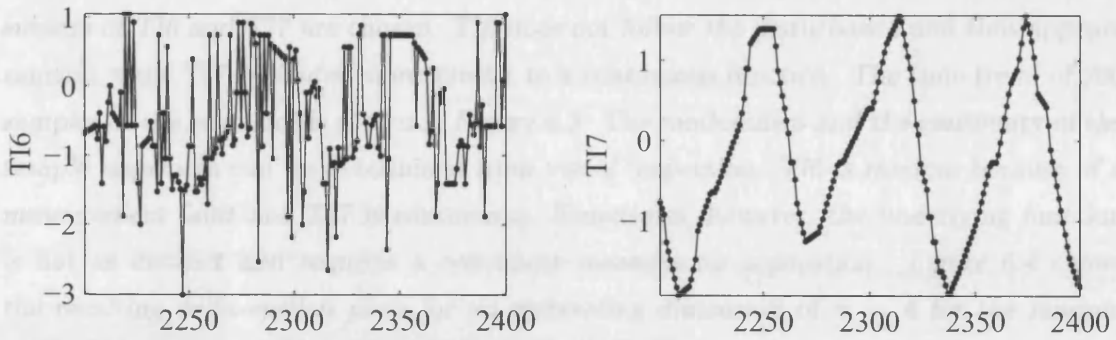


Figure 6.3: Extract of time trend of reference case study, temperature measurements TI6 and TI7, for delta-epsilon continuity analysis. TI6 is chosen to represent a random function while TI7 is more continuous as it seems to follow an oscillatory function.

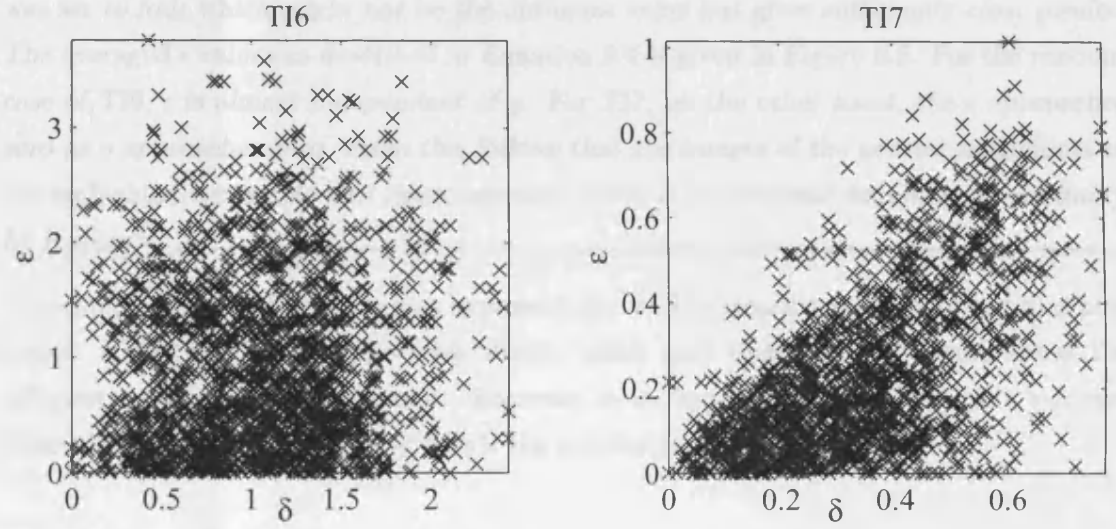


Figure 6.4: Delta-epsilon plots for the time trends as shown in Figure 6.3; embedding dimension $m = 4$.

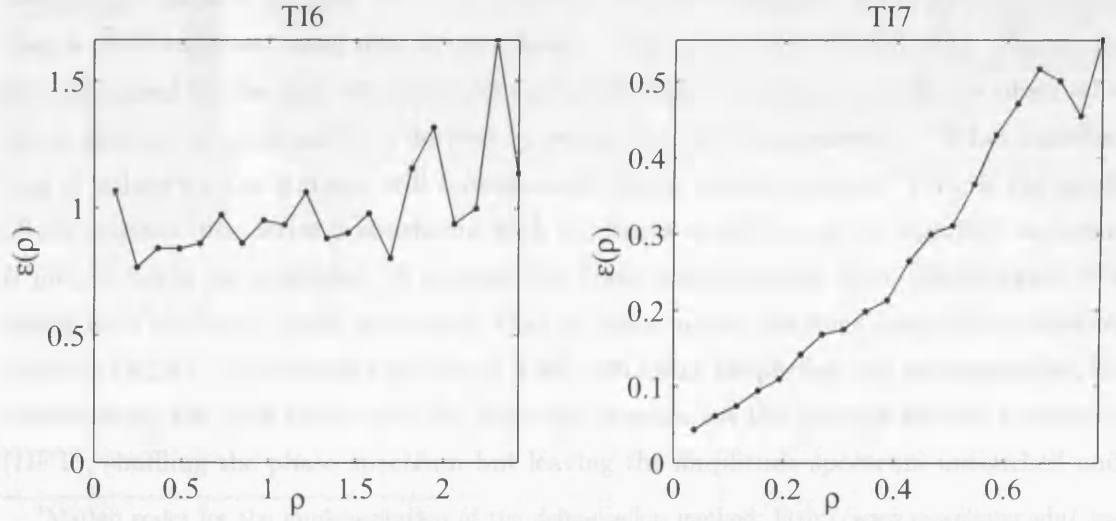


Figure 6.5: Plots of averaged ϵ values as defined in Equation 6.4 over distance parameter ϱ for time trends as shown in Figure 6.3; embedding dimension $m = 4$.

Case Study: *An example is given using data from Example II in Section 4.2. Data subsets of TI6 and TI7 are chosen. TI6 does not follow the disturbance and thus appears random while TI7 oscillates more similar to a continuous function. The time trend of 200 samples of the sequence is shown in Figure 6.3. The randomness and the continuity of the sample sequences can be determined from visual inspection. TI6 is random because of a measurement fault and TI7 is continuous. Sometimes, however, the underlying function is not as distinct and requires a continuity measure for separation. Figure 6.4 shows the resulting delta-epsilon plots for an embedding dimension of $m = 4$ for the random and continuous time trend. While the left panel appears more like a random blur, the right panel seems to lie under a boundary curve that is almost linear. This curve is the approximated and linear upper boundary as described in Equation 6.5. The delta-epsilon plot for TI7 looks therefore more structured. In both cases the embedding dimension m was set to four which might not be the optimum value but gives sufficiently clear results. The averaged ϵ values as described in Equation 6.4 is given in Figure 6.5. For the random case of TI6, ϵ is almost independent of ϱ. For TI7, on the other hand, the ϵ approaches zero as ϱ approaches zero. From this follows that the images of the nearest neighbours of the embedded vectors lie also close together, which is the original definition of continuity by Kaplan*[4]. ———————————————————————

The computational effort increases exponentially with increasing number of samples processed. Here, the 200 samples were chosen which gave unambiguous results about the different nature of the two signals. However, it is then important to choose a relevant interval of the time trend during which the continuity can be observed.

6.2.3 Tests for Nonlinearity

The measures for continuity, determinism and predictability will be directly applied to two variables in the next section. There is, however, a further property of the dynamic system that is often analysed using nearest neighbours. This property is nonlinearity. The statistics computed by the nearest neighbours principle can be used as a nonlinear observable and a measure of nonlinearity is derived by asking the following question: "What distribution of values for the statistic will a comparable linear model produce?"[57]. If the result of the original time series is consistent with the linear model it can be regarded as linear. If not, it might be nonlinear. A comparable linear model results from construction of a signal with the same linear properties, that is, linear mean, variance and autocorrelation function (ACF). A surrogate time series [130] with these properties can be constructed by transforming the time signal into the frequency domain via the discrete Fourier transform (DFT), shuffling the phase spectrum but leaving the amplitude spectrum untouched and

[4]Matlab codes for the implementation of the delta-epsilon method; http://www.macalester.edu/ kaplan/Software/, July 2004.

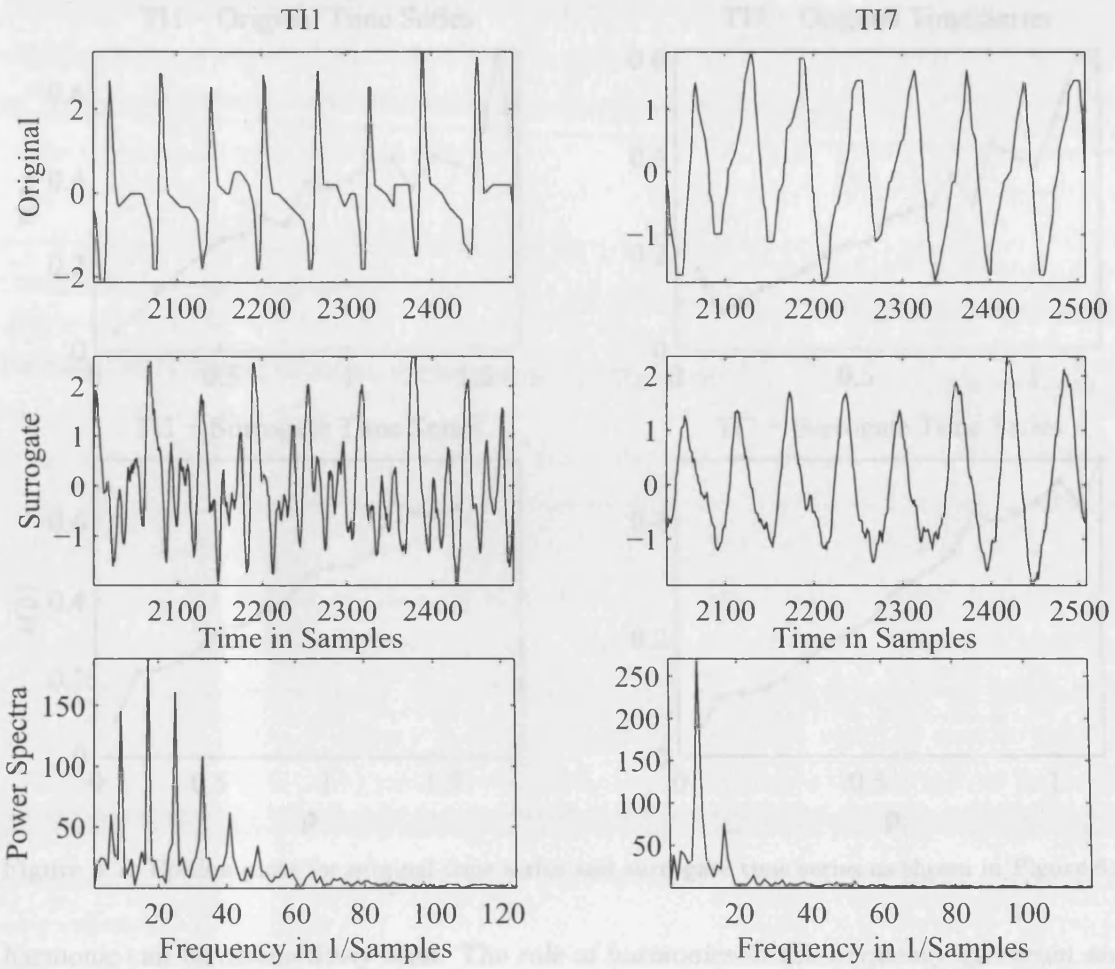


Figure 6.6: Original time series (upper panels) of time trends from reference case study, their surrogates derived from phase shuffling (middle panels) and the corresponding power spectra (lower panels).

transforming the signal back into the time domain[5]. When carrying out the surrogate data analysis, it is important to ensure that the data is end-matched. End-matching requires the starting value and first order derivative to match the value and first order derivative of the end of the sequence. The reason for this is that the sequence is finite and conversion into the frequency domain using the DFT can result in spectral leakage. Stam et al. [125] suggest an algorithm that minimises the frequency mismatch error between a time series and the delayed time series.

Case Study: *Figure 6.6 shows the effect of phase shuffling on two time trends from reference case study in Section 4.2, TI1 and TI7. The regular low frequency pattern of TI1 (left hand panels) is completely destroyed after computing the surrogates while the same low frequency persists, to some extent, for TI7 (right hand panels). Looking at the power spectra in the lower panels it is apparent that TI1 has more harmonics of the main oscillation, at least four are clearly recognisable, than TI7 for which only one*

[5]Thereby, the ACF of the signal remains unchanged as the power spectrum and the ACF are related through the Wiener-Kinchin theorem.

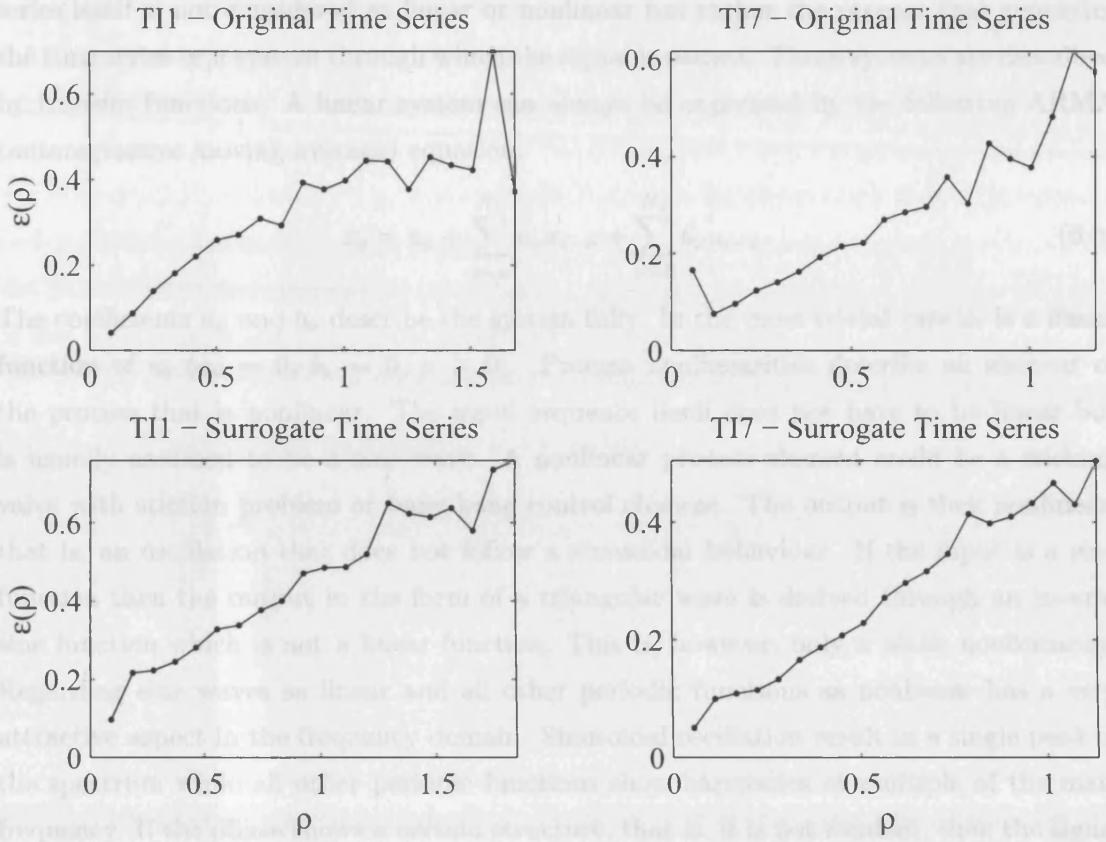


Figure 6.7: Epsilon plots for original time series and surrogate time series as shown in Figure 6.6

harmonic can be distinctively seen. The role of harmonics in the frequency spectrum and the relationship of nonlinearity and predictability will be explained later on in this section. First, the application of the delta-epsilon method by Kaplan to the original and surrogate time series from Figure 6.6 will show how a test for the nonlinearity can be derived. The shuffled data series in the middle left hand panel is less structured and thus less predictable than the original time series. This should result in an averaged ϵ value, Equation 6.4, that does not approach zero as δ, or ϱ, approaches zero. Figure 6.7 gives the plots for the averaged ϵ values over the distance ϱ of δ. The plots for both original time series and the surrogate of TI7 approach zero for small ϱ. The surrogate time series of TI1, however, which appears unstructured in the middle left hand panel of Figure 6.6, has a value of 0.3 for the smallest interval of ϱ and $\varrho + \Delta\varrho$. TI1 might therefore be nonlinear, although not only one but a number of surrogates should be considered for statistical significance. —

A similar approach to the delta-epsilon algorithm is the delay vector variance method by Gautama et al. [31]. Instead of the distances, variances of the embedded vectors and the variance of the images are computed to retrieve a statistic similar to the averaged ϵ plots. This provides a measure for both nonlinearity and noise.

Nonlinearity in the Context of Process Monitoring: The definition of nonlinearity in time series is by no means trivial. A signal that is often regarded as nonlinear is, for example, a triangular periodic wave. The reason for non-triviality is the fact that the time

series itself is not considered as linear or nonlinear but rather the process that generates the time series or a system through which the signal is passed. Those systems are described by transfer functions. A linear system can always be expressed by the following ARMA (autoregressive moving average) equation:

$$x_i = a_0 + \sum_{i=\nu}^{N_1} a_\nu x_{i-\nu} + \sum_{\nu=0}^{N_2} b_\nu u_{i-\nu}. \tag{6.6}$$

The coefficients a_ν and b_ν describe the system fully. In the most trivial case x_i is a linear function of u_i ($a_\nu = 0$, $b_\nu = 0$, $\nu > 0$). Process nonlinearities describe an element of the process that is nonlinear. The input sequence itself does not have to be linear but is usually assumed to be a sine wave. A nonlinear process element could be a sticking valve with stiction problem or bang-bang control element. The output is then nonlinear, that is, an oscillation that does not follow a sinusoidal behaviour. If the input is a sine function then the output in the form of a triangular wave is derived through an inverse sine function which is not a linear function. This is, however, only a *static* nonlinearity. Regarding sine waves as linear and all other periodic functions as nonlinear has a very attractive aspect in the frequency domain. Sinusoidal oscillation result in a single peak in the spectrum while all other periodic functions show harmonics at multiple of the main frequency. If the phase shows a certain structure, that is, it is not random, then the signal is considered nonlinear. The phenomena of showing structure in the phase spectrum is also called phase coupling.

Nonlinearity in time series is very relevant in the context of process monitoring and fault diagnosis because the underlying nonlinearity is often the root cause. A nonlinearity method based on nearest neighbours and surrogates was therefore successfully implemented as a diagnostic tool by Thornhill et al. [133, 135, 137]. The methods are based on similar approaches as described above based on methods as described by Sugihara and May [127] and Schreiber [112]. Choudhury [19] uses the bispectrum and bicoherence to exploit the effect of phase coupling in the frequency spectrum. A method based on Ljapunov exponents as described by Kantz and Schreiber [57] was applied in [155] to simulated data of two interacting control loops and to industrial data. Stam et al.[125], on the other hand, measure the nonlinearity of oscillating system by time reversibility and amplitude asymmetry, arguing that all linear signals are time reversible and amplitude symmetrical. The reverse conclusion is not valid as for example a square wave is both time reversible and symmetrical but has distinct harmonics and is therefore not considered as linear.

6.3 A Measure of Causality Using Nearest Neighbours

In this section, methods based on the concept of nearest neighbours that measure the predictability from one time series to a second time series are proposed for causality analysis in chemical processes. First the concept of predictability for two time series

is explained and the underlying mechanism of causality, generalised synchronisation or coupling, is described. A review of methodologies that estimate the interdependence of two time series using nearest neighbour principles is given next. In Section 6.3.2, two main frameworks of the causality measure are given, one based on causality and one on one step ahead prediction. The variations to the main framework are then implemented and applied to the case study in Section 6.3.3. The most successful method is chosen and the parameters are optimised (Section 6.3.7). A significance level that gives an estimate of the reliability of the measure is developed in Section 6.3.4.

6.3.1 Underlying Mechanism of Causality

In Section 6.2, methods to measure predictability were introduced. The definition of predictability for a single time series was to what extent the past can be used to determine the future. The same question can be extended to the case of two variables, rephrasing the predictability definition by Salvino et al. [106]:

> Predictability in the context of comparing two variables indicates to what extent the past of one variable can be used to determine the future of a second variable.

This definition results in a causality measure which can be derived by comparing the predictability from a first variable to the second with the predictability of the second variable to the first (see Section 4.1).

The underlying principle that is exploited to find causality using nearest neighbours is the principle of generalised synchronisation. Two systems are called synchronised in a generalised way if a causal functional relationship exists between the states of both systems. Therefore, if two time series are described by

$$y(t) = f(x(t)) \tag{6.7}$$

then $x(t)$ is identified as the driver and $y(t)$ as the response if f is a non-invertible function. When trajectories in the state space representation of systems $x(t)$ and $y(t)$ are connected by generalised synchronisation as in Equation 6.7, then two close states of the driver are connected to two close states of the response [103]. Taken's theorem [128] translates this property to the state space representation of embedded vectors. The existence of the relationship between driver and response is evaluated by finding the indices of the nearest neighbours of the driver at time i. The indices gained this way should give good estimates of the response values at time i. Generalised synchronisation is further studied by Kocarev and Parlitz [63] and Kocarev et al. [64].

The development of statistical measures for coupling strength, interdependency and hence causality started in the early 1990s as powerful computers and workstations became more readily available. One of the earliest approaches by Čenys et al. [13] is the mean conditional dispersion (MCD) which finds the nearest neighbours within a radius ϵ, similar to the delta-epsilon method by Kaplan in Section 6.2.2. The Euclidean distance between the original embedded vector of the response and the embedded vectors of the response with the indices of nearest neighbours of the driver are calculated. The dependency of the sum of the distance on ϵ is a measure of causality. A later application of MCD to human and animal electroencephalogram (EEG) signals was reported by Ulbikas et al. [140].

Rather than using the concept of nearest neighbours, Rulkov et al., [103] applied the principle of false nearest neighbours, see Section 6.1.1, to establish a causality measure. The mutual false nearest neighbours give then a measure of interdependency by comparing an increased embedding dimension of the driver while fixing the embedding dimension of the response variable. In 1996, Schiff et al. [110] introduced a method similar to [13]. Instead of considering the nearest neighbours within a radius, the number of nearest neighbours is fixed to a value K. Additional scaling has the effect that the resulting measure is smaller than one with one corresponding to independency and zero to dependency between the two variables.

Variations and generalizations of the method by Schiff et al. have been successfully applied to EEG signals and discrete dynamic systems [5, 114, 100]. Arnhold et al. [2] in particular suggest a number of generalisations and alternatives of the measure of dependency. Bhattacharya et al. [6] propose a modification which omits the first K nearest neighbours but considers the following $K+1,\ldots,2K$ arguing that noise can make the nearest neighbours of the driver better predictors of the response than the response itself. By considering only $K+1,\ldots,2K$ neighbours, the resulting measure will be more robust against noise. A further variation was proposed by Le Van Quyen [72]. Instead of summing the distance between embedded vectors, the distance between future images is summed. Thus, the predictability of a future value of the response is estimated by the embedded vector of the driver. Further approaches that consider a future value, referred to as "one step ahead prediction", are by [73, 11, 150, 28]. Goodridge et al. [37] investigated the use of variances instead of simple averaging over the number of nearest neighbours, similar to the DVV method for detection of nonlinearity [31, 32].

6.3.2 Embedding-Prediction and One-Step-Ahead-Prediction

In this section, the algorithm proposed for calculating a predictability based measure for dependency will be developed. A further, modified version of the algorithm will be given later in this section. The modified algorithm involves a structural difference compared to the algorithm in this section, the prediction of a step ahead in the future rather than the

prediction of an embedded vector. All other modifications that might improve the method are investigated in the next section. The proposed algorithm computes $H(X|Y)$ as a measure of interdependence, that is, how Y influences X. The inverse case, $H(Y|X)$ can be derived by exchanging time series X and Y. For notational purposes D_i is introduced as an interim measure for constructing H which has to be calculated for all embedded vectors. The proposed algorithm can be summarised in the following steps which will be explained below in more detail.

Algorithm - Embedding Prediction

Step 1. Construct embedded vectors $\mathbf{x}_i$, $\mathbf{y}_i$ and get distances between all pairings of embedded vectors $||\mathbf{x}_i - \mathbf{x}_j||$, $||\mathbf{y}_i - \mathbf{y}_j||$.

Step 2. Find indices $r_{i,j}$ and $s_{i,j}$ of K nearest neighbours of $\mathbf{x}_i$ and $\mathbf{y}_j$.

Step 3. Compute $D_i(X|X)$, $D_i(X|Y)$ as functions of $||\mathbf{x}_i - \mathbf{x}_j||$, $||\mathbf{x}_i - \mathbf{x}_{s_{i,j}}||$ over all $j = 1 \dots K$.

Repeat Step 2 to 3 for $i = 1 \dots N^*$

Step 4. Compute $H(X|Y)$ as function of $D_i(X|X)$, $D_i(X|Y)$) over all $i = 1 \dots N^*$.

Step 5. Compute causality measure $h_{x \to y} = H(X|Y) - H(Y|X)$.

Step 1. First, embedded vectors are constructed from time series x and y: $\mathbf{x}_i = [x_i, x_{i-\kappa}, \dots x_{i-(m-1)\kappa}] \in \mathrm{R}^{1,m}$ and $\mathbf{y}_i = [y_i, y_{i-\kappa}, \dots y_{i-(m-1)\kappa}] \in \mathrm{R}^{1,m}$. If the time series consists of N samples then $N^* = N - (m-1)\kappa$ embedded vectors are formed. For implementation, the embedded vectors are arranged in a matrix such that the embedded vectors form m rows:

$$\mathbf{X} = \begin{bmatrix} x_{(m-1)\kappa+1} & \cdots & x_1 \\ x_{(m-1)\kappa+2} & \cdots & x_2 \\ \vdots & & \vdots \\ x_N & \cdots & x_{N-(m-1)\kappa} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_{(m-1)\kappa+1} & \cdots & y_1 \\ y_{(m-1)\kappa+2} & \cdots & y_2 \\ \vdots & & \vdots \\ y_N & \cdots & y_{N-(m-1)\kappa} \end{bmatrix}. \tag{6.8}$$

$\mathbf{X} \in R^{N^*,m}$, $\mathbf{Y} \in R^{N^*,m}$. As described in Section 6.1.1, m is the embedding dimension and κ the time delay between two successive samples. The embedding dimension m as well as κ are sometimes chosen differently for x and y which does not affect the structure of the algorithm for causality, see for example [110]. However, as the signals that are investigated in the background of root cause analysis are similar due to the pre-selection of process variables, identical parameters for m are assumed. Optimization issues for m and κ are addressed in Section 6.3.3. The Euclidean distance for all combinations of the

embedded vectors $i \neq j$ is then computed by

$$\begin{aligned} d_{i,j} &= ||\mathbf{x}_i - \mathbf{x}_j|| \\ &= \sqrt{(x_i - x_j)^2 + (x_{i-\kappa} - x_{j-\kappa})^2 + \ldots + (x_{i-(m-1)\kappa} - x_{j-(m-1)\kappa})^2}. \end{aligned} \tag{6.9}$$

The generation of all distances at this early stage helps to implement a efficient computation since the distances are used in both Step 2 and 3, first to find the nearest neighbours and then to compute the statistic D_i. Additionally, the distances are symmetrical so that only $\frac{1}{2}N^*(N^* - 1)$ distances have to be calculated. This can be achieved efficiently if the distances are stored in a matrix $\mathbf{D} = \{d_{i,j}\} \in \mathrm{R}^{N^*,N^*}$. Thus, elements $d_{i,i}$ on the main diagonal equal zero and $\mathbf{D} = \mathbf{D}^{\mathrm{T}}$ holds.

Step 2. The K nearest neighbours for every embedded vector $\mathbf{x}_i$ and $\mathbf{y}_i$ result from the minimum value of the distances such that for embedded vector $\mathbf{x}_i$

$$r_{i,j} := \{ j \mid \min_{j \neq i} ||\mathbf{x}_i - \mathbf{x}_j|| \} \tag{6.10}$$

and for embedded vector $\mathbf{y}_i$

$$s_{i,j} := \{ j \mid \min_{j \neq i} ||\mathbf{y}_i - \mathbf{y}_j|| \}. \tag{6.11}$$

The minimum values and their indices can be found easily by sorting all distances, choosing only the K smallest distances and gathering $r_{i,j}$ and $s_{i,j}$. The indices can be summarised in matrices $\mathbf{R} \in \mathrm{R}^{N^*,K}$ and $\mathbf{S} \in \mathrm{R}^{N^*,K}$. The adjustment of parameter K, the number of nearest neighbours, is described in the following Section 6.3.7. This step has to be repeated for all $i = 1 \ldots N^*$.

Step 3. Next, the statistic D_i is computed from the distances of the nearest vectors. A simple solution is that the statistic is the averaged sum of the distances of all K nearest neighbours,

$$D_i(X|Y) = \frac{1}{K} \sum_{j=1}^{K} ||\mathbf{x}_i - \mathbf{x}_{s_{i,j}}||. \tag{6.12}$$

To scale this measure to sequence x, the average distance of all embedded vectors $\mathbf{x}_j$ with $j = 1 \cdots N^*$ is defined:

$$D_i(X|X) = \frac{1}{N^* - 1} \sum_{j=1}^{N^*} ||\mathbf{x}_i - \mathbf{x}_j||. \tag{6.13}$$

Both $D_i(X|Y)$ and $D_i(X|X)$ increase with increasing embedding dimension m since the square of the average distance of embedded vectors is the square root of the sum of all $(x_i - x_j)^2$ for $j = 1 \ldots m$ as described in Equation 6.9. This effect can be eliminated by multiplying D_i by factor $\frac{1}{m}$. Thus, D_i will be approximately in the interval $0 \leq D_i \leq 2$ if x is scaled to unit variance. It will lie there only approximately because of the stochastic nature of the measure. A benefit of the statistic is that there are no parameters to adjust in this step. The computation of D_i has to be repeated for all i.

Step 4. After repeating Step 2 and 3 over all embedded vectors an average value of D_i must be calculated to retrieve a single statistic. A straight forward averaging results in

$$H(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{D_i(X|Y)}{D_i(X|X)} \tag{6.14}$$

as the interdependence measure. The resulting measure H is an indication of interdependency between X and Y. If $H(X|Y)$ is much smaller than one, then Y influences X and vice versa, or in descriptive form

$$\begin{aligned} H(X|Y) < 1 \quad &\Rightarrow \quad Y \rightarrow X \\ H(Y|X) < 1 \quad &\Rightarrow \quad X \rightarrow Y. \end{aligned} \tag{6.15}$$

Other options for calculating statistic H will be discussed in the next section. No parameters are required for the calculation of $H(X|Y)$.

Step 5. In the last step, the two measures of x influencing y and y influencing x are compared to turn the predictability measure into a causality measure. This addresses the question whether x is a better predictor of y than y is of x.

The resulting measure of interdependence introduced here with definitions for D_i and $H(X|Y)$ is based on the measure defined by Schiff et al. [110]. More recently, Arnhold et al. [2] introduced a measure by defining $D_i(X|X)$ as

$$D_i^{(1)}(X|X) = \frac{1}{K} \sum_{j=1}^{K} \|\mathbf{x}_i - \mathbf{x}_{r_{i,j}}\| \tag{6.16}$$

by considering only the K nearest neighbours for $D_i^{(1)}(X|X)$ and then consequently they define

$$H^{(1)}(X|Y) = \sum_{i=1}^{N^*} \frac{D_i^{(1)}(X|X)}{D_i(X|Y)}. \tag{6.17}$$

This measure has been picked up subsequently by further authors[100, 114, 5, 6]. However, a major drawback comes with the definition of $D_i(X|X)$. By taking the nearest neighbours of the sequence and comparing it to its own embedded vectors gives a measure that not only depends strongly on the number of nearest neighbours but also cannot be easily scaled to a unit value for independency. The reason for this is that the distance between embedded vector $\mathbf{x}_i$ and its nearest neighbours $x_{r_{i,j}}$ is not the average value but a smaller value, even if x is completely random because the minimum is always taken. Thus, if K is increased, the average value $D_i(X|X)$ will increase, too. Also, since the data here is strongly periodic, the self predictability $D_i(X|X)$ will always be very small compared to the predictability due to the second variable, $D_i(X|Y)$, that is, X is a good predictor for itself. The detailed mathematical derivation of the expected value of $D_i(X|X)$ for the case that $m = 1$ is given in Equation B.11 Appendix B.1, the expectation value results in:

$$\mathrm{E}\{D_i(X|X)\} = \frac{4N}{\sqrt{3}K}\left[\sqrt{\frac{N-K}{N}} - 1\right] + 2\sqrt{3}\left[1 - \frac{2}{3}\sqrt{\frac{N-K}{N}}\right]. \tag{6.18}$$

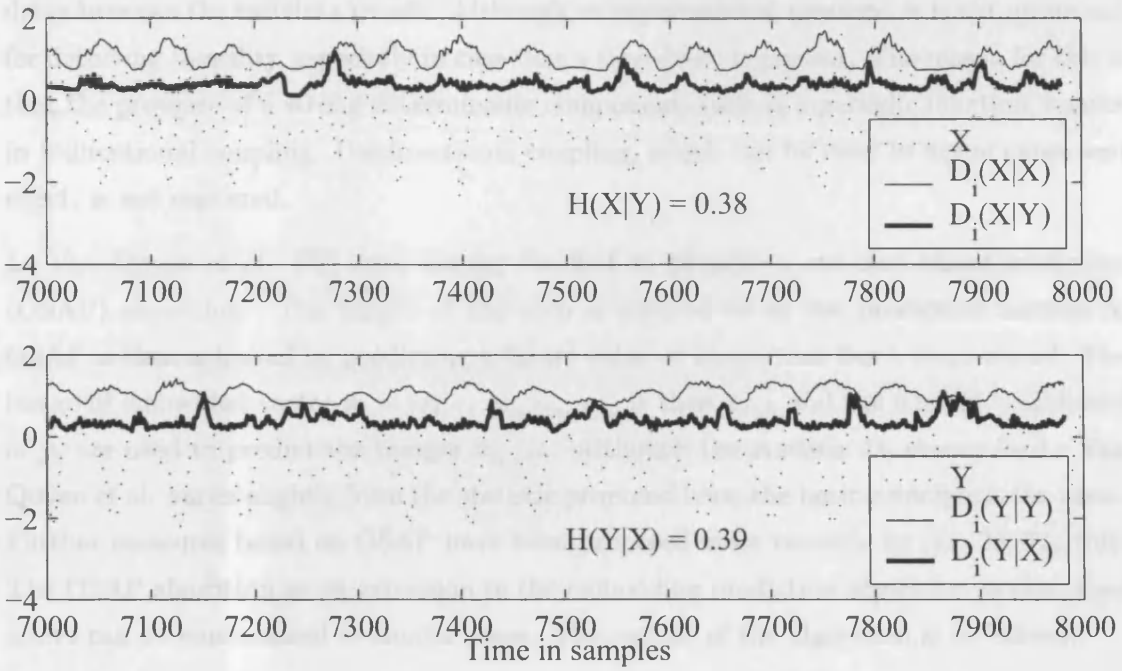


Figure 6.8: Construction and value of the embedding prediction for two time trends of the reference case study. X = TI4, Y = TI5, $N = 1000$, $m = 4$, $\kappa = 6$, $K = 20$.

This is an almost linear dependency on the ratio of K/N, as shown in the Appendix, and therefore has uncontrolled effects on the causality measure.

Case Study: *The algorithm as described above is applied to the case study of Section 4.2, again the relationship between temperature TI4 and TI5 is investigated here. The causality measure $H(TI4|TI5)$ results in a value of 0.38 and $H(TI5|TI4)$ in 0.39 when choosing an embedding dimension of $m = 5$, a time lag of $\kappa = 6$ and K, the number of nearest neighbours, equals 20. Both values of $H(TI4|TI5)$ and $H(TI5|TI4)$ are smaller than 1 and therefore TI4 influences TI5 and vice versa. To understand the construction of H, the trends of $D_i(X|X)$ and $D_i(X|Y)$ are plotted in the upper plot of Figure 6.8 for TI4 and TI5. The first observation is that $D_i(X|Y)$ is smaller than $D_i(X|X)$ throughout, thus the indication of interdependence. In case of independence, $D_i(X|Y)$ and $D_i(X|X)$ are identical. Interestingly, $D_i(X|X)$ oscillates. The oscillation exists because the time series is not stationary. Thus, every time x is close to its mean value for a longer time, $D_i(X|X)$ is at a minimum, that is, the average distance to all other embedded vectors is then minimal. The lower plot of Figure 6.8 shows a similar oscillatory behaviour. The main problem of the interdependence measures $H(TI4|TI5)$ and $H(TI5|TI4)$, however, is that their values though smaller than one are almost identical (0.38 and 0.39) and thus do not imply causality. This is mainly because there is a time delay of 10 samples between the oscillations in x and y, see Table 5.2. This time delay cannot be incorporated in the measure of interdependence. Thus, the equations for D_i, Equation 6.12 and 6.13, will be modified as suggested first by Le Van Quyen et al. [72] as shown in the next paragraph.*

The reason why the causality measure did not work is the inability to incorporate a time

delay between the two data trends. Although an asymmetrical measure, it is not optimised for detecting causality, especially in case that a time delay is present. The reason for this is that the presence of a strong deterministic component, such as a periodic function, results in bidirectional coupling. Unidirectional coupling, which can be used to argue cause and effect, is not captured.

Le Van Quyen et al. [72] were among the first to propose a one-step-ahead-prediction (OSAP) algorithm. The length of the step is referred to as the prediction horizon h. OSAP is then achieved by predicting a future value or image that lies h steps ahead. The image of embedded vector $\mathbf{x}_i = [x_i, \ldots x_{i-(m-1)\kappa}]$ is then x_{i+h} and the nearest neighbours of $\mathbf{y}_i$ are used to predict the images $x_{s_{i,j}+h}$. Although the statistic D_i chosen by Le Van Quyen et al. varies slightly from the statistic proposed here, the basic principle is the same. Further measures based on OSAP have been proposed more recently by [11, 28, 73, 150]. The OSAP algorithm as an extension to the embedding prediction algorithm as described above can be summarised in similar steps. The outline of the algorithm is as follows.

Algorithm - One Step Ahead Prediction

Step 1. Construct embedded vectors $\mathbf{x}_i$, $\mathbf{y}_i$, their images x_{i+h}, y_{i+h} and
- get distances $||\mathbf{x}_i - \mathbf{x}_j||$, $||\mathbf{y}_i - \mathbf{y}_j||$ as well as
- distances of images $|x_{i+h} - x_{j+h}|$, $|y_{i+h} - y_{j+h}|$.

Step 2. Find indices $r_{i,j}$ and $s_{i,j}$ of K nearest neighbours of $\mathbf{x}_i$ and $\mathbf{y}_i$.

Step 3. Compute $D_i^o(X|X)$, $D_i^o(X|Y)$ as functions of
- $|x_{i+h} - x_{j+h}|$, $|x_{i+h} - x_{s_{i,j}+h}|$ over all $j = 1 \ldots K$.

Repeat Step 2 and 3 for $i = 1 \ldots N^*$

Step 4. Compute $H^o(X|Y)$ as function of $D_i^o(X|X)$, $D_i^o(X|Y)$).

Step 5. Compute causality measure $h_{x \to y} = H^o(X|Y) - H^o(Y|X)$.

The norm $|\cdot|$ is used since the images are single values. The measures $D_i^o(X|Y)$ and $D_i^o(X|X)$ are here defined by

$$D_i^o(X|Y) = \frac{1}{K} \sum_{j=1}^{K} |x_{i+h} - x_{s_{i,j}+h}|. \tag{6.19}$$

To scale this measure to sequence x, the average distance of all images is computed.

$$D_i^o(X|X) = \frac{1}{N^* - 1} \sum_{j=1}^{N^*} |x_{i+h} - x_{j+h}|. \tag{6.20}$$

Since now no embedded vectors are used for the computation of $D_i^o(X|X)$, it be interpreted as the standard deviation of the time series where the mean of the time series is x_{i+h}. The computation of $H^o(X|Y)$ is identical to the computation as described in Equation 6.14.

$$H^o(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{D_i^o(X|Y)}{D_i^o(X|X)} \tag{6.21}$$

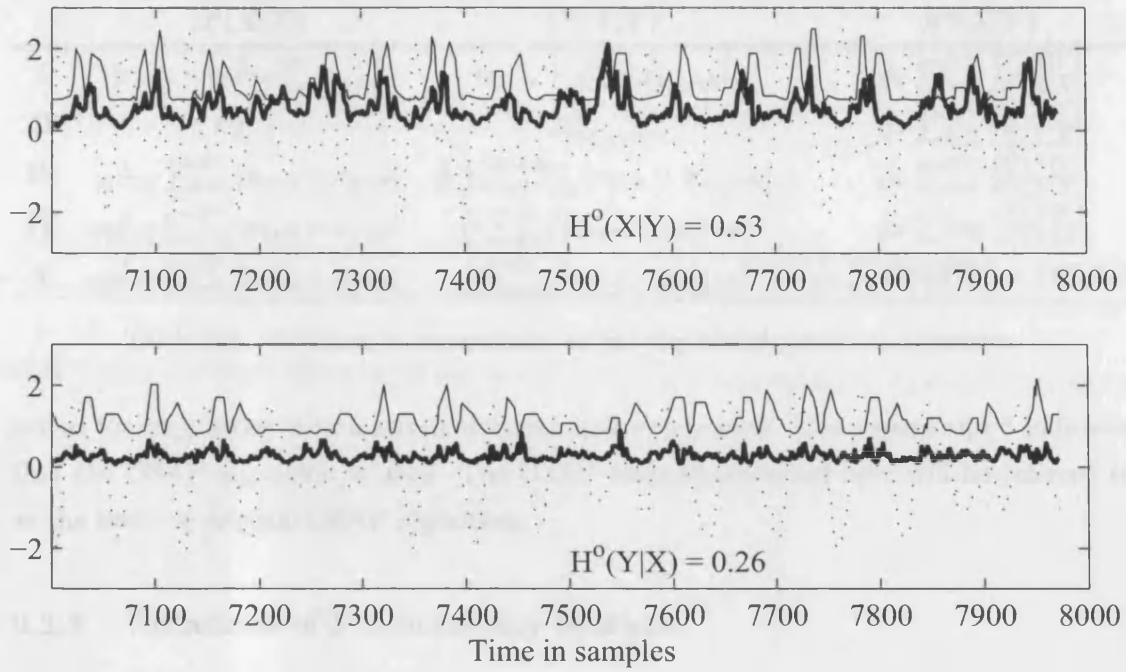


Figure 6.9: Construction and value of the interdependence measure for time trends of the case study. Upper plot: dotted line = X, solid line = $D_i^o(X|X)$, thick line = $D_i^o(X|Y)$. Lower plot: dotted line = Y, solid line = $D_i^o(Y|Y)$, thick line = $D_i^o(Y|X)$. H^o is larger if solid line is well above the thick line. X = TI4, Y = TI5, $N = 1000$, $m = 4$, $h = 10$, $\kappa = 6$, $K = 20$.

Prediction horizon: If the time lag between two time series is known by means such as the cross-correlation function, the prediction horizon is set to the detected time lag. If this is done then the nearest neighbours method is in a way the replacement of the linear correlation coefficient of the algorithm described in Section 5.2. However, if the prediction horizon is not available or uncertain, the nearest neighbours method will give a useful and more robust result due to the embedding. An optimisation of the prediction horizon in case that the time delay is not available is given below.

Case Study: *The OSAP algorithm is applied to the same time trends as the embedding prediction, that is TI4 and TI5 of the case study from Section 4.2 but with the prediction horizon set to 10. Figure 6.9 shows the time trend of TI4 and TI5 together with the OSAP values D_i^o. In the upper plot, $D_i^o(X|Y)$ is on average smaller than $D_i^o(X|X)$ but not considerably which means that Y can predict X to some extent but is not a very good predictor. In the lower plot, however, $D_i^o(Y|X)$ is very small while $D_i^o(Y|Y)$ is between 1 and 2 for all times i. Thus, a very strong dependence from X to Y exists. This is reflected in the values for the interdependence measures $H^o(X|Y) = 0.53$ which twice as big as $H^o(Y|X) = 0.26$. Thus, the measure clearly detects a dependence from $X = TI4$ to $Y = TI5$ rather than vice versa. If X would be a perfect predictor of Y than $D_i^o(Y|X)$ would be constantly zero.* ———————————————

In the following only the OSAP method will be pursued since the algorithm is better

	$D_i^o(X\|X)$	$D_i^o(X\|Y)$	$H^o(X\|Y)$
I	$\lvert x_{i+h} - \frac{1}{N^*-1}\sum x_{j+h}\rvert$	$\lvert x_{i+h} - \frac{1}{K}\sum x_{s_{i,j}+h}\rvert$	$\frac{1}{N^*}\sum_{i=1}^{N^*}\frac{D_i^o(X\|Y)}{D_i^o(X\|X)}$
II	$\sigma^2_{x_{i+h}}$	$\sigma^2_{x_{s_{i,j}+h}}$	$\frac{1}{N^*}\sum_{i=1}^{N^*}\frac{D_i^o(X\|Y)}{D_i^o(X\|X)}$
III	$\frac{1}{N^*-1}\sum_{i=1}^{N^*}\lvert x_{i+h} - x_{j+h}\rvert$	$\frac{1}{K}\sum_{j=1+K_1}^{K+K_1}\lvert x_{i+h} - x_{s_{i,j}+h}\rvert$	$\frac{1}{N^*}\sum_{i=1}^{N^*}\frac{D_i^o(X\|Y)}{D_i^o(X\|X)}$
IV	$\frac{1}{N^*-1}\sum_{i=1}^{N^*}\lvert x_{i+h} - x_{j+h}\rvert$	$\frac{1}{K}\sum_{j=1}^{K}\lvert x_{i+h} - x_{s_{i,j}+h}\rvert$	$\frac{1}{N^*}\sum\log\frac{D_i^o(X\|Y)}{D_i^o(X\|X)}$
V	$\frac{1}{N^*-1}\sum_{i=1}^{N^*}\lvert x_{i+h} - x_{j+h}\rvert$	$\frac{1}{K}\sum_{j=1}^{K}\lvert x_{i+h} - x_{s_{i,j}+h}\rvert$	$\frac{1}{N^*}\sum D_i^o(X\|Y) - D_i^o(X\|X)$

Table 6.1: Variations in computation of one-step-ahead-prediction algorithm.

suited for time series with a strong deterministic component. The superscript ^{o} indicates that the OSAP algorithm is used. The OSAP method described here will be referred to as the basic or original OSAP algorithm.

6.3.3 Variations of Predictability Statistic

The variations in the statistic calculation of D_i and H are numerous and documented in the literature. For the one step ahead prediction, not as many variations have been tested so far and not in a systematic manner. In this section, the impact of five variations will be tested on process data. The alternative statistics are listed in Table 6.1. Variation I to III concern D_i^o while variation IV and V are alternatives for computing H^o. The variations are referred to as:

- Robustness for outliers (I);
- Variance instead of average (II);
- Nearest neighbours outside K_1 (III);
- Added Logarithm (IV);
- Difference instead of division (V).

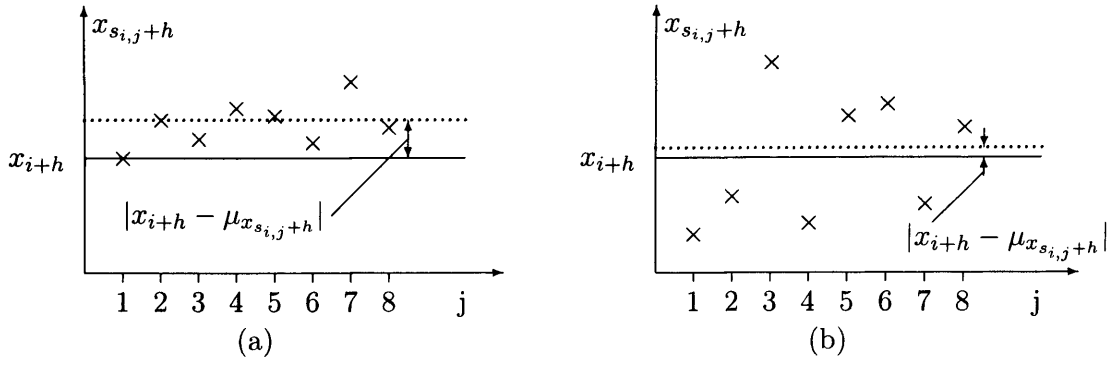


Figure 6.10: Constructional drawback of Variation I: good prediction (a) gives a larger value for the statistic $D_i^o(X|Y)$ than poor prediction (b).

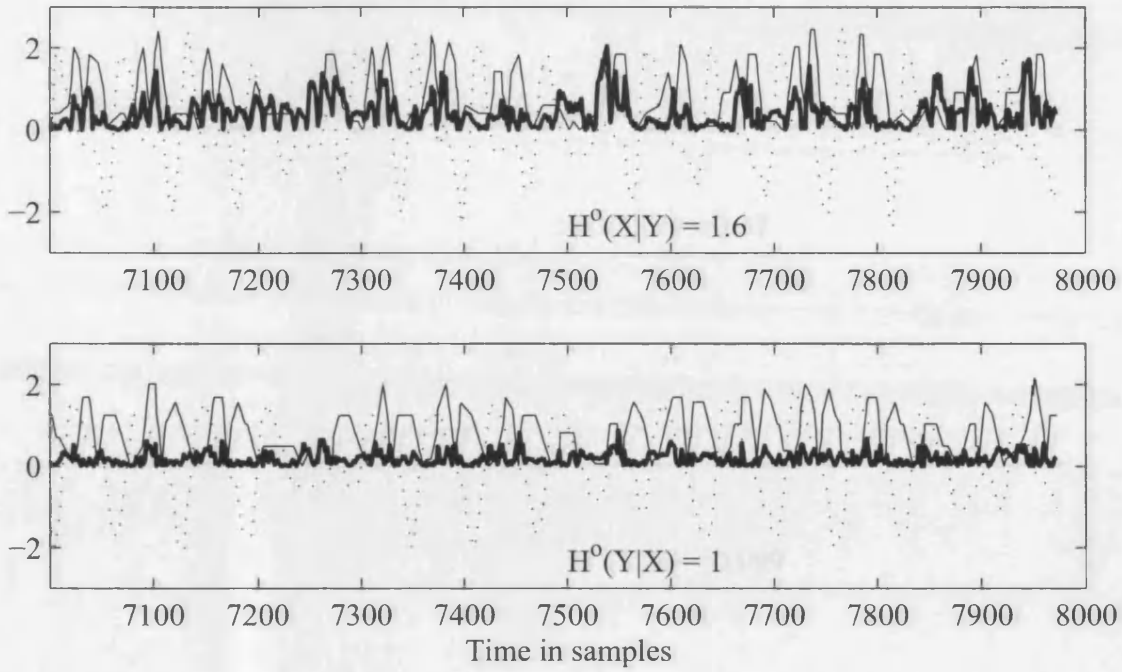


Figure 6.11: Construction of nearest neighbours statistic, Variation I. Upper plot: dotted line = X, solid line = $D_i^o(X|X)$, thick line = $D_i^o(X|Y)$; lower plot: dotted line = Y, solid line = $D_i^o(Y|Y)$, thick line = $D_i^o(Y|X)$. X = TI4, Y = TI5, $N = 1000$, $m = 4$, $h = 10$, $\kappa = 6$, $K = 20$.

Variation I: Robustness for Outliers The averaging of the distance when computing $D_i^o = \frac{1}{N^*-1} \sum_{j=1}^{N^*} |x_{i+h} - x_{j+h}|$ can result in misleading estimates if the estimation is distorted by outliers. Schiff et al. [110] therefore calculate the distance between x_{i+h} and the mean value of the nearest neighbours instead of summing the distance between x_{i+h} and each predicted nearest neighbour. This has the advantage of robustness since outliers have not as big an impact. A tradeoff when taking the distance to the average is that the result can be distorted for larger number of nearest neighbours. This effect is shown in Figure 6.10. Both plots show predictions of the value indicated by a solid line, the dotted line indicates the average of the nearest neighbours. The distance between the solid line and the dotted line is the predication measure of Variation I. The estimated values in the left hand plot are in fact the better predictions but the average distance gives a misleading result. Figure 6.11 shows the trend of D_i^o for the example of TI4 and TI5 from the case study. Comparing the trends of D_i^o with the same trends for the basic OSAP algorithm in Figure 6.9 one can notice that the trends of Variation I are noisier than the trends of the basic algorithm. While $D_i^o(Y|X)$ is always well below $D_i^o(Y|Y)$ in Figure 6.9 this is not the case for Variation I. On the other hand, the resulting interdependency measures H^o indicate the causality more distinctively for Variation I where $H^o(X|Y) = 1.6$ is larger than $H^o(Y|X) = 1.0$.

Variation II: Variance Instead of Average Both the original OSAP algorithm as well as Variation I take the mean of the distance of the nearest neighbours. In Variation II, the mean is replaced by the variance of the distances, arguing that a good predictor will

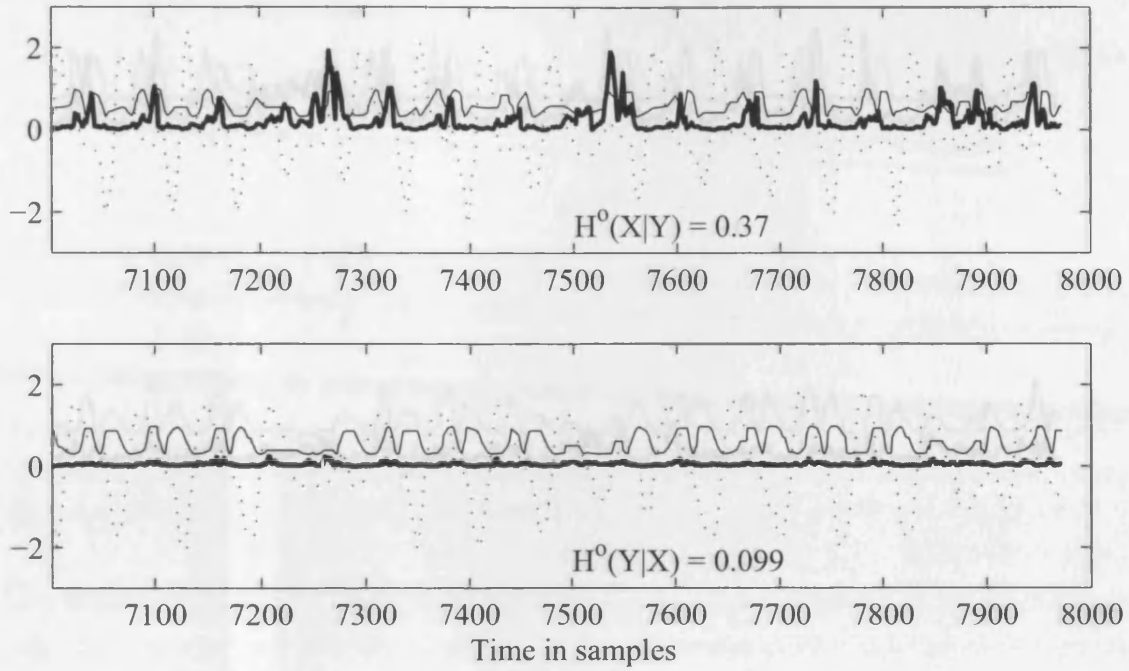


Figure 6.12: Construction of nearest neighbours statistic, Variation II. Upper plot: dotted line = X, solid line = $D_i^o(X|X)$, thick line = $D_i^o(X|Y)$; lower plot: dotted line = Y, solid line = $D_i^o(Y|Y)$, thick line = $D_i^o(Y|X)$.X = TI4, Y = TI5, $N = 1000$, $m = 4$, $h = 10$, $\kappa = 6$, $K = 20$.

not only estimate a future value correctly but will do so with small variations. By taking the variance, the impact of outliers is reduced without the drawback of false prediction that occurred for Variation I. The idea of using variance rather than the mean has to the author's knowledge not previously been applied to the OSAP algorithm. There are, however, parallel variations of the delta-epsilon method for one variable by Gautama et al. [33], the delay vector variance, and also for the embedding prediction by Goodridge et al. [37]. Figure 6.12 shows the trends of D_i^o. Here, $D_i^o(Y|X)$ is smaller than $D_i^o(Y|Y)$ for all times. This is not, as desired, the case for $D_i^o(X|Y)$ which is at some points much larger than $D_i^o(X|X)$. Unlike for the original method and Variation I, $D_i^o(X|Y)$ and $D_i^o(Y|X)$ show some periodic behaviour in phase with the periodicity of the time trend. The interdependence measure $H^o(X|Y)$ is larger than $H^o(Y|X)$ by a factor of 3.6. This shows at a first glance a stronger causality than the original method and Variation I.

Variation III: Nearest Neighbours Outside K_1 So far, K nearest neighbours have been considered for the averaging or variance calculation of D_i^o. There might, however, be the chance that the nearest neighbours are due to random effects rather than coupling if the noise-to-signal ratio is very high. To overcome this problem, the K_1 nearest neighbours are omitted and all further K nearest neighbours are considered. This approach was proposed by Bhattacharya et al. [6] for the case of embedding prediction. The time series investigated in the case study, however, generally have a strong deterministic part and a low noise level. Figure 6.13 shows the trend of D_i^o, K_1 is chosen equal to K. The trend is almost identical to the trend of the original method in Figure 6.9. As expected since

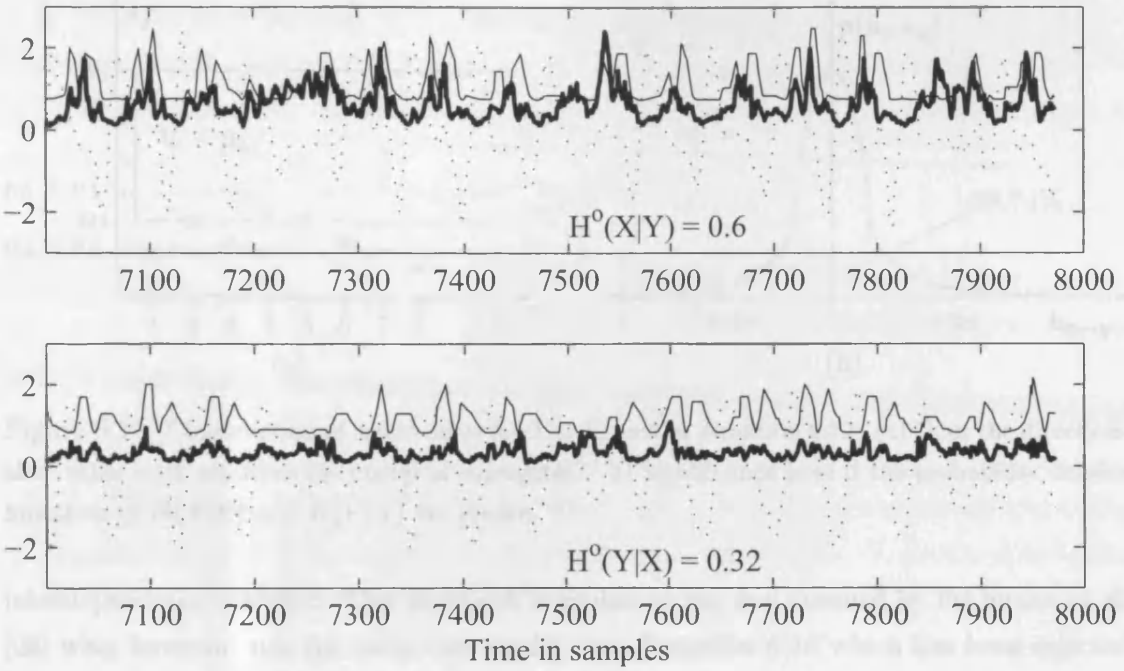


Figure 6.13: Construction of nearest neighbours statistic, Variation III. Upper plot: dotted line = X, solid line = $D_i^o(X|X)$, thick line = $D_i^o(X|Y)$; lower plot: dotted line = Y, solid line = $D_i^o(Y|Y)$, thick line = $D_i^o(Y|X)$. X = TI4, Y = TI5, $N = 1000$, $m = 4$, $h = 10$, $\kappa = 6$, $K = K_1 = 20$.

the signal has a low noise component, the impact of excluding the K_1 nearest neighbours is only marginal. The interdependence measures are in the range of the measure for the original method: $H^o(X|Y) = 0.60$ (original method: 0.53) and $H^o(Y|X) = 0.32$ (original method: 0.26).

Variation IV: Logarithm Variations I to III involves changes in the computation of D_i^o. In Variation IV, on the other hand, the summation for H^o is changed to summing the logarithm of $D_i^o(X|Y)/D_i^o(X|X)$. The argument of using the logarithm is that the asymmetry is more pronounced compared to using a direct summation. The interpretation of the results remains unchanged since the logarithm is a monotonically increasing function. However, H^o will now be negative so that large negative values indicate interdependence while values close to or larger than zero indicate independence. The final directionality measure $h_{x\to y} = H^o(X|Y) - H^o(Y|X)$ is again positive for x influencing y. The use of the logarithm in the summation has been tested in [2, 100, 6, 114] and is similar to the structure of transfer entropy, see next chapter. The value of $H^o(TI4|TI5)$ with identical parameter setting as chosen for Variation I to III is -0.79 while the opposite case is much smaller, $H^o(TI4|TI5) = -1.5$.

Variation V: Difference Instead of Division Using logarithmic function in fact can be interpreted as taking the difference of $D_i^o(X|Y)$ and $D_i^o(X|X)$ since $\log \frac{a}{b} = \log a - \log b$. Thus, the difference without taking the logarithm is a viable option when computing the

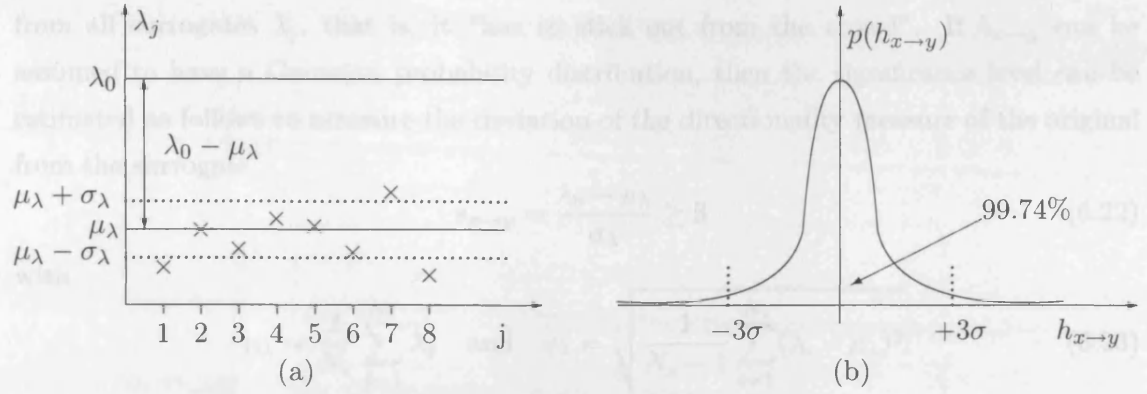


Figure 6.14: Construction of significance level as defined in Equation 6.22: (a) Does the directionality value stick out from the crowd of surrogates? (b) Significance level if the probability density functions of $H(X|Y)$ and $H(Y|X)$ are known.

interdependence statistic. This approach is similar to the one pursued by Feldmann et al. [28] who, however, use the definition for D_i^o from Equation 6.16 which has been rejected here because of its strong dependence on the number of nearest neighbours, see argument above. Again, the value of $H^o(TI4|TI5)$ with same parameter setting as before is negative and -0.54 while the opposite case is also negative but much smaller, $H^o(TI4|TI5) = -0.86$.

No recommendations will be made at this point. The five proposed variations have their right of existence as argued in the paragraphs above. A comparison of the five variations is carried out below.

6.3.4 Significance Level

For comparing the different implementation of the nearest neighbours one step ahead prediction algorithm (OSAP), a measure is needed for evaluating the performance of the algorithm. And maybe more importantly, once the algorithm is specified, a threshold should be established that defines above which level the detected directionality is a valid and not only coincidence. For the latter purpose, Theiler et al. [130] suggest Monte Carlo methods to establish a significance or confidence level. Monte Carlo methods are any methods that solve a problem by generating suitable random numbers or sequences and observing that a fraction of the random numbers obey some property[6]. The problem is to verify or reject a null hypothesis. Here, in the case of the directionality estimation, the null hypothesis is that x does not influence y. The null hypothesis is denoted by $\lambda_0 = h_{x\to y}$. Suitable random numbers are values of the directionality measure that are generated by computing the measure from surrogate time series x_{surr} and y_{surr}, that is $h_{x_{\text{surr}}\to y_{\text{surr}}} =: h^{\text{surr}}_{x\to y}$. The method of surrogates is briefly described in Section 6.2.3, for details see [114]. Altogether, N_s surrogate measures are constructed and denoted by $\lambda_j = h^{\text{surr},j}_{x\to y}$ with $j = 1 \ldots N_s$. The null hypothesis is now rejected if λ_0 varies considerably

[6]See definition on http://mathworld.wolfram.com/Monte-Carlo-Method.html, November 2004.

from all surrogates λ_j, that is, it "has to stick out from the crowd". If $h_{x\to y}$ can be assumed to have a Gaussian probability distribution, then the significance level can be estimated as follows to measure the deviation of the directionality measure of the original from the surrogate,

$$s_{x\to y} = \frac{\lambda_0 - \mu_\lambda}{\sigma_\lambda} \geq 3 \tag{6.22}$$

with

$$\mu_\lambda = \frac{1}{N_s}\sum_{j=1}^{N_s}\lambda_j \quad \text{and} \quad \sigma_\lambda = \sqrt{\frac{1}{N_s - 1}\sum_{i=1}^{N_s}(\lambda_i - \mu_\lambda)^2}. \tag{6.23}$$

The principle of significance level $s_{x\to y}$ was suggested by Theiler et al. and is illustrated in Figure 6.14. The threshold of the significance above which a measurement is considered as valid is 3 for a two sided test in Theiler et al., that is, the distance of the directionality measure to the mean of the measure for the surrogates is three times the standard deviation of the surrogates. Assuming a Gaussian distribution, this would be equivalent to a 99.74% certainty. A three sigma test is still valid even the distribution is not Gaussian but the 99.74% significance level cannot be defined.

In case that probability density functions of $H(X|Y)$ and $H(Y|X)$ and their joint probability density function are accessible, the probability density function of $h_{x\to y}$ can be deduced,

$$p(h_{x\to y}) = \frac{\mathrm{d}}{\mathrm{d}h_{x\to y}}\int_{-\infty}^{\infty}\int_{-\infty}^{h_{x\to y}+H(Y|X)} p(H(X|Y), H(Y|X))\mathrm{d}H(X|Y)\mathrm{d}H(Y|X) \tag{6.24}$$

The significance level can then be conveniently defined by a threshold value of $h_{x\to y}$ as illustrated in Figure 6.14 (b). The problem is that even though the probability density functions might be available for some reference signals, the joint probability will be difficult to derive. One cannot assume that $H(X|Y)$ and $H(Y|X)$ are independent as the two values tend to lie in the same range or are even equal in most cases of independency of x and y.

If the significance level in Equation 6.22 cannot be adopted because the measure does not follow a Gaussian distribution, a rank test can be considered instead to accept or reject a directionality value. It does however not provide a good means of comparison for a variation of methods. In a rank test, the directionality measure is computed for N_s surrogates λ_j [57]. The probability that the original signal has the largest value is $\alpha = \frac{1}{N_s+1}$; if 19 surrogates are considered then the chances are α=5%. The directionality value can be accepted with a 95% certainty if it is the largest among all surrogates, that is, $\lambda_0 > \lambda_j$ for all $j = 1 \ldots N_s$.

In the following, $h_{x\to y}$ will be considered as Gaussian distributed and both significance level and rank test will be carried out to compare the variations of the method.

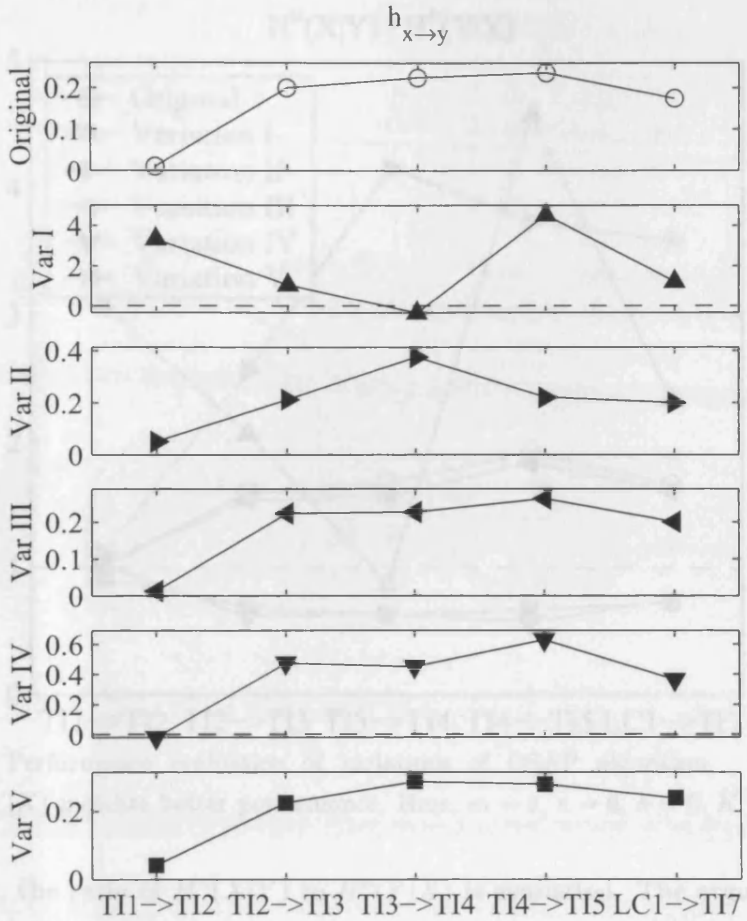


Figure 6.15: Performance evaluation of variations of OSAP algorithm. For correct identification of directionality $H^o(X|Y)$ should be larger than $H^o(Y|X)$ where --- represents $H^o(X|Y)$ and -*- represents $H^o(Y|X)$; $m = 4$, $\kappa = 6$, $h = 6$, $K = 20$, $N = 1000$.

6.3.5 Selecting the Best One-Step-Ahead-Prediction Variation

The five variations of the original OSAP method are compared using the data trends from the case study in Section 4.2. The relationships selected are the temperature measurements along the distillation column, TI1 to TI5, that is whether TI1 influences TI2, TI2 influences TI3 etc. The relationship between LC1 and TI7, with LC1 being further upstream and closer to the root cause, is also included because both show similarities in the time trend. Thus, the direction of dependency is considered as known for these five relationships:

$$\text{TI1}\rightarrow\text{TI2}; \quad \text{TI2}\rightarrow\text{TI3}; \quad \text{TI3}\rightarrow\text{TI4}; \quad \text{TI4}\rightarrow\text{TI5}; \quad \text{LC1}\rightarrow\text{TI7}.$$

The results for H^o using the OSAP algorithm and Variation I to V are shown in Figure 6.15. The parameters are chosen as before with $N = 1000$, $m = 4$, $h = 6$, $\kappa = 6$, $K = K_1 = 20$. All methods give good results in the sense that the direction of dependency is measured correctly, that is, $H^o(X|Y) > H^o(Y|X)$ with the only exception of Variation I where the relationship between TI3 and TI4 is detected incorrectly. The representation in Figure 6.15 is, however, not suited for comparing the algorithms since the maximum and minimum values differ for each method.

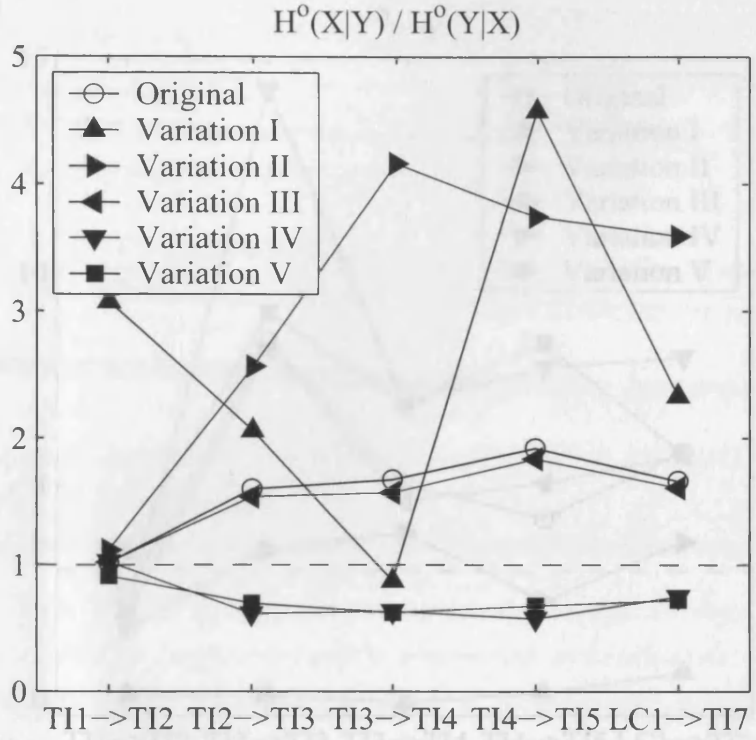


Figure 6.16: Performance evaluation of variations of OSAP algorithm. Higher ratios of $H^o(X|Y)/H^o(Y|X)$ indicate better performance. Here, $m = 4$, $\kappa = 6$, $h = 6$, $K = 20$, $N = 1000$.

As a first shot, the ratio of $H^o(X|Y)$ to $H^o(Y|X)$ is evaluated. The argument for this is that the larger the difference between the two interdependence measures the better the estimation of the direction. Figure 6.16 shows the ratio of $H^o(X|Y)$ over $H^o(Y|X)$. The ratios of Variation III to V lie in the same range as the rations of the original algorithm while Variation I and II show significant larger values. A suspicion that the ratio is not an accurate mean of comparison is raised by the circumstance that Variation I does not detect the relationship between TI3 and TI4. The problem with the comparison of the ratio is that not only the maximum and minimum value of algorithm variations I to V are different but also the standard deviations of the measure.

The significance level constructed from surrogate data as described in Equation 6.22 overcomes this problem by only assuming a Gaussian distribution of the directionality measure and scaling to standard deviation of the measure. The significance level of the relationships of the case study are shown in Figure 6.17. The threshold value of 3 that indicates a deviation larger than 3σ is shown by a dashed line. Here, Variation I and II show a very poor performance while Variation IV and V are clearly better performing than the original algorithm. Variation IV has an overall higher value of $s_{x\to y}$ but the attractiveness of Variation V lies in the accurate detection of the relationship between TI1 and TI2. TI1 and TI2 are very similar and a dependency is therefore more difficult to detect. The decision to chose Variation V over Variation IV is also supported by the rank test. Table 6.2 summarises the results of the rank test, the numbers indicating the rank of the original value of $h_{x\to y}$ among the surrogates. Although Variation IV does not detect the TI1 and

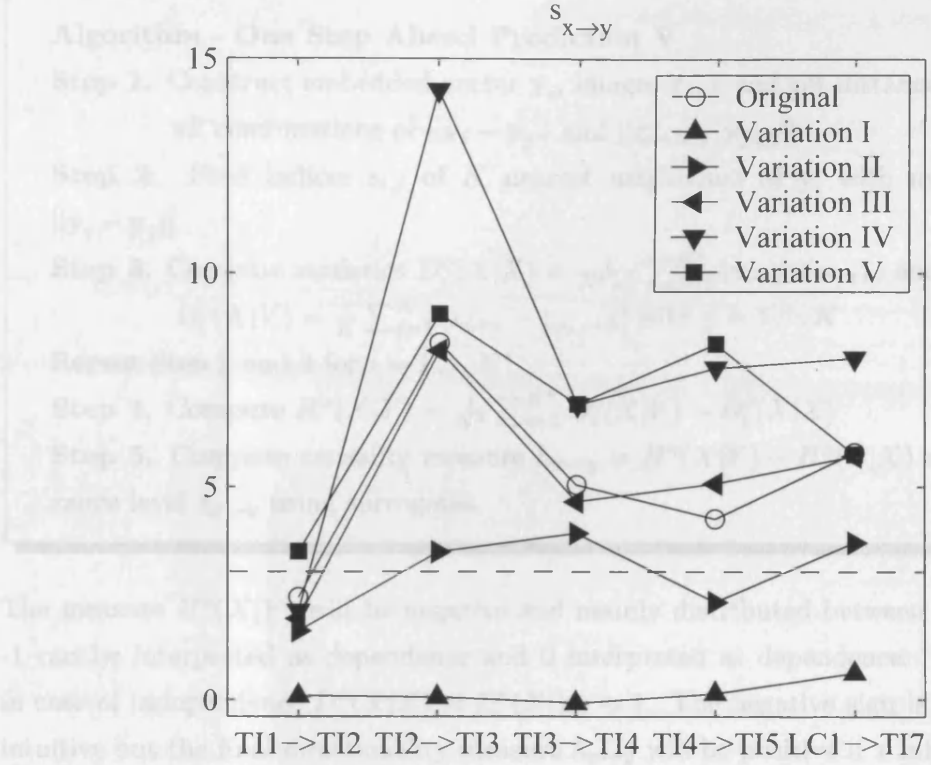


Figure 6.17: Performance evaluation of variations of OSAP algorithm: significance level $s_{x\to y}$. Higher values indicate better performance. Here, $m = 4$, $\kappa = 6$, $h = 6$, $K = 20$, $N = 1000$.

TI2 relationship as the highest rank, it is the second highest which outperforms Variation V for which it is only the seventh highest value. Variation I is performing poorly also in the rank test.

6.3.6 Best One-Step-Ahead-Prediction Algorithm

As a result of the previous analysis, the nearest neighbours algorithm with one step ahead prediction (OSAP) using the difference between $D_i^o(X|X)$ and $D_i^o(X|Y)$ of the prediction is chosen for future work since it showed best results when applied to the case study (see previous section). The algorithm can then be summarised as follows.

$\lambda_0 \leqq \lambda_i$	Original	Var I	Var II	Var III	Var IV	Var V
TI1 → TI2	7	9	7	6	7	2
TI2 → TI3	1	4	1	1	1	1
TI3 → TI4	1	4	1	1	1	1
TI4 → TI5	1	6	1	1	1	1
LC1 → TI7	1	8	1	1	1	1

Table 6.2: Rank test: rank of λ_0 among λ_i, $i = 1 \ldots N_s$ in terms of largest value. Here, $m = 4$, $\kappa = 6$, $h = 6$, $K = 20$, $N = 1000$.

Algorithm - One Step Ahead Prediction V

Step 1. Construct embedded vector $\mathbf{y}_i$, images x_{i+h} and get distances of all combinations of $||\mathbf{y}_i - \mathbf{y}_j||$ and $||x_{i+h} - x_{j+h}||$

Step 2. Find indices $s_{i,j}$ of K nearest neighbours of $\mathbf{y}_i$ with minimum of $||\mathbf{y}_i - \mathbf{y}_j||$.

Step 3. Compute statistics $D_i^o(X|X) = \frac{1}{N^*-1} \sum_{j=1}^{N^*} |x_{i+h} - x_{j+h}|$ and $D_i^o(X|Y) = \frac{1}{K} \sum_{j=1}^{K} |x_{i+h} - x_{s_{i,j}+h}|$ with $j = 1 \ldots K$

Repeat Step 2 and 3 for $i = 1 \ldots N^*$

Step 4. Compute $H^o(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} D_i^o(X|Y) - D_i^o(X|X)$

Step 5. Compute causality measure $h_{x \to y} = H^o(X|Y) - H^o(Y|X)$ and significance level $s_{x \to y}$ using surrogates.

The measure $H^o(X|Y)$ will be negative and mainly distributed between -1 and 0 where -1 can be interpreted as dependence and 0 interpreted as dependence. This is because in case of independence $D_i^o(X|X) \approx D_i^o(X|Y) \approx 1$. The negative sign is not particulary intuitive but the final directionality measure $h_{x \to y}$ will be positive if x influences y which is the measure of interest.

Computational Effort: The number of computations required for the described algorithm is given in Table 6.3. The majority of computations are summations or differences which depend most strongly on N^{*2} and on the embedding dimension m. In the algorithm, $N^* = N-(m-1)\kappa$ is the number of embedded vectors. Simplifying the dependency results in a dependency of the computational effort on $N^{*2}m$. The number of nearest neighbours only has a marginal impact on the total number of computations since it has to be by definition a fraction of N^*. Implementation with vector oriented software such as Matlab allows the computation to be carried out efficiently. On a Pentium IV PC, the computation for $N = 1000$, thus $N^* \approx 1000$ and an embedding dimension $m < 5$ the computation of $h_{x \to y}$ requires a few seconds. Longer computation time however are experienced for the calculation of the significance level which requires the computation to be carried out N_s times. With $N_s = 20$ as a minimum number, the computation of the significance level takes a minute or two, again on a Pentium IV processor.

	$a \pm b$	$(a)^2$	$\sqrt{a}$	$\lvert a \rvert$
$\lvert\lvert \mathbf{y}_i - \mathbf{y}_j \rvert\rvert$	$\frac{m}{2}N^*(N^*-1)$	$\frac{m}{2}N^*(N^*-1)$	$\frac{1}{2}N^*(N^*-1)$	-
$\lvert x_{i+h} - x_{j+h} \rvert$	$\frac{1}{2}N^*(N^*-1)$	-	-	$\frac{1}{2}N^*(N^*-1)$
$D_i^o(X\|X)$	N^{*2}	-	-	N^{*2}
$D_i^o(X\|Y)$	N^*K	-	-	N^*K
$H^o(X\|Y)$	$2N^*$	-	-	-

Table 6.3: Number of computations for nearest neighbours one-step-ahead-prediction algorithm, variation V.

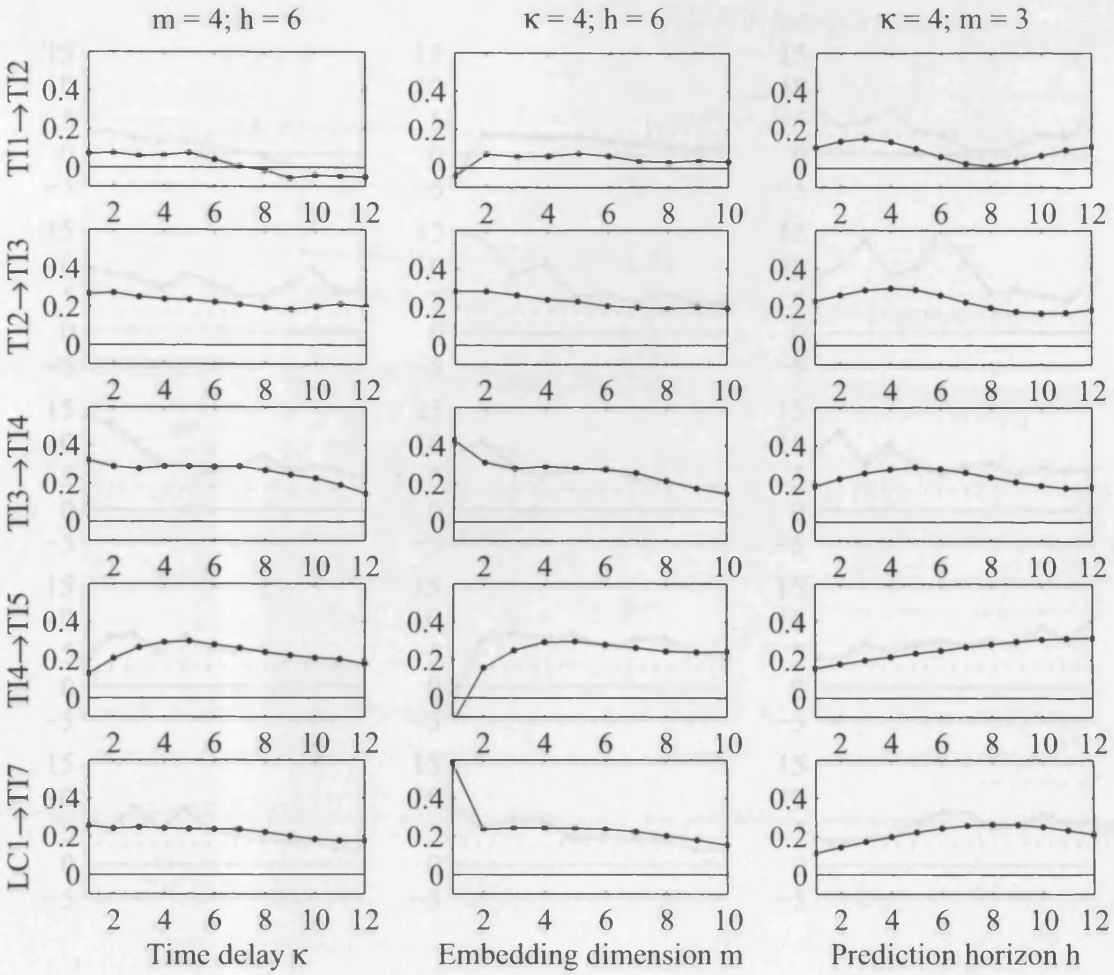


Figure 6.18: Parameter optimisation through maximising $h_{x\to y} = H^o(X|Y) - H^o(Y|X)$ for embedding dimension, time lag and prediction horizon for validation that the directionality is identified correctly. Here, $K = 20$, $N = 1000$.

6.3.7 Parameter Optimisation

Apart from variations in calculating the statistics D_i^o and H^o, parameters of the algorithm can be adjusted to detect the coupling strength with an optimised value. The parameters that are used for the nearest neighbours method are:

- m Embedding dimension,
- κ Time lag between embedding,
- h Prediction horizon,
- K Number of nearest neighbours,
- N Number of samples.

Both time lag κ and prediction horizon h depend on the sampling rate. The sampling interval of the reference case study from Section 4.2 is 20 seconds and treated as a reference guide. Time lag and prediction horizon have to be adjusted when dealing with other data sets. The number of samples N and number of nearest neighbours K are not expected to have a large impact on the directionality measure above certain thresholds since they are

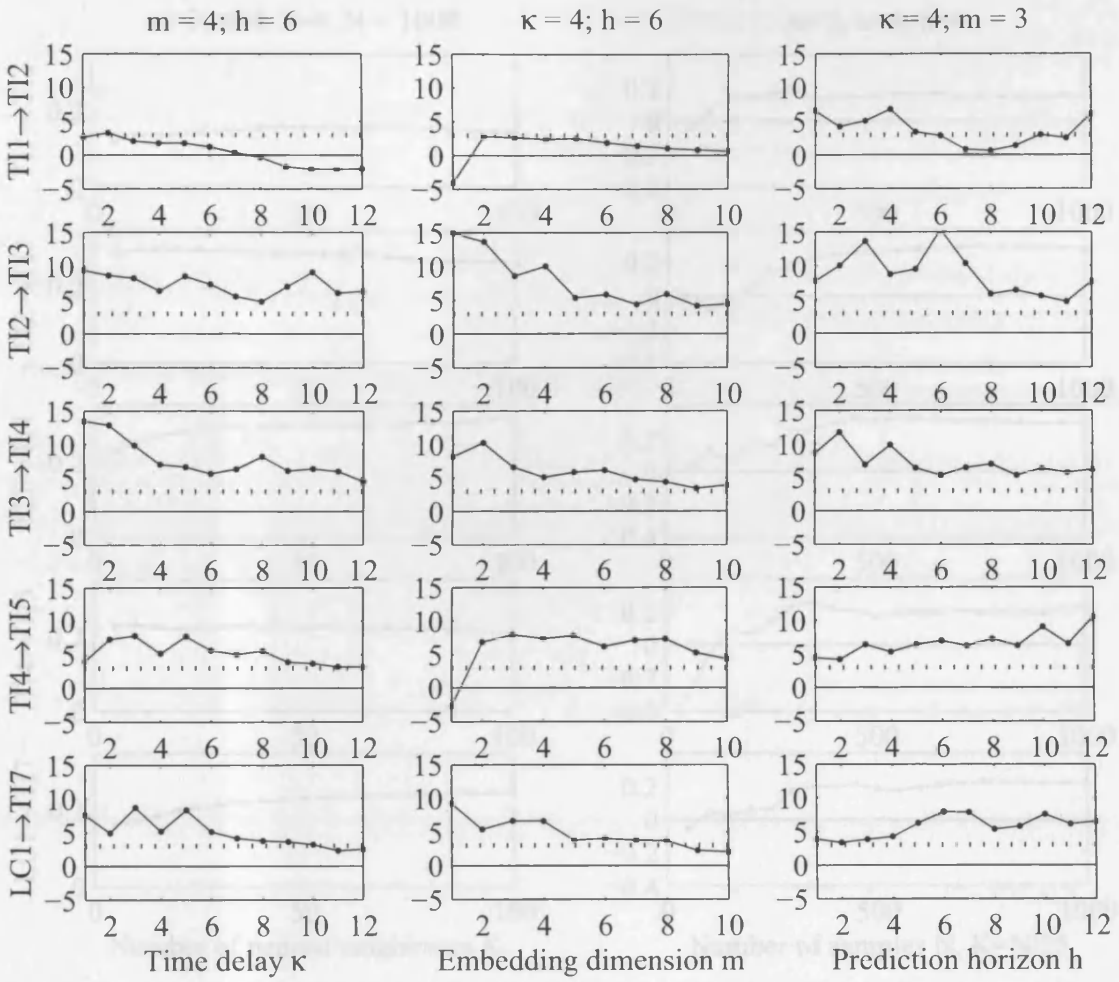


Figure 6.19: Parameter optimisation through maximising $s_{x\to y}$ for embedding dimension, time lag and prediction horizon. Here, $K = 20$, $N = 1000$.

only the number of statistical averages. Therefore, embedding dimension, time lag and prediction horizon are optimised for fixed length and number of nearest neighbours. The optimisation is done using the examples of the case study for which the directionality is known, as done in the previous section. Thus, the five relationships

$$\text{TI1}\to\text{TI2};\quad \text{TI2}\to\text{TI3};\quad \text{TI3}\to\text{TI4};\quad \text{TI4}\to\text{TI5};\quad \text{LC1}\to\text{TI7}.$$

are investigated for reference. The optimisation is conducted in three steps.

- First, embedding dimension m and prediction horizon h are fixed while the time delay κ is optimised.
- In the second step, prediction horizon h and the optimised time delay κ are fixed while the embedding dimension m is optimised.
- In the third step, the prediction horizon h is varied while h and m are set to their optimised values.

The aim of the optimisation is to find the highest significance level while ensuring that the causality measure $h_{x\to y}$ is positive. Both the directionality measure and the significance

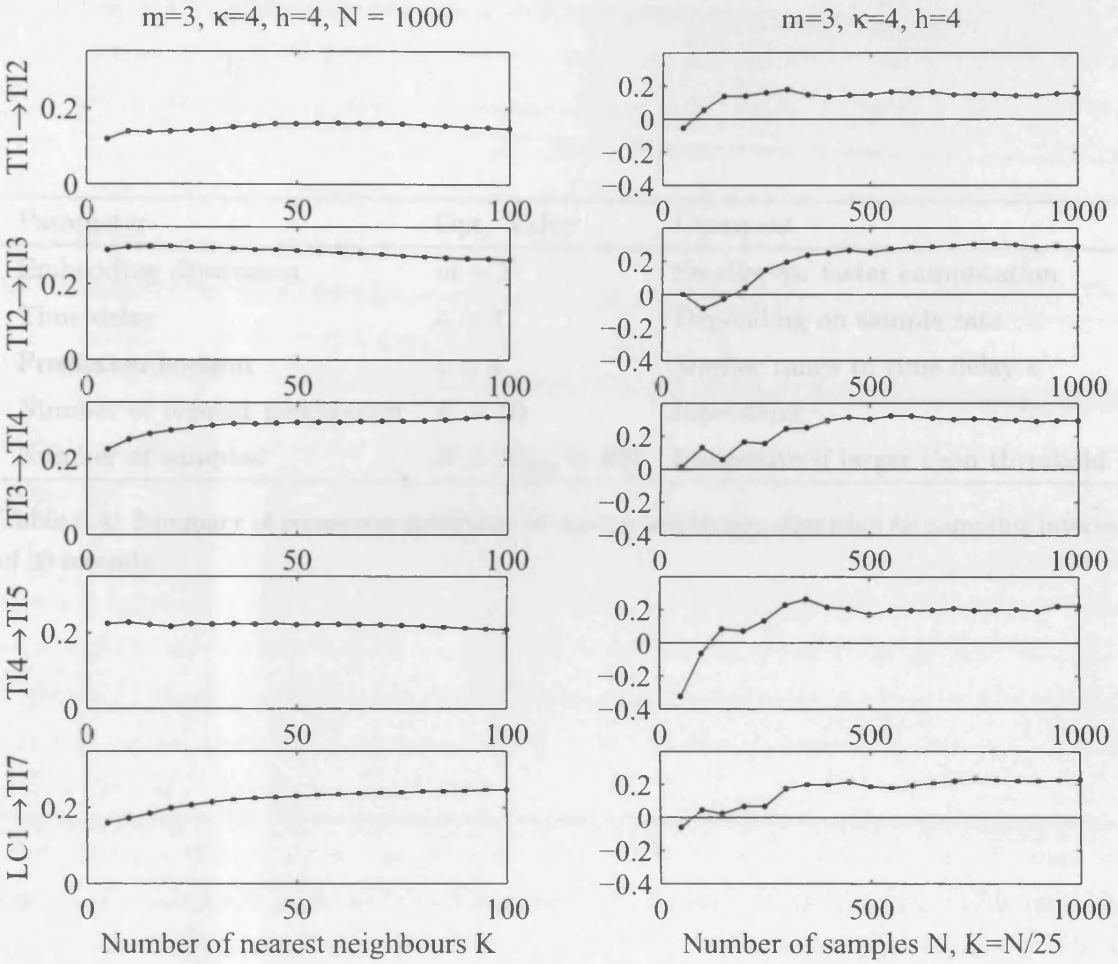


Figure 6.20: Parameter optimisation by maximising $h_{x\to y}$ for the number of nearest neighbours and the number of samples.

level are therefore considered for optimisation. The results of the procedure are shown in Figures 6.18 and 6.19. The left hand column shows $h_{x\to y}$ for the first step. Time delay κ is adjusted while m and h are fixed. The tendency of the measure for the five investigated relationship is that $h_x\to y$ decreases with increasing time delay. An exception is the relationship between TI4 and TI5 which initially increases. The corresponding significance level is shown in the left hand column of Figure 6.19. The trends are similar to the directionality measure but the decreasing and increasing trends are noisier. It is suspected that this is due to the randomisation of the surrogates.

Optimising κ in the sense that the sum of all relationships TI1 to TI7 is maximum for the $h_{x\to y}$ results in a value of $\kappa = 4$. Similar considerations are undertaken for embedding dimension m which is shown in the column in the middle of Figure 6.18 and 6.19 for which κ is set to 4 and h fixed to 6. The optimum value is then $m = 3$. Prediction horizon h is optimised in the right hand columns to $h = \kappa = 4$.

The same optimisation procedure is used for the number of nearest neighbours and the number of samples. The number of nearest neighbours K is thought to have no major

Parameter	Opt. Value	Comment
Embedding dimension	$m = 3$	Smaller for faster computation
Time delay	$\kappa = 4$	Depending on sample rate
Prediction horizon	$h = 4$	Similar range to time delay κ
Number of nearest neighbours	$K = 20$	Insensitive
Number of samples	$N > N_{\min} \approx 400$	Insensitive if larger than threshold

Table 6.4: Summary of parameter guidelines of nearest neighbours algorithm for sampling interval of 20 seconds.

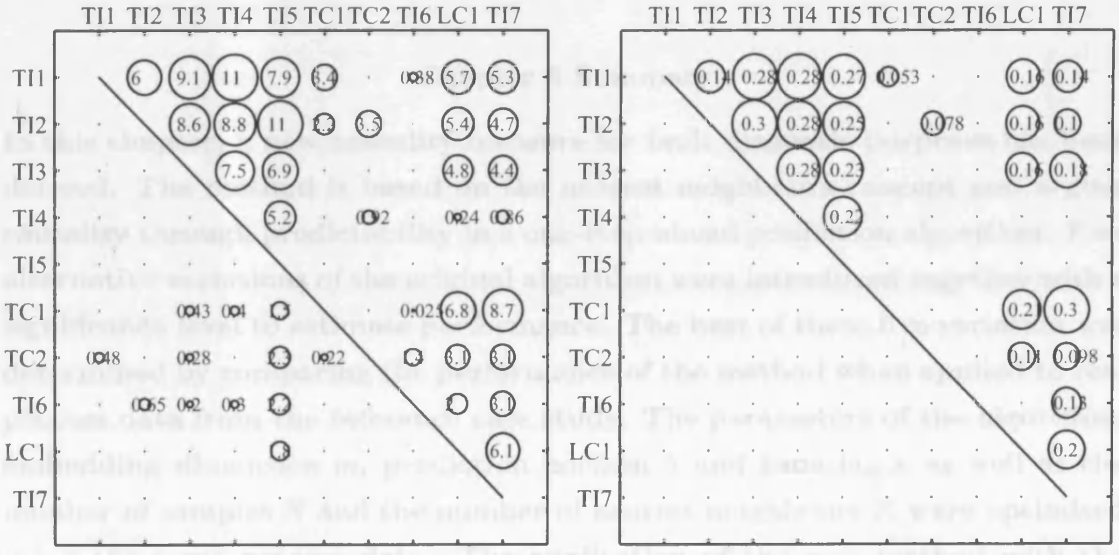


Figure 6.21: Case study results of for significance level $s_{x \to y}$ (left) and causality values $h_{x \to y}$ if significance level exceeds threshold (right); $N_s = 20$, $N = 1000$.

impact on the directionality measure due to the averaging nature of D_i^o and H^o. The left hand column of Figure 6.20 shows that this is true for the sample trends of the case study. K is also linked to the number of samples N and it is useful to define K as a fraction of N. The fraction of K/N is kept constant for investigating the impact of the number of samples on the directionality measure. It is expected that the measure is independent of N at least for $N > N_{\min}$ where $N_{\min}$ is a minimum number of samples required for the computation. The right hand column of Figure 6.20 confirms this expectation. The directionality measure varies significantly until $N_{\min} \approx 400$.

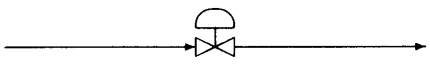
The parameters are optimised for Variation V of the OSAP algorithm since this is the algorithm selected for future use. A summary of the values of all parameters is given in Table 6.4.

Case Study: *The proposed algorithm including the guidelines for the parameter estimation are applied to the full reference case study from Section 4.2. The directionality measure and the significance level are computed for all combinations of the process variables. Figure 6.21 shows the bubble chart. All relationships that are detected are consistent with the root cause hypothesis. This can be seen from the fact that all $h_{x\to y}$ and $s_{x\to y} > 3$ lie above the main diagonal.*

The efficiency of the nearest neighbours method will be investigated in Part III and the results of the investigation will be discussed and summarised in Chapter 11. A comparison to the CCF method introduced in Chapter 5 and to transfer entropy which will be introduced in the chapter will be made.

Chapter 6 Summary

In this chapter, a new causality measure for fault diagnosis purposes has been derived. The method is based on the nearest neighbours concept and argues causality through predictability in a one-step-ahead prediction algorithm. Five alternative variations of the original algorithm were introduced together with a significance level to estimate performance. The best of these five variation was determined by comparing the performance of the method when applied to real process data from the reference case study. The parameters of the algorithm, embedding dimension m, prediction horizon h and time lag κ as well as the number of samples N and the number of nearest neighbours K were optimised using the same process data. The application of the new method with the guideline values for the parameters was successful and the root cause could be identified.



Chapter 7

Transfer Entropy

In this chapter, the method of transfer entropy is proposed for fault diagnosis. Transfer entropy is a statistical tool based on probability density functions that evaluates the predictability of one variable from a second variable. The underlying mechanism is similar to the one used by the nearest neighbours method. The chapter discusses the estimation of the probability density function and the application of transfer entropy to process data. Required parameters are optimised and a significance level established.

Transfer entropy was proposed by Schreiber in 2000 [113] as a recent advancement in the area of information theory. In 1948, Shannon [117] introduced entropy to quantify information in a mathematically defined sense. Further entropy measures, such as mutual information or conditional entropy, investigate the common amount of information contained in two or more signals. Transfer entropy, as its name suggests, quantifies the amount of information transferred from one signal to a second signal. It is an asymmetric measure and can therefore be transformed into a causality measure specifying whether A influences B or B influences A. Successful applications of transfer entropy are to physiological data such as the interdependency between heart and breath rate [113], financial data like the interaction between the American Dow Jones and the German DAX [78] or neurological data [55]. In this chapter, the application of transfer entropy to chemical processes is investigated and applied to the example of the industrial case study from Section 4.2. Parameter settings and implementation issues are discussed.

The chapter is structured as follows. In the first section, the concepts of probability distributions and probability density functions are reviewed. Two alternative estimation methods for the probability density functions, histograms and the Kernel method, are given in Section 7.2. Entropy measures and their application to fault diagnosis are introduced in Section 7.3. Section 7.4 introduces the causality measure of transfer entropy. A significance level is established and the computational effort estimated. The parameters used in the transfer entropy calculation are optimised in Section 7.5

7.1 Probability Distributions

A signal whose trend can be described at any time by a function is called deterministic. Most time signals of measurements from chemical processes are random to some degree and cannot or can only partially be expressed by a deterministic function. All concepts from functional theory for examining the time series, such as the Fourier transform, have shortcomings when investigating a random or stochastic signal because they do not capture the stochastic nature. Thus, a different approach has to be chosen for analysis. Random signals are expressed in terms of probabilities. In this sense, the probability distribution of a random variable is defined as the probability of a random variable X being smaller or equal than a value x.

$$P_X(x) = P(X \leq x) \tag{7.1}$$

Here, P denotes 'probability'. The probability function is monotonically increasing and the limit values are $P_X(x \to -\infty) = 0$, $P_X(x \to +\infty) = 1$. But more importantly, the following statement has to be kept in mind when analysing any measurements from practical applications, such as from chemical processes.

> The probability distribution gives an exact description of the stochastic or random part of a time signal.

Probability Density Function:

The probability density function (PDF) of a random signal is the derivative of the probability distribution.

$$p_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} P_X(x) \tag{7.2}$$

In the following, the subscript X is omitted. Properties of the PDF are $0 \leq p(x) \leq 1$, $p(x \to -\infty) = p(x \to \infty) = 0$ and $\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$. Also, mean and variance can be derived from the PDF, such that

$$\mu_x = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x \quad \text{and} \quad \sigma_x^2 = \int_{-\infty}^{\infty} x^2 p(x)\mathrm{d}x - \mu_x^2. \tag{7.3}$$

Mean and variance are also called expectation value of first order. An important property of random signals is stationarity. The definition of stationarity can be found, for example, in [51]

> A random process is **strictly stationary** if the PDF is invariant to time shifts.
> A random process is **weakly stationary** if the expectation values of first order, such as mean and variance, are invariant to time shifts.

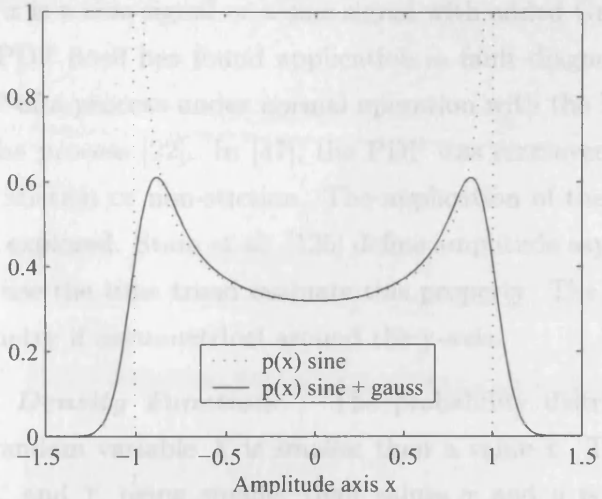


Figure 7.1: Probability density functions of a sine wave $x(t) = \sin(t)$ (dotted line) and of a sine wave with added Gaussian noise $x(t) = \sin(t) + \eta(0, 0.1)$ (solid line).

Testing a random signal for stationarity is not a trivial task. To prove strict stationarity the PDF during all time instances has to be known. Usually, this is not the case but the PDF is estimated from a sample time trend instead. Fortunately, for most applications proving that the random signal is weakly stationary is sufficient. A good indication of weak stationarity is a nearly constant moving average for the period of investigation.

Describing a signal by its PDF discards all time information, that is, the deterministic part of the system. The description of deterministic signals is nevertheless useful since most signals of practical interest have a deterministic and random component. A purely deterministic signal, however, cannot be reconstructed from the PDF alone. For example, the PDF of a triangular periodic signal has a uniform distribution, that is, $p(x) = 0.5$ for $-1 \leq x \leq 1$. This is, accidentally, the same PDF as uniform random noise. A triangular oscillation is therefore one possible, though unlikely, outcome of a random white noise process. More useful is the PDF of a sine wave because of its original shape but even more because of the widespread occurrence of a mixture of sine waves and random noise. The PDF of a pure sine wave is

$$p(x) = \left\{ \begin{array}{ll} \left(\pi\sqrt{A^2 - x^2}\right)^{-1}, & \text{for } -A \leq x \leq A; \\ 0, & \text{elsewhere.} \end{array} \right\} \tag{7.4}$$

where, A is the amplitude of the sine wave. The PDF of a sine wave with added Gaussian random noise with mean μ and standard variance σ is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi^3}} \int_{-A}^{A} \frac{\exp\left(-\frac{(x-z-\mu)^2}{2\sigma^2}\right)}{\sqrt{A^2 - z^2}} \mathrm{d}z \tag{7.5}$$

For a complete derivation of 7.4 and 7.5 see [46] which uses the following theorems:

$$p(z) = p_X(f^{-1}(z))|\frac{\mathrm{d}}{\mathrm{d}z} f^{-1}(z)| \quad \text{where} \quad z = f(x) \tag{7.6}$$

$$p(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z - x) \mathrm{d}x \quad \text{where} \quad z = x + y. \tag{7.7}$$

The trend of $p(x)$ if x is a sine signal or a sine signal with added Gaussian noise is shown in Figure 7.1. The PDF itself has found application in fault diagnosis by comparing the estimate of the PDF of a process under normal operation with the PDF estimate when a fault is present in the process [22]. In [47], the PDF was employed to decide whether a loop fault is due to stiction or non-stiction. The application of the PDF as nonlinearity measure can also be explored. Stam et al. [125] define amplitude asymmetry as a measure of nonlinearity and use the time trend evaluate this property. The PDF is an indication of amplitude asymmetry if asymmetrical around the y-axis.

Joint Probability Density Function: The probability distribution describes the probability that a random variable X is smaller than a value x. The probability of two random variables X and Y being smaller than values x and y is the joint probability distribution

$$P_{XY}(x,y) = P(X \leq x \cap Y \leq y) \tag{7.8}$$

from which the joint probability density function (joint PDF) can be derived

$$p_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} P(x,y). \tag{7.9}$$

Again, $p_{XY}(x,y)$ is abbreviated by $p(x,y)$ if it is unambiguous. The PDF of x and y can be derived from the joint PDF, $p(x) = \int_{-\infty}^{\infty} p(x,y)\mathrm{d}y$, $p(y) = \int_{-\infty}^{\infty} p(x,y)\mathrm{d}x$, and the integral over both x and y must equal to 1. An important property of two random variables is whether the two variables are dependent, that is, whether the knowledge of x has any implication for y. Two variables are independent if the joint PDF is the product of the two PDFs.

$$p(x,y) = p(x)p(y) \tag{7.10}$$

The joint PDF has been defined here for two variables x and y. The concept can be extended to a number of variables dependent in time, say $\mathbf{x}_i = [x_i, x_{i-\kappa}, \ldots x_{i-(k-1)\kappa}]$ and $\mathbf{y}_i = [y_i, y_{i-\kappa}, \ldots y_{i-(l-1)\kappa}]$. The joint PDF is then denoted by $p(\mathbf{x}_i, \mathbf{y}_i)$. The length of the time vectors $\mathbf{x}_i$ and $\mathbf{y}_i$, that is, k and l, are referred to as embedding dimensions for consistency with the same vectors defined for the nearest neighbours method.

Conditional Probability Density Function: Joint PDF gave a measure of dependency between two random variables. The conditional PDF now indicates whether the PDF of x and thus the outcome of x varies if y is known. The conditional PDF is defined by the joint PDF and the PDF of y.

$$p(x|y) = \frac{p(x,y)}{p(y)} \tag{7.11}$$

Transition Probability: The concept of conditional PDF can be extended to consecutive time variables. By doing so, time information is preserved by capturing the dynamics of the underlying system. Thus, the concept of transition probability is defined as follows.

$$p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i) = \frac{p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i)}{p(\mathbf{x}_i, \mathbf{y}_i)} \tag{7.12}$$

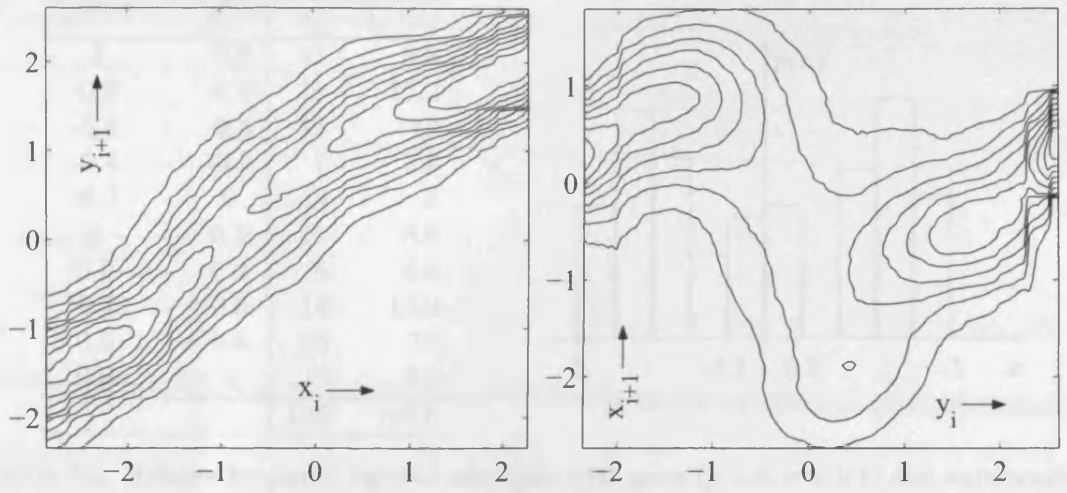


Figure 7.2: Transition probabilities of causal relationship $p(y_{i+1}|x_i)$ (left panel) and non-causal relationship $p(x_{i+1}|y_i)$ (right panel) for $k = 0$ and $l = 1$. The transition probabilities are estimated from time trends of the reference case study with x=TI3 and y=TI4.

Transition probabilities describe the probability of a future value of x if past values of x and y are known. Parameter h is called the prediction horizon. If x_{i+h} is independent of $\mathbf{x}_i$ and $\mathbf{y}_i$ then $p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i) = p(x_{i+h})$ using Equation 7.10. Transition probabilities provide the basis for transfer entropy.

The concept of causality using transition probability is illustrated in Figure 7.2 for two time trends taken from the reference case study in Section 4.2. Here, the embedding dimensions are set to $k = 0$ and $l = 1$ and the prediction horizon is set to $h = 1$. The parameters k and l are chosen in order to display the transition probability in a three dimensional contour plot. Time trend x is temperature measurement TI3 and trend y is TI4 from the example in Section 4.2 of this work. The plots can be interpreted as follows. In the left plot cutting through $p(y_{i+1}|x_i)$ at $x_i = 2$ gives the PDF $p(y_{i+1}|2)$. The expectation value is then $\mu_{y_{i+1}}^{x_i=1} = \int_{-\infty}^{\infty} x_{i+1} p(x_{i+1}|1) \mathrm{d}x_{i+1}$. If the expectation value varies for different values of x_i then y depends on x.

The left panel of Figure 7.2 shows the transition probability $p(y_{i+1}|x_i)$, that is, y as a cause of x. Cutting for example at $x_i = -2$ results in $\mu_{y_{i+1}}^{x_i=-2} \approx -1.8$ while cutting at $x_i = +2$ results in $\mu_{y_{i+1}}^{x_i=+2} \approx +2.0$. Thus, x influences or causes y. On the other hand, cutting at $y_i = -2$ in the right panel results in $\mu_{x_{i+1}}^{y_i=-2} \approx 0.2$ which is very similar to $\mu_{x_{i+1}}^{y_i=+2} \approx 0.0$. The conclusion is that x does not influence y.

7.2 Estimation of Probability Density Function

In almost every application, PDF, joint PDF and transition probabilities of the random processes investigated are not known. The PDFs however can be estimated from a time trend realisation of the process. PDFs are estimated by histograms in a straight forward

$x^{(j-1)}$	-	$x^{(j)}$	n_j	$\overline{h}_j$ [%]
-1	-	-0.8	11	8.8
-0.8	-	-0.6	19	15.2
-0.6	-	-0.4	15	12
-0.4	-	-0.2	7	5.6
-0.2	-	0	10	8
0	-	0.2	11	8.8
0.2	-	0.4	6	4.8
0.4	-	0.6	14	11.2
0.6	-	0.8	20	16
0.8	-	1	12	9.6
$\sum$			125	100.0

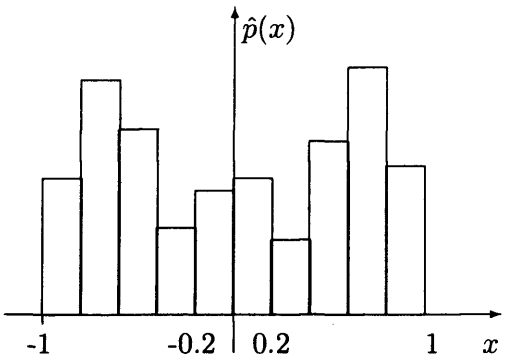


Figure 7.3: Relative frequency table of sine signal with noise ($\mu = 0$, $\sigma = 0.1$) and corresponding histogram estimation, $N = 125$, $n = 10$.

manor or by Kernel estimations. Both techniques are established and well understood. They will be introduced here briefly, though only the Kernel estimator is used for the calculation of transfer entropy in the following sections due to its higher accuracy.

7.2.1 Histogram

In histograms, also referred to as relative frequency, the number of measurements that lie within a measurement interval are counted and divided by the total number of measurements. Discrete measurement intervals are constructed by partitioning the amplitude axis into n so-called amplitude bins. The jth interval of a signal x is then defined as

$$\begin{aligned} \Delta x_j &= [x_{min} + (j-1)\overline{h}\,;\, x_{min} + j\overline{h}] \\ &= [x^{(j-1)}\,;\, x^{(j)}]. \end{aligned} \tag{7.13}$$

Here, $\overline{h}$ is the bin width defined by $(x_{max} - x_{min})/n$ with x_{min} and x_{max} minimum and maximum value of the time series. If N is the total number of samples available, then the relative frequency of bin Δx_j is the number of samples that fall into this bin: $\overline{h}_j = \frac{n_j}{N}$ with $n_j = \{x_i | x_i \in \Delta x_j\}$. The estimated PDF is then defined by

$$\hat{p}(x) = \sum_{i=1}^{n} \overline{h}_j \{\Theta(x - x^{(j-1)}) - \Theta(x^{(j)})\} \tag{7.14}$$

where Θ is the step function. An example of a PDF estimate using histograms is shown in Figure 7.3. Comparing Figure 7.3 to the analytical solution shown in Figure 7.1 shows that the estimation with histograms is only a poor representation of the real trend.

The parameter to adjust when estimating the PDF is the optimal bin width h which controls the number of bins $n = [(x_{max} - x_{min})/\overline{h}]_{\text{round}}$. If the bin width is chosen too large then the histogram will appear as a block of squares and if the bin width if set too small then appear very granular with empty bins next to bins with a large number of samples. For an optimal bin width, Scott [115] proposes to minimise the integral mean

square error between the estimated and the actual underlying PDF. The minimisation problem results in the following optimal bin width

$$\overline{h}_{\text{opt}} = \left[N \int_{-\infty}^{\infty} p^2(x) \mathrm{d}x / 6 \right]^{-1/3}. \tag{7.15}$$

Assuming a Gaussian distribution $p(x)$, the optimal bin width is then given by

$$\overline{h}_{\text{opt}} = 2\sigma_x \left(\frac{3\sqrt{\pi}}{N} \right)^{1/3}. \tag{7.16}$$

The number of bins follows accordingly. The estimation of the transition probabilities using histograms is not recommended. To illustrate this, Equation 7.16 is solved for N: $N = \left(\frac{2\sigma_x}{\overline{h}_{\text{opt}}}\right)^3 3\sqrt{\pi}$. To achieve a good estimate of the sine wave signal with added Gaussian noise, from Equation 7.5, 10 bins are assumed, resulting in $\overline{h}_{\text{opt}} \approx 0.2$. The standard deviation of sine with noise is $\sigma_x \approx 0.71$. The number of samples required for the estimation is then $N = \left(\frac{2 \cdot 0.71}{0.2}\right)^3 3\sqrt{\pi} \approx 1903$. Estimating transition probabilities requires the power of that number, depending on embedding dimension k and l. Setting $k = 0$ and $l = 2$ requires $N = 1903^{0+2+1} = 6.810^9$ samples for a good estimation.

7.2.2 Kernel Estimation

The Kernel method gives a more precise estimation of the PDF than histograms by considering the exact values of a time series x. A Kernel function K is centered around every sample point and summed to give an estimate $\hat{p}$.

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^{N} K(x - x_i) \tag{7.17}$$

The Kernel function K has to fulfill $\int_{-\infty}^{\infty} K(x)\mathrm{d}x = 1$, its maximum value must be at $x = 0$ and the limit values for plus and minus infinity are zero. Here, a Gaussian Kernel function is used which satisfies all requirements.

$$K(x - x_i) = \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{(x - x_i)^2}{2\overline{h}^2}\right) \tag{7.18}$$

where $\overline{h}$ is the estimator width which is adjusted to the number of samples N and the standard deviation of the time series x after Silverman [119]. The following example explains the construction of the PDF. Consider the example of the sine signal with added noise as given in Equation 7.5. Figure 7.4 shows the construction of the estimated PDF $\hat{p}(x)$ using the Kernel method as well as the actual PDF $p(x)$. Even for a small number of samples the Kernel method gives a good estimation of the PDF.

In [119], the amplitude axis is considered as continuous. For implementation, however, the x-axis has to be discretised similar to histograms. The difference is that the number of amplitude bins can be set arbitrarily high and is not restricted by the number of samples

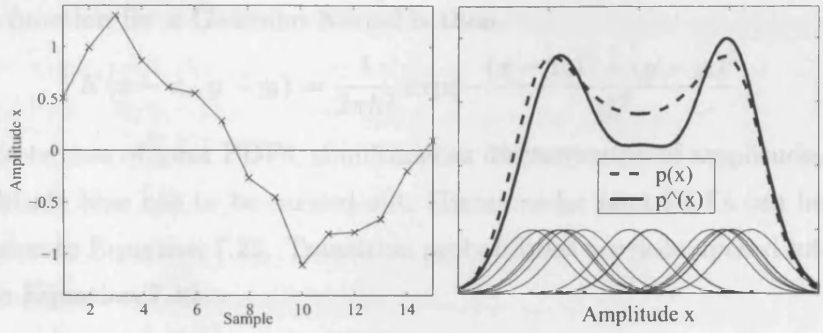


Figure 7.4: Example of Kernel estimation, (a) discrete sine function with additive noise, construction of Kernel estimator $\hat{p}(x)$ using 15 samples and (b) actual PDF $p(x)$.

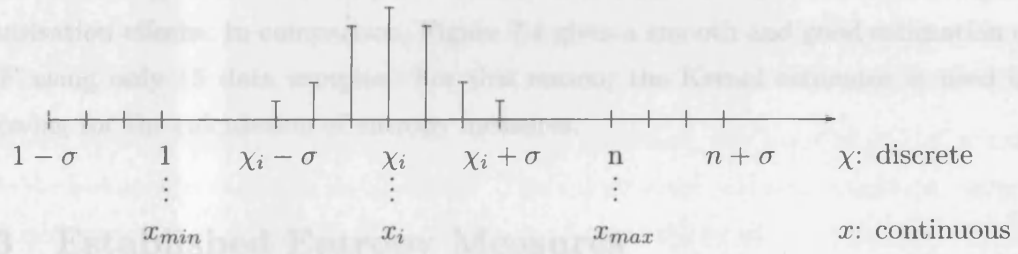


Figure 7.5: Kernel estimation: Transformation from continuous to discrete amplitude axis according to Equation 7.19. The total number of bins is $n + 2\sigma$.

N or the standard deviation σ_x. A discrete value χ_i is introduced that translates a sample value of the time series, x_i, onto a discrete grid $\chi_i \in [1; n]$.

$$\chi_i = \left[(n-1) \frac{x_i - x_{min}}{x_{max} - x_{min}} \right]_{round} + 1 \tag{7.19}$$

The number of amplitude bins is denoted by n and can be set independently. The PDF is estimated according to Equation 7.17 with discrete Kernel function

$$K\left[\chi - \chi_i\right] = \frac{1}{a\sigma\sqrt{\pi}} \exp(-(\chi - \chi_i)^2/\sigma^2) \tag{7.20}$$

with $\chi = \chi_i - \sigma, \chi_i - \sigma + 1, \ldots, \chi_i + \sigma$. Due to the finite length of $K\left[\chi - \chi_i\right]$ a scaling by factor $a = \sum_{\chi=-\sigma}^{\sigma} \frac{1}{\sigma\sqrt{\pi}} \exp(-\chi^2/\sigma^2)$ is required. For the discrete Kernel, estimator width h is replaced by the discrete Kernel width σ. The optimal Kernel width is

$$\sigma = \left[cN^{-1/5} \sigma_x n \right]_{round}, \quad \text{with } c \approx 0.2. \tag{7.21}$$

The translation from continuous to discrete x-axis for the PDF estimation is shown in Figure 7.5. The total number of bins is $n + 2\sigma$ with $\sigma \ll n$ per definition.

The estimation of the joint PDF can be constructed parallel to the PDF. The estimation of the joint PDF for x and y using the Kernel method is denoted by

$$\hat{p}(x, y) = \frac{1}{N} \sum_{i=1}^{N} K(x - x_i, y - y_i). \tag{7.22}$$

The Kernel function for a Gaussian Kernel is then

$$K(x - x_i, y - y_i) = \frac{1}{2\pi h^2} \exp(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\overline{h}^2}). \tag{7.23}$$

For the discrete case of joint PDFs, simultaneous discretization of amplitude axes x and y into n amplitude bins has to be carried out. Higher order joint PDFs can be constructed as an extension to Equation 7.22. Transition probabilities are decomposed into joint PDFs according to Equation 7.12.

The increased estimation quality can be clearly seen when comparing the estimation of the sine wave signal with noise for both histogram and Kernel estimation. Figure 7.3 shows the histogram for 125 samples available. The estimate is poor and shows significant quantisation effects. In comparison, Figure 7.4 gives a smooth and good estimation of the PDF using only 15 data samples. For this reason, the Kernel estimator is used in the following for the calculation of entropy measures.

7.3 Established Entropy Measures

The concept of measuring information was introduced shortly after the development of communication systems to quantify the amount of information transferred from the source to the receiver in a communication system. In 1948, Shannon and Weaver [117] and Wiener [149] introduced in parallel developments a logarithmic measure of information, referred to as entropy.

> In the context of statistical communication theory, **information** is understood as the amount of **uncertainty** or **randomness** of a random process and entropy is a measure of the information contained in that process.

In his early textbook on information theory and statistics Kullback [70] points out that information theory is a branch of mathematical theory of probability and statistics rather than a synonym to communication theory. Entropy measures are applicable to any random process described by a probability density function. More recent textbooks by Gray [39] or Ihara [51] give further applications for entropy measures such as hypothesis testing and discrete maximum entropy analysis.

Entropy: The entropy of random variable X or the entropy of the probability density function $p(x)$ is defined as

$$G(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) \mathrm{d}x \tag{7.24}$$

to measure the amount of information or uncertainty contained in random variable X. The logarithm may be taken to base e, the natural logarithm, or base 2. For the natural

logarithm, entropy is measured in 'nats' while for the base 2 logarithm entropy is measured in 'bits'. The signal with the highest uncertainty is uniform random noise. Every outcome of the random process is equally likely and therefore has a maximum uncertainty. Using the notational convention $0 \cdot \log 0 = 0$ gives the entropy $G(X) = \log 2a$ from Equation 7.24 for uniform random noise distributed between $-a$ and a. The minimum value of the entropy measure is achieved by a constant signal which has an entropy value of $G(X) = 0$. Thus, the boundaries of the entropy value are

$$0 \leq G(X) \leq \log 2a. \tag{7.25}$$

The maximum value therefore depends on the interval defined by a over which the amplitude of random signal X is distributed. For discrete probability density functions the integral in Equation 7.24 can be replaced by the sum over all discrete amplitude bins.

Mutual Information: Entropy on its own measures the uncertainty of a random variable before the outcome is observed. The uncertainty however might be reduced if a second variable is observed. The reduction of uncertainty of two random variables is described by mutual information as follows.

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y \tag{7.26}$$

Here, $p(x,y)$ is the joint PDF of the two random variables. The independence property from Equation 7.10 can be exploited to explain the concept of mutual information. If X and Y are independent, that is, $p(x,y) = p(x) \cdot p(y)$, then the mutual information can be simplified to $I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log 1\mathrm{d}x\mathrm{d}y = 0$. This property leads to the observation that mutual information is a single number that expresses the strength of dependency between two variables described by their PDFs and joint PDF.

Conditional Entropy: Conditional entropy describes the remaining uncertainty of a random variable after the outcome of another variable is known and is denoted by

$$\begin{aligned} G(X|Y) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x|y)\mathrm{d}x\mathrm{d}y \\ &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\tfrac{p(x,y)}{p(y)}\mathrm{d}x\mathrm{d}y \end{aligned} \tag{7.27}$$

$G(Y|X)$ is calculated in an analogous way. Conditional entropy is an asymmetric measure. However, the comparison between $G(X|Y)$ and $G(Y|X)$ does not give a measure of causality because the difference is a function of the entropy measures $G(X)$ and $G(Y)$ of the time sequence. In other words, the difference of conditional entropies equals to the difference of the entropies. This becomes more clear when considering that:

$$G(X) = I(X,Y) + G(X|Y) \quad \text{and} \quad G(Y) = I(X,Y) + G(Y|X) \tag{7.28}$$

For a derivation of these equation see Appendix C. Subtracting these two equations results in

$$G(X) - G(Y) = G(X|Y) - G(Y|X) \tag{7.29}$$

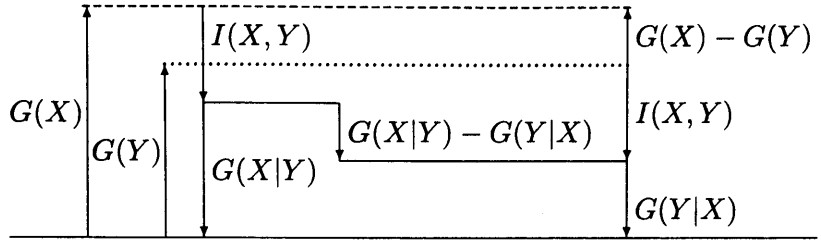


Figure 7.6: Entropy measures: relationship between entropy, mutual information and conditional entropy of two random variables.

and thus shows that the difference of the conditional entropies equals to the difference of the two entropies. The relationship is illustrated in Figure 7.6 where the uncertainty of X is larger than the uncertainty of Y and the X and Y are dependent. Conditional entropy can be calculated for discrete PDFs by replacing the integrals in Equation 7.27 by sums of the discrete amplitude bins.

7.4 Transfer Entropy

Transition probabilities as introduced in Equation 7.12 contain information on causality. Unlike conditional entropy and mutual information, transition probabilities capture the time dependent behaviour and can therefore distinguish which variable causes the other. The recently proposed method by Schreiber [113] of transfer entropy measures the amount of information transferred from variable X to Y similar to the measure of dependency expressed by mutual information. Transfer entropy is calculated as follows.

$$T(X|Y) = \int_{\mathrm{R}^k} \int_{\mathrm{R}^l} \int p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i) \cdot \log \frac{p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i)}{p(x_{i+h}|\mathbf{x}_i)} \mathrm{d}x_{i+h} \mathrm{d}\mathbf{y}_i \mathrm{d}\mathbf{x}_i \tag{7.30}$$

The integrals are taken over all k past values of x and l past values of y as well as over the future value x_{i+h} and can be replaced by sums for discrete transition probability and joint PDF. The numerator of the logarithmic term, describing the probability of x_{i+h} if past values $\mathbf{x}$ and $\mathbf{y}$ are known, is compared to the denominator, describing the probability of x_{i+h} if only $\mathbf{x}$ is known. Transfer entropy therefore incorporates a time dependency with a functional dependency and is therefore suited to capture fault propagation dependencies as described in Section 2.4.3, that is, time lags, magnitude decrease and functional attenuation.

Transfer entropy has been previously applied to physiological and financial data. Schreiber [113] investigates whether the heart rate affects the breath rate or vice versa. Marschinski and Kaiser [78] and Kaiser and Schreiber [55] analyse cause and effect between various stock exchange indices.

7.4.1 Causality Measure

The transition probabilities in Equation 7.30 for the definition of transfer entropy can be replaced according to Equation 7.12 by the joint PDFs. Replacing the transition probabilities by joint PDFs and the definition of independence in Equation 7.10 give the lower threshold of $T(X|Y)$. If x_{i+h} is independent from $\mathbf{y}_i$, then no information is transferred from y to x. In this case, the log term in Equation 7.30 is

$$\frac{p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i)}{p(x_{i+h}|\mathbf{x}_i)} = \frac{\frac{p(x_{i+h}|\mathbf{x}_i)p(\mathbf{y}_i)}{p(\mathbf{y}_i)}}{p(x_{i+h}|\mathbf{x}_i)} = 1, \tag{7.31}$$

and $\log(1) = 0$. Thus, if y does not influence x then the transfer entropy measure is zero. The maximum value is not trivial to compute and depends on the amplitude over which x and y are distributed. Nevertheless, for causality analysis for plant-wide disturbances, the difference between $T(X|Y)$ and $T(Y|X)$ gives a measure of directionality.

$$t_{x\to y} = T(Y|X) - T(X|Y). \tag{7.32}$$

Thus, large values of $t_{x\to y}$ indicate a strong causality from x to y. The directionality measure treats the case of strong bidirectional coupling identical to the case of no presence of coupling which will both result in a small value of $t_{x\to y}$. For the directionality problems investigated, this is of main interest. Bidirectional coupling can be interpreted in a further analysis step.

7.4.2 Significance Level

Small values of transfer entropy $t_{x\to y}$ suggest no causality or direction of influence while large values do. To establish a threshold above which $t_{x\to y}$ is recognized as a valid result Kantz and Schreiber [57] suggest Monte Carlo methods using surrogate data. The threshold is referred to as significance or confidence level. Here, the problem is to verify or reject a null hypothesis. In the case of the directionality estimation, the null hypothesis is that the transfer entropy measure $t_{x\to y}$ is valid, that is, it is large enough to imply that x influences y. The null hypothesis is denoted by $\lambda_0 = t_{x\to y}$. Suitable random numbers are values of the directionality measure that are generated by computing the measure from N_s surrogate time series such that $\lambda_j = t_{x\to y}^{surr,j}$ with $j = 1 \ldots N_s$. For surrogate time series construction, the iterative amplitude adjusted Fourier transform (iAAFT) method as described by Schreiber and Schmitz [114] is used in the following computations. The significance level is then defined as

$$s_{x\to y} = \frac{\lambda_0 - \mu_\lambda}{\sigma_\lambda} > 6 \tag{7.33}$$

where μ_λ and σ_λ are mean and standard deviation of λ_j. A six sigma threshold for the significance level is chosen here rather than a two or three sigma test as in [57] since $t_{x\to y}$ does not necessarily follow the Gaussian distribution assumed by Equation 7.33.

Joint PDF	Dimension	$k=1, l=2$
$p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i)$	$k+l+1$	4
$p(\mathbf{x}_i, \mathbf{y}_i)$	$k+l$	3
$p(x_{i+h}, \mathbf{x}_i)$	$k+1$	2
$p(\mathbf{x}_i)$	k	1

Table 7.1: Types and dimensions of joint probability density functions required for implementation of transfer entropy in Equation 7.30.

Marschinski and Kantz [78] use surrogate time series for transfer entropy referred to as effective transfer entropy. It is defined as the difference of transfer entropy from original data sets and shuffled data, such that

$$ET(X|Y) = T(X|Y) - T(X, Y^{surr}). \tag{7.34}$$

The data is shuffled once for reference. Effective transfer entropy gave good results for financial data. In the following, the significance level is used rather than effective transfer entropy as it gave good and, more importantly, robust results.

7.4.3 Computational Effort

Before calculating the causality value in Equation 4.3 from transfer entropy in Equation 7.30, joint PDFs and transition probabilities have to be constructed from time series. Replacing the transition probability through joint PDFs after Equation 7.12 gives the joint PDFs as summarised in Table 7.1 required for the computation of $t_{x\to y}$. The computational effort is significant and a computation with modern PC technology can take up to a couple of seconds for each transfer entropy calculation. This, and the finite number of samples limits choice of embedding dimensions k and l. The estimation of the computational effort is separated into two parts. Firstly, the construction of the PDF has to be considered and secondly the calculation of the transfer entropy value after Equation 7.30. For each joint PDF of dimension k, $(2\sigma+1)^k N$ summations are required. Considering the joint PDFs required after Table 7.1, this adds up to $q^k(q+1)(q^l+1)N$ summations with $q = 2\sigma+1$ as the length of the Kernel. After Equation 7.21, σ increases with the number of bins n and with the number of samples N. The computation of transfer entropy requires $(n+2\sigma)^{k+l+1}$ summations. Embedding dimensions k and l have to be chosen carefully as the number of computations increases with their power. In the following, they are set to $k=0$ and $l=2$. This results in a total of $4N(2\sigma^3+4\sigma^2+3\sigma+1)+(n+2\sigma)^3$ summations.

7.5 Parameter Optimisation

The calculation of transfer entropy requires a number of parameters to be set. It will be shown in the following that when applied to chemical process data the transfer entropy

measure is insensitive to the parameter setting within bounds. Parameters will be optimized by applying them to data of a reference case study. The transfer entropy parameters are optimized using these relationships parallel to the optimisation procedure in Chapter 6. The parameter settings are then applied to a test case study for which the root cause was unknown prior to analysis. The sampling interval in the reference case study is 20 seconds. This has to be considered when using the results for other data sets.

Fixed parameters are the embedding dimensions which are set to $k = 0$ and $l = 2$ due to computational constraints as argued above. The special case $k = 0$ means that only past values of y and not x are regarded when estimating a future value of x so that for instance T(TI1|TI2) in Eq. 7.30 would be calculated using past values of TI2 and the current value of TI1 while T(TI2|TI1) uses past values of TI1 and the current value of TI2. Adjustable parameters are the time lag κ between past value of y, such that $\mathbf{y} = [y_i, y_{i-\kappa}]$, and the prediction horizon h of the future value of x, x_{i+h} in Eq. 7.30. The optimal parameter of both time lag and prediction horizon are a function of the process dynamics. If the process dynamics are known, the parameters can be set accordingly. If a dead time is detected between two measurements both κ and h should have the optimum value when set equal to the dead time. However, if the process dynamics are unknown, smaller values of κ and h should give good results as the transfer entropy measure should be robust for parameter changes. The minimum number of sample N required for computation is also investigated in the following.

7.5.1 Algorithm Time Lag

In a first step, the time lag is varied while the other parameters are fixed. Time lag κ was defined for the transition probabilities in Equation 7.12 which uses embedded vectors $\mathbf{x}_i = [x_i, x_{i-\kappa}, \ldots, x_{i-(k-1)\kappa}]$ and $\mathbf{y}_i = [y_i, y_{i-\kappa}, \ldots, y_{i-(k-1)\kappa}]$. A change of the time lag corresponds to sub-sampling of the data series with the benefit of no data being disregarded. Adjusting the time lag ensures that dynamics of the underlying system are represented accurately. Figure 7.7 shows transfer entropy values and significance level for the five relationships of the reference case study. The significance level above the threshold of six sigma for most time lags which shows that transfer entropy is robust. The maximum value of $s_{x\to y}$ is at different values of κ for the five plots, varying from $\kappa = 2$ for TI1→TI2 to $\kappa = 9$ for LC1→TI7. When summing all $s_{x\to y}$ over the five relationships the maximum lies at $\kappa = 4$. Thus, the time lag is set to $\kappa = 4$ which gives a significance level well above the threshold for all five relationships. With a sampling rate of 20 seconds time lag κ is equivalent to 80 seconds. In general, time lag κ is dependent on the process dynamics of the process. However, Figure 7.7 shows that the result is robust to variations in κ so that a certain mismatch between process dynamics and time lag can still give good results.

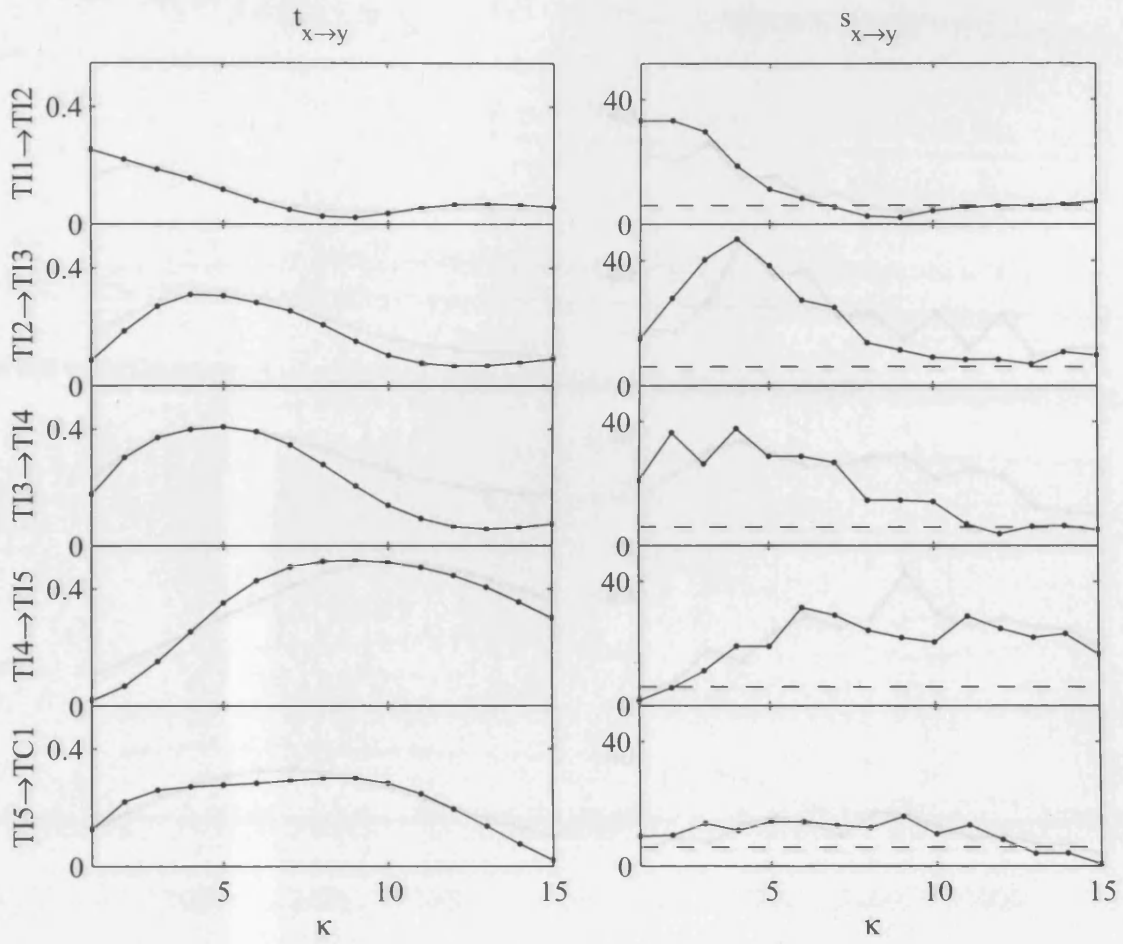


Figure 7.7: Optimizing time lag κ for five dependencies; left hand plots: transfer entropy value $t_{x\to y}$, right hand plots: significance value $s_{x\to y}$; with $h = \kappa$ and $N = 4100$.

7.5.2 Prediction Horizon

The prediction horizon was set to the κ while optimizing time lag κ. It is expected that the value of the prediction horizon h lies in the same range as the time lag since dynamics of the underlying system are also reflected by the prediction horizon. Figure 7.8 shows transfer entropy and significance value as a function of h. The dependency of transfer entropy on h is similar to the dependency on κ as in Figure 7.7. The values of $t_{x\to y}$ and $s_{x\to y}$ are robust against the choice of h. The maximum value of $s_{x\to y}$ varies for the five relationships but the sum of all significance levels has its maximum at $h = 4$. Thus, the prediction horizon is chosen to be equal to the time lag.

7.5.3 Minimum Number of Samples

For practical application the number of samples to be considered for computation is a crucial parameter. The question is how many samples are required to get a significant result, that is, the minimum number of samples $N_{\min}$. The impact of the number of samples on the transfer entropy measure $t_{x\to y}$ is shown in Figure 7.9. Average value and

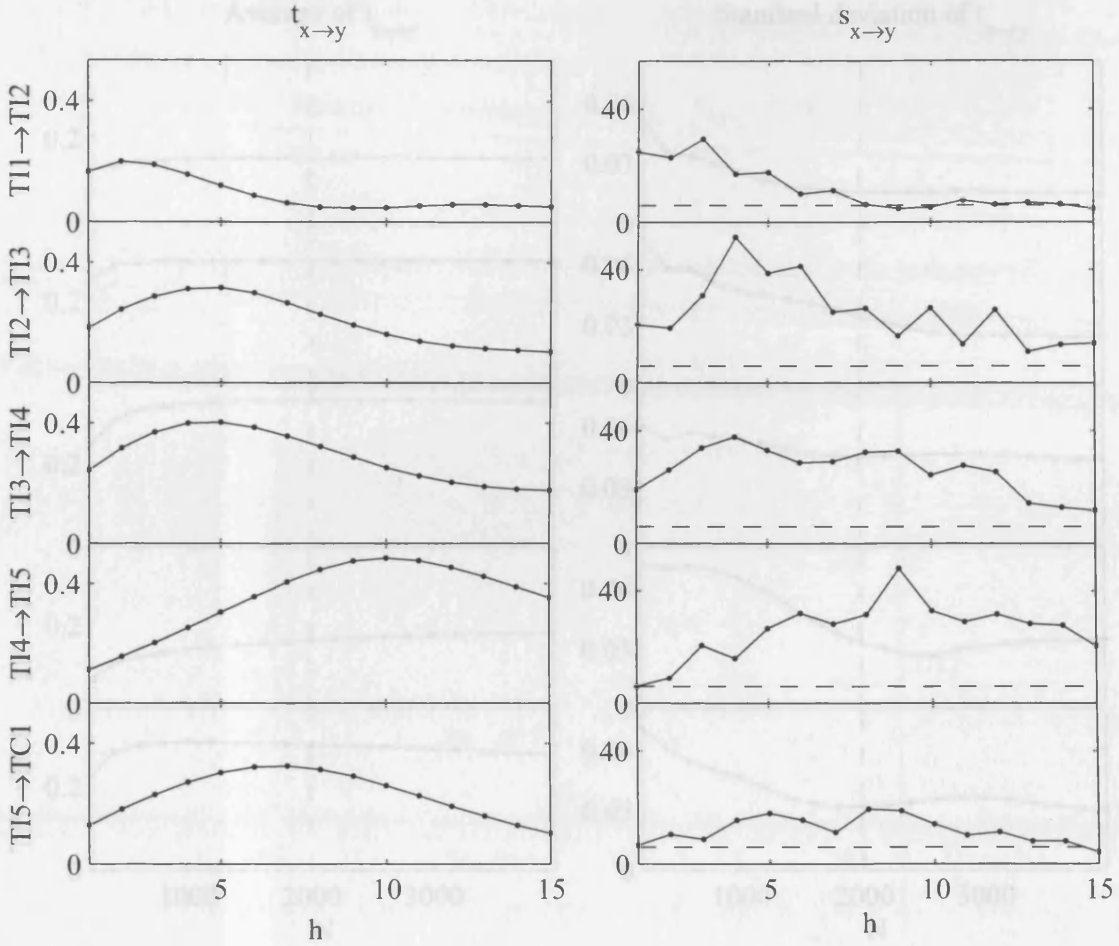


Figure 7.8: Optimizing prediction horizon h for five dependencies; left hand plots: transfer entropy value $t_{x\to y}$, right hand plots: significance value $s_{x\to y}$; with $\kappa = 4$ and $N = 4100$.

standard deviation of $t_{x\to y}$ are calculated using subsets of the original data set of length N. The subsets are overlapping and delayed by 100 samples. For example, with the total length being 4100 samples, 40 subsets are constructed for $N = 200$ consisting of samples 1 to 200, 101 to 300, 201 to 400 and so on, down to 2 subsets for $N = 4000$. The left hand plots of Figure 7.9 shows that the average value of $t_{x\to y}$ does not vary significantly when $N \geq 600$. However, the results of $t_{x\to y}$ for the subsets can vary as shown in the right hand plots. The standard deviation only levels out for approximately $N > 2000$ samples.

The recommendation is that the minimum number of samples should be set to $N_{\text{min}} = 2000$ if possible. If, however, fewer than 2000 samples are available, it is still worth carrying out transfer entropy analysis down to 400 or 500 samples. The confidence level in the result will then be lower as a consequence.

7.5.4 Guidelines

The guidelines for setting the parameters resulting from the reference case study are summarized in Table 7.2. The parameters are dependent on the dynamics of the underlying

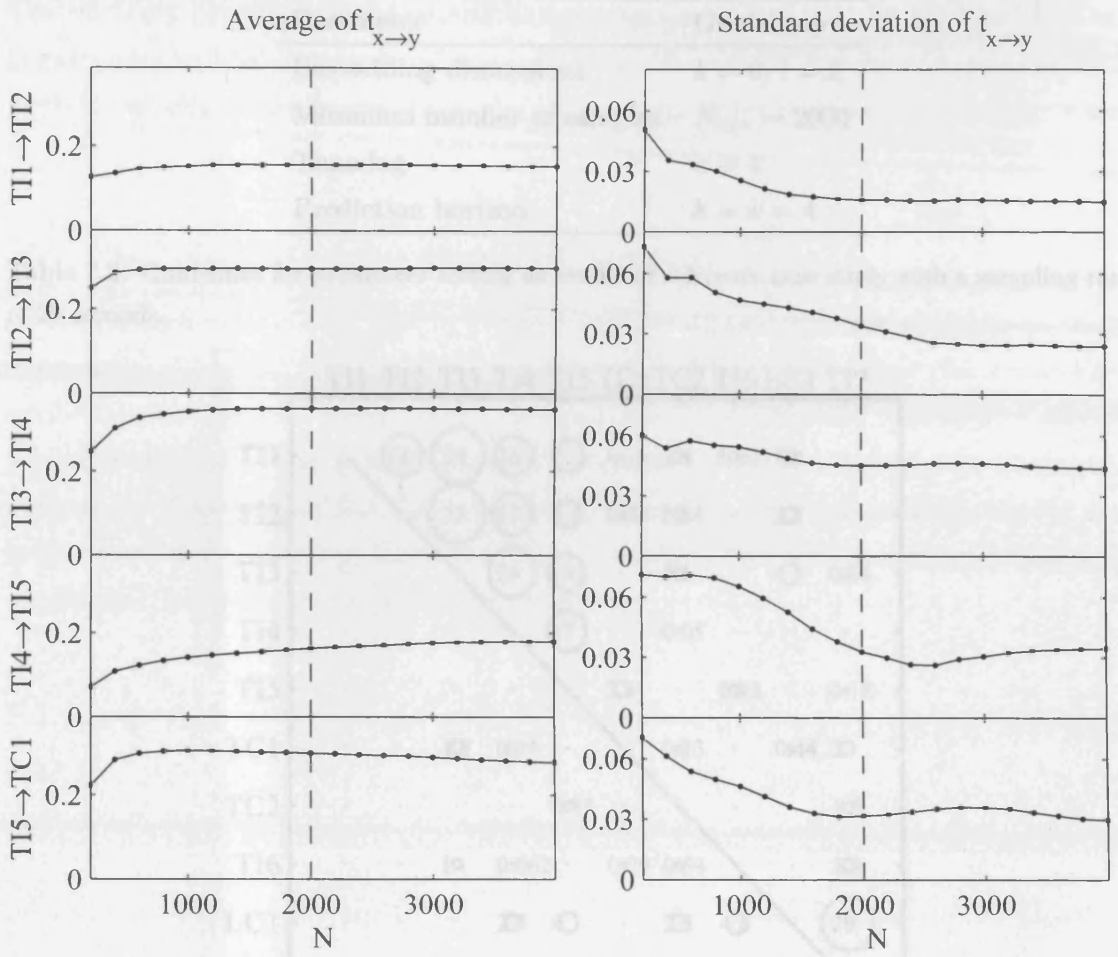


Figure 7.9: Finding minimum number of samples $N_{\min}$ for five dependencies; left hand plots: average of transfer entropy $t_{x\to y}$, right hand plots: standard deviation of $t_{x\to y}$; with $h = \kappa = 4$.

nature of the data. The data investigated in the reference case study is cycling with a period of oscillation of 60 samples. Setting time lag and prediction horizon to $\kappa = h = 4$ corresponds to sub-sampling by factor 4. Thus, the number of samples of oscillation period are reduced to 15. This must be taken into account when investigating other case studies. The transfer entropy measure however is robust to changes in both time lag and prediction horizon. It is therefore expected that the parameters guidelines given in Table 7.2 also give good results if the underlying dynamics differ from the ones in the reference case study.

Case Study: *Transfer entropy for the parameter estimation is applied to the full reference case study from Section 4.2 using the proposed guidelines for time lag and prediction horizon. The directionality measure and the significance level are computed for all combinations of the process variables. Significance level and transfer entropy value for all combinations of process variables of the reference case study is shown in Figure 7.10. A significance level above the threshold of six can be detected for the relationships between TI1 to TI5 and LC1 and TI7.*

Parameter	Guidelines
Embedding dimensions	$k = 0$, $l = 2$
Minimum number of samples	$N_{\min} = 2000$
Time lag	$\kappa = 4$
Prediction horizon	$h = \kappa = 4$

Table 7.2: Guidelines for parameter setting as results of reference case study with a sampling rate of 20 seconds.

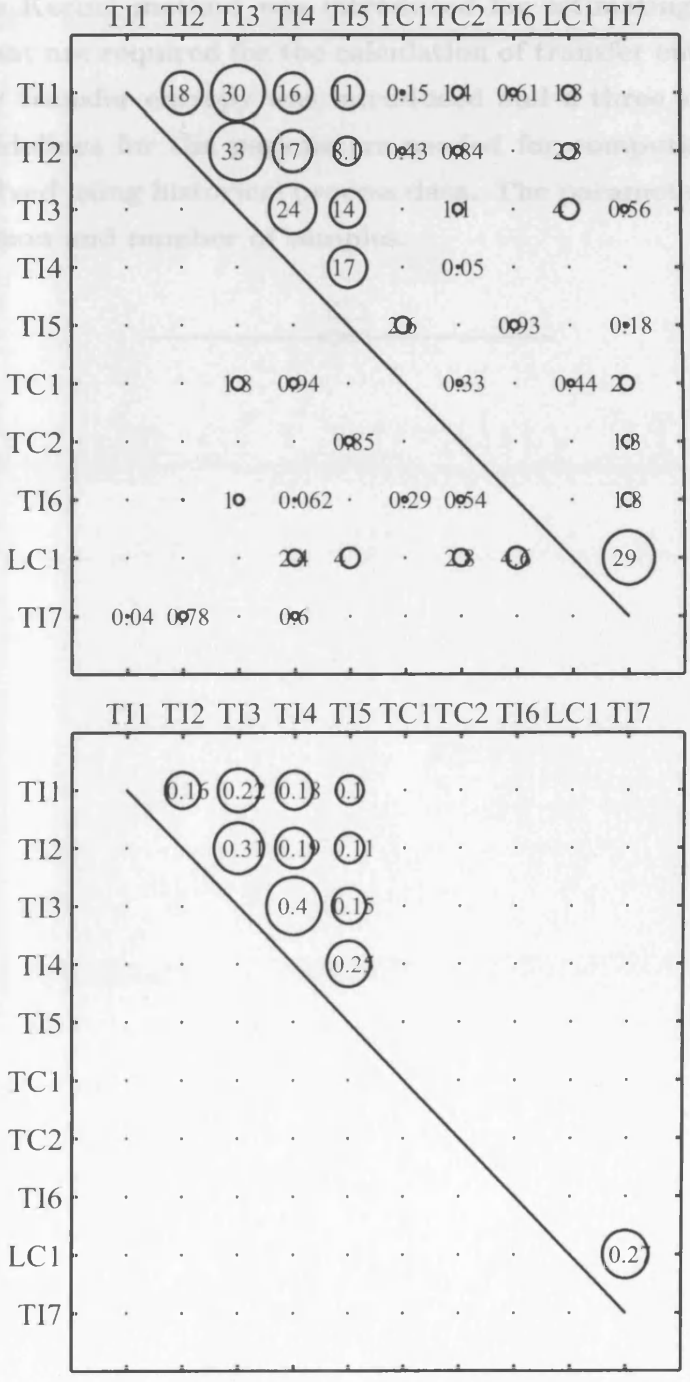
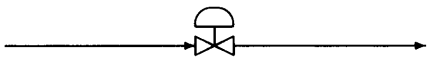


Figure 7.10: Transfer entropy results of reference case study, significance level $s_{x \to y}$ (top) and causality values $h_{x \to y}$ if significance level exceeds threshold (bottom).

The efficiency of transfer entropy will be investigated in Part III and the results of the investigation will be discussed and summarised in Chapter 11. A comparison to the CCF method introduced in Chapter 5 and the nearest neighbours method from Chapter 6 will be made.

Chapter 7 Summary

In this chapter, transfer entropy has been developed for the purpose of fault diagnosis. The Kernel method was introduced for estimating the transition probabilities that are required for the calculation of transfer entropy. A significance level for transfer entropy was introduced and a three sigma threshold proposed. Guidelines for the parameters needed for computing transfer entropy were derived using historical process data. The parameters are time lag, prediction horizon and number of samples.



Part III

Application of Causality Measures

PART III - APPLICATION OF CAUSALITY MEASURES

In this part, the causality methods of cross-correlation, nearest neighbours and transfer entropy developed in the previous part are applied to a number of simulated and industrial data sets. The purpose of the application to various case studies is to provide means of comparison of the developed tools among each other and in relation to other data-driven methods. Since the methods are all statistical tools, the impact of changes in the nature of data is difficult to predict. Thus, an experimental approach is pursued throughout this part by analysing the impact that different natures of data have on the causality measures.

The fault propagation effects that were introduced and described in Section 2.4.3 are investigated here and the impact they have on the causality measures are studied in detail in Chapter 8. The effects that can be used to argue cause and effect are dead time (because the disturbance propagates from one measurement to the next with a time delay) and low pass filtering (because most processes act as a low pass filter). The impact of additive noise is also studied though it distorts the causality results rather than arguing cause and effect. In Chapter 9, two complete industrial case studies are investigated in which the cause and effect relationships investigate the root cause of a disturbance. The case studies are from plants at Eastman Chemical Company and BP Chemicals sites. Finally in Chapter 10, the impact of quantisation and compression on the three methods is investigated using real process data that is then quantised and compressed.

Chapter 8

Fault Propagation Effects

In this chapter, the causality measures are applied to benchmark data sets that model the fault propagation effects. The effects investigated are dead time, low pass filtering and additive noise. The aims of applying the causality measures to simulated data are as follows:

- **To investigate the impact of an isolated change in the nature of the data, that is the impact of a time delay and low pass filtering, on the causality measures;**
- **To compare the performance of the three causality measures in relation to these changes to the nature of the data;**
- **To find out whether further insight into the problem studied can be gained through cause and effect analysis.**

The investigated simulated data in this chapter is specially designed to test the response of the methods to fault propagation effects that can occur in chemical processes. Simulated data does not contain any structural noise that is essential for the statistical methods investigated in this work, however, it serves the purpose of getting more insights into the methods. Investigating simulated data thus serves the purpose of investigating the impact of changes in the data trend that occur when the signal propagates in the process (see Section 2.4.3). The four effects commonly observed are dead time or time delay, low pass attenuation and added noise, as listed in Table 2.5. The effect of each propagation will be investigated in turn for the cross-correlation, transfer entropy and nearest neighbours method. Since the signals are normalised prior to analysis, a decrease in magnitude cannot be captured with the causality measures. A decrease in magnitude is only of significance if identical measurements are compared, for example, flow measurement with flow measurement. It can be estimated by calculating the standard deviation as a percentage of the mean. This quantity is only valid if no equipment acts on the variables, for example, a buffer tank might decrease the variations in flow rate in general so that the disturbance has a larger impact on the standard deviation / mean percentage than before the buffer

Effect	Methods	Won't Work	Comment
Dead Time	CCF, TE, NN	Var	see 8.1
Low Pass Filter	TE, NN	Var, CCF	see 8.2
Additive Noise	CCF, TE, NN	Var	see 8.3
Amplitude Attenuation	Var	TE, NN, CCF	Only for same types

Table 8.1: Fault propagation effects and measures of causality; CCF: cross-correlation function, NN: nearest neighbours, TE: transfer entropy; Var: variability analysis using variance or standard deviation.

tank.

The effects acting on a process quantity such as flow, temperature or level are briefly described in Section 2.4.3, that is, dead time, low pass filtering, amplitude attenuation and additive noise. These effects and the methods applied to them are summarised in Table 8.1. In the following, the effect of dead time will be discussed in Section 8.1. Arguing cause and effect from the variance contained in the signal will give no result as the variance will be similar for both the delayed and the original time series. Section 8.2 discusses the impact of a low pass filter element on the causality measure. The cross-correlation causality method will not be able to detect the fault propagation since no time delay is introduced through the filtering. The impact of additive noise on cross-correlation, nearest neighbours and transfer entropy is discussed in Section 8.3. The case of amplitude attenuation will not be discussed in this chapter since amplitude attenuation can be best measured using a variability measure and arguing that the variability is strongest for the measurement closest to the root cause. This argument, however, can be misleading. Only measurements of the same types, that is flow and flow or pressure and pressure, can be compared.

The questions addressed in this chapter are as follows:

- Which of the three developed causality measures detects the direction of interdependence best through the presence of a time delay?
- Can transfer entropy or nearest neighbours argue the direction of interdependence better through the presence of a low pass filtering element?
- Which of the three causality measures is most robust to additive noise?

8.1 Signal Dead Time

As the product is processed in the plant, it travels through along the equipment with a certain speed or throughput. The propagation time between two measurement points is not known in general and is usually larger than the sampling time of the measurements.

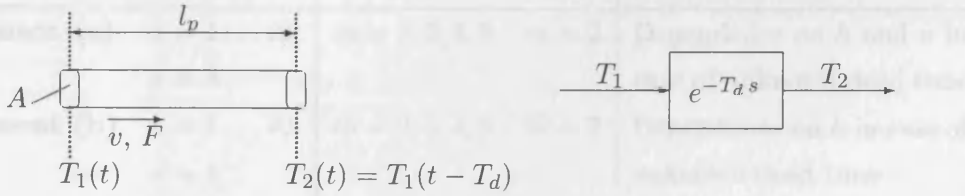


Figure 8.1: Signal dead time: Temperatures along a tube with flow rate F, velocity v. Right: block diagram.

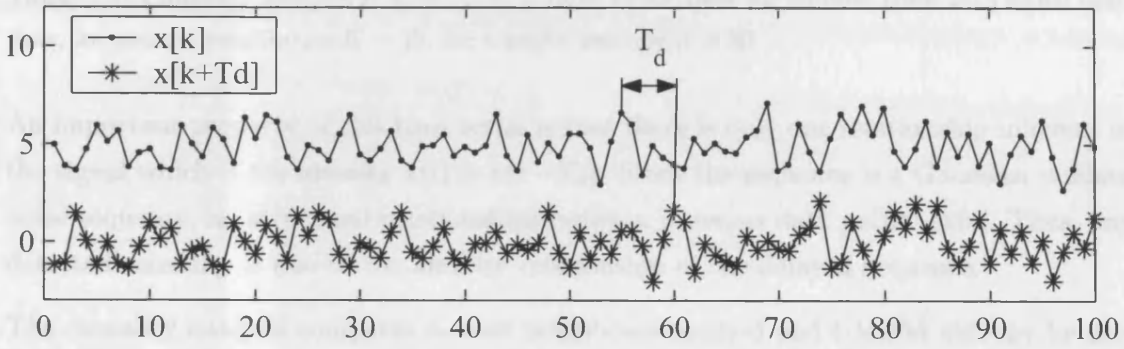


Figure 8.2: Gaussian random noise signal with signal dead time of $T_d = 5$ samples.

Thus, an event happening first in a measurement upstream can be observed a certain time later in a measurement further downstream. The signal dead time T_d between the event in the first and second measurement is usually unknown.

For example, consider a pipeline of length l_p and area A, as shown in Figure 8.1. The signal dead time T_d depends on the volumetric flow rate $F = A\,v$, where v is the velocity of the liquid, as well as l_p and A. If all three parameters are constant over the area of time and the transported mass is a compressible liquid, then the signal dead time can be expressed as

$$T_d = \frac{A\,l_p}{F}. \tag{8.1}$$

In general, the flow rate F is unknown and also not necessarily constant. Determining all parameters through measurement would be feasible for some conditions but the throughput in the plant might change thus affecting the flow rate. Other equipment such as tanks and distillation columns also act as dead time elements to some extent. As a note it is remarked here that pressure measurements usually do not exhibit a signal dead time since pressure variations propagate at the speed of sound. Thus, the dead time is a fraction of a second which cannot be measured with standard sampling rates of a few seconds to a few minutes.

The time trend of a random noise signal delayed by a signal dead time T_d of five samples is shown in Figure 8.2. Shown here are 100 samples, for transfer entropy 2000 samples and for the nearest neighbours method 400 samples are analysed. This simulated data is used for investigating the impact of signal dead time on the causality measures. The result for all measurements should be that the original time series causes the delayed time series.

		NN	TE	Aim
Experiment (a)	$h = 1 \ldots 10$; $\kappa = h$	$m = 2, 3, 4, 5$	$m = 2$	Dependence on h and κ in case of unknown dead time
Experiment (b)	$h = 1 \ldots 10$; $\kappa = 1$	$m = 2, 3, 4, 5$	$m = 2$	Dependence on h in case of unknown dead time
Experiment (c)	$h = T_d$ $\kappa = 1 \ldots 10$	$m = 2, 3, 4, 5$	$m = 2$	Dependence on κ in case of known dead time

Table 8.2: Causality analysis as structured in three experiment for random noise with signal dead time, for nearest neighbours $K = 15$, for transfer entropy $n = 50$.

An important property of this time series is that there is only one relationship inherent in the signal which is the identity $x(t) = y(t - T_d)$. Since the sequence is a Gaussian random noise sequence, no structural relationships between previous data points exist. Thus, any detected causality is due to the identity relationship of the delayed sequence.

The causality analysis compares nearest neighbours method and transfer entropy by carrying out the three experiments listed in Table 8.2 and referred to as Experiment (a), (b), (c). The causality measure using the nearest neighbours method is calculated for an embedding dimension of $m = 2, 3, 4, 5$. The transfer entropy method implemented here is structurally similar to the nearest neighbours method with an embedding dimension $m = 2$ since the transition probability length is set to $l = 2$ in definition 7.30. The results for transfer entropy and nearest neighbours are expected to be similar since the two methods have a related structure and both use the concept of predictability. In the experiment, the prediction horizon h and the algorithm time delay κ are varied to find out if the directionality can be detected by the presence of signal dead time. The prediction horizon h for both transfer entropy and nearest neighbours method was the distance of the future value to be predicted to the embedded vector which is used to predict the future value. The algorithm time delay κ is used for the construction of the embedded vectors and specifies the distance between the values of the embedded vector as given in Equations 6.1 and 7.30.

Experiment (a) In this experiment, prediction horizon h and algorithm time delay κ are dependent and varied jointly from one to ten. The consequence is that the causality is only captured if h and κ are exactly the signal dead time $T_d = 5$ since there is only a relationship between the ith sample of the first sequence and the $i + T_d$th sample of the second sequence. This effect can be seen in Figure 8.3. An exception is the causality measure using the nearest neighbours method for embedding dimension $m = 5$ and a prediction horizon $h = 1$. The reason for this is that the embedded vector captures just the previous measurement that is able to predict the future value of the delayed time series. It can be observed for the nearest neighbours method that the lower the embedding dimension the higher the significance level. The reason is that due to the special construction of the

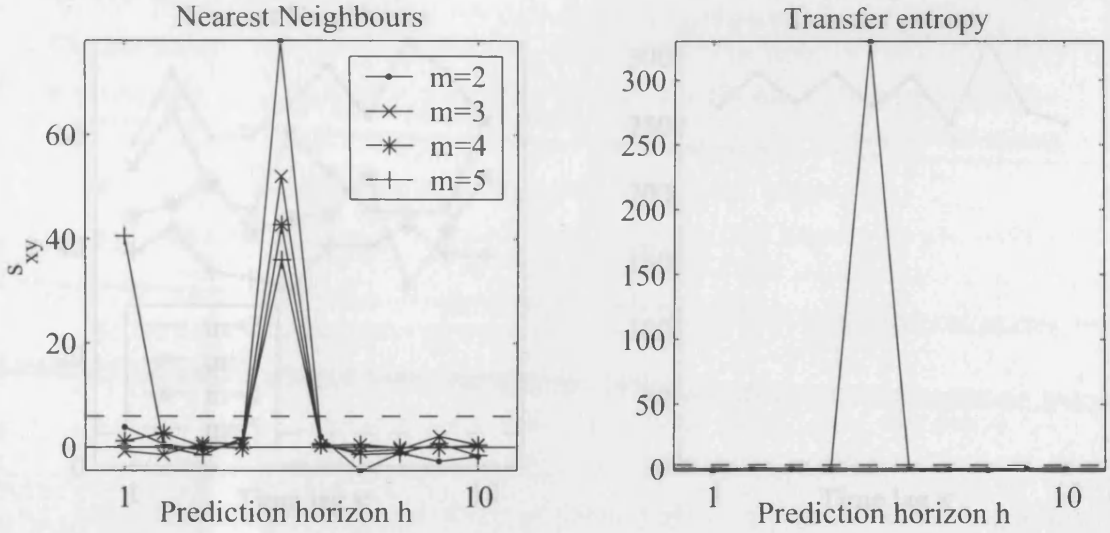


Figure 8.3: Experiment (a): Results of nearest neighbours and transfer entropy significance level for random noise with signal dead time; $h = \kappa$ and $T_d = 5$ samples.

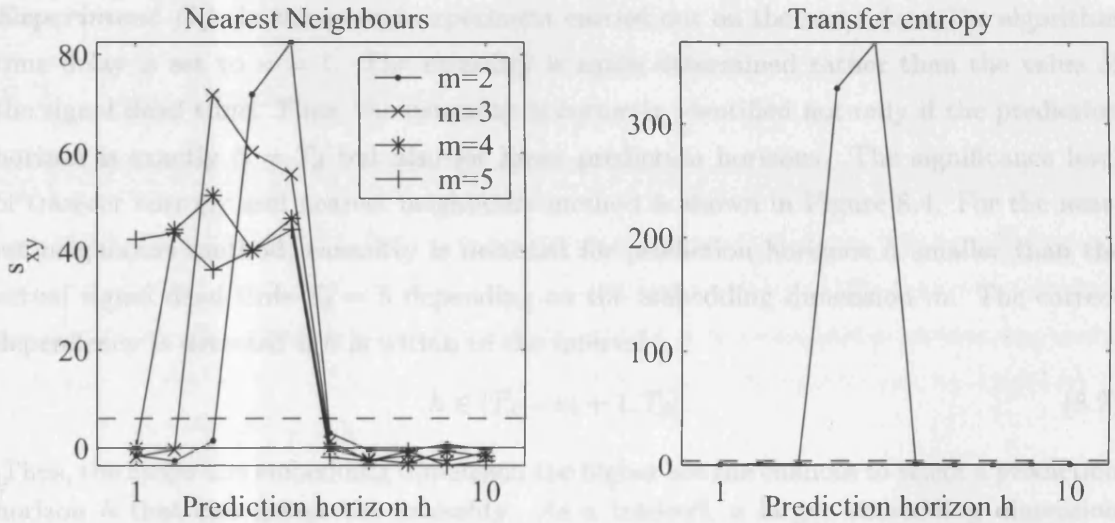


Figure 8.4: Experiment (b): Results of nearest neighbours and transfer entropy significance level for random noise with signal dead time; $\kappa = 1$ and $T_d = 5$ samples.

measurements, only one data sample predicts the delayed sequence. Any additional sample interferes the prediction and worsens the result. The amplitude of the significance level is extremely high for both nearest neighbours and transfer entropy, exceeding up to ten times the threshold in case of nearest neighbours and 100 times the threshold in case of transfer entropy. The reason for the better result of transfer entropy could be that more samples, 2000 instead of 400 for the nearest neighbours. The number of samples was chosen differently so that the computation time is constant for both methods. To verify this hypothesis, 2000 samples were used for the nearest neighbours method. The causality measure of the nearest neighbours method increased to the same dimension as the transfer entropy measure, for a $h = \kappa = T_d$ the significance level of the nearest neighbours method is $s_{x \to y} = 228$ and for the transfer entropy method $s_{x \to y} = 321$.

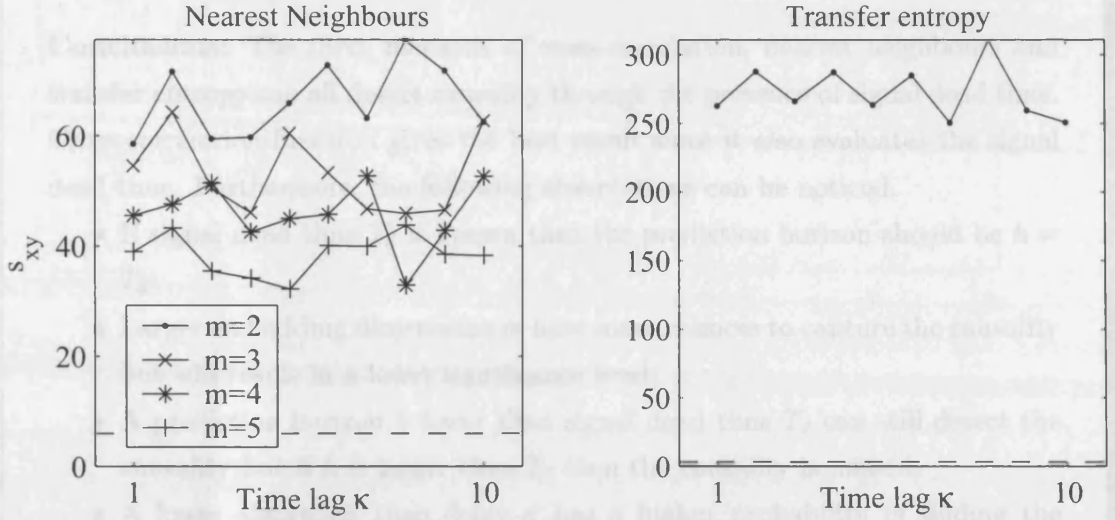


Figure 8.5: Experiment (c): Results of nearest neighbours and transfer entropy significance level for random noise with signal dead time; $h = T_d = 5$ samples.

Experiment (b) In the second experiment carried out on the same data, the algorithm time delay is set to $\kappa = 1$. The causality is again determined rather than the value of the signal dead time. Thus, the causality is correctly identified not only if the prediction horizon is exactly $h = T_d$ but also for lower prediction horizons. The significance level of transfer entropy and nearest neighbours method is shown in Figure 8.4. For the nearest neighbours method, causality is detected for prediction horizons h smaller than the actual signal dead time $T_d = 5$ depending on the embedding dimension m. The correct dependency is detected if h is within in the interval

$$h \in [T_d - m + 1; T_d]. \tag{8.2}$$

Thus, the larger the embedding dimension the higher are the chances to select a prediction horizon h that can detect the causality. As a tradeoff, a larger embedding dimension has a lower significance level. Since transfer entropy here is equivalent to an embedding dimension of $m = 2$, a correct causality can only be detected if $h = T_d - 1$ or $h = T_d$. The similar shape of both results for $m = 2$ in Figure 8.4 again highlights the relationships between the two methods.

Experiment (c) In the final experiment, the prediction horizon is set to the signal dead time T_d. The algorithm time delay κ is now varied from one to ten as shown in Figure 8.5. For these settings, the causality is correctly identified independently of algorithm time delay κ. The reason for this is that the only dependency is the identity $x(t) = y(t - T_d)$ which is captured with the setting of the prediction horizon to the signal dead time. The same result is achieved for both nearest neighbours and transfer entropy. Again, a lower embedding dimension m results in a higher significance level.

When assuming that the signal dead time is the only indication of a cause and effect relationship then the following guidelines should be considered when applying transfer entropy and the nearest neighbours measure.

Conclusions: The three methods of cross-correlation, nearest neighbours and transfer entropy can all detect causality through the presence of signal dead time. Cross-correlation function gives the best result since it also evaluates the signal dead time. Furthermore, the following observations can be noticed.

- If signal dead time T_d is known then the prediction horizon should be $h = T_d$;
- Larger embedding dimensions m have more chances to capture the causality but will result in a lower significance level;
- A prediction horizon h lower than signal dead time T_d can still detect the causality but if h is larger than T_d then the causality is missed;
- A lower algorithm time delay κ has a higher probability of finding the dependency if embedding dimension m is large enough.

8.2 Low Pass Filtering

In this section, the effect of the low pass filtering of process equipment on the causality measures are investigated. Items of process equipment that show low pass behaviour are also referred to as first-order systems. In Stephanopoulos [126], first-order systems are described. An example of such a system is a tank with an inflow stream w_{in} and a resistance in the outflow. Thus, the dynamic behaviour of the system can be described by the following differential equation:

$$A\,R\frac{\mathrm{d}h_l}{\mathrm{d}t} + h_l = R\,w_{\text{in}} \tag{8.3}$$

where h_l is the level in the tank, A the area, R the resistance of the outflow and w_{in} the inflow. Transforming the differential equation into the Laplace domain with h_l as output and w_{in} as input gives the following transfer function that has the form of a first order system:

$$G_c(s) = \frac{H_l(s)}{W_{\text{in}}(s)} = \frac{K_c}{\tau_D s + 1}, \qquad \text{where } K_c = R,\ \tau_D = AR. \tag{8.4}$$

Low pass filtering of time data for the purposes of simulation can be realised in a number of ways, for example through transformation into the frequency domain followed by elimination of high frequency components and back transformation into the time domain. For discrete time samples, a straightforward time domain moving average (MA) realisation can be expressed as follows

$$\text{filt}(x_i, n_f) = \frac{1}{n_f}\sum_{\nu=-n_f/2}^{n_f/2} x_{\nu-i} \tag{8.5}$$

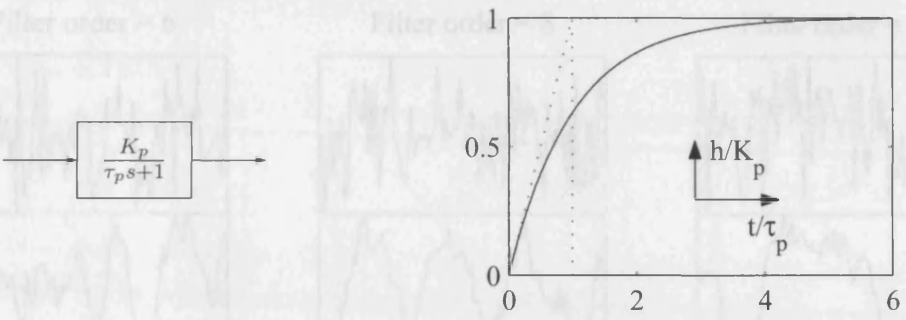


Figure 8.6: Transfer function (left) and step response (right) of a first order low pass system.

where n_f is the order of the filter. The filter is acausal since future values of x are used to compute the average. The acausality does not cause any difficulties since the computation is done offline. This filter implementation ensures that there is no time delay introduced which is useful when studying only the filtering effect on the causality measure.

The simulated data set which will be used for testing the causality measures is constructed in the following way. A random noise signal with zero mean and unit variance ($N(0,1)$) is filtered with a low pass filter. Three filter orders are implemented, $n_f = 6$, $n_f = 8$ and $n_f = 12$. The filtered signal is filtered twice again so that four signals in total exist, x_0, x_1, x_2 and x_3:

$$\begin{aligned} x_0 &= N(0,1) \\ x_1 &= \text{filt}(x_0, n_f) \\ x_2 &= \text{filt}(x_1, n_f) \\ x_3 &= \text{filt}(x_2, n_f) \end{aligned} \tag{8.6}$$

The time trends of these signals are shown in Figure 8.7. The time trend of the random signal x_0 is similar in all three cases. The filtered signal after the second filtering has features of a sinusoidal oscillation. For higher filter order, the oscillation period is different depending on the cut-off frequency of the filter is a function of the filter order. The filtered signals x_2 and x_3 have a very similar shape since all high frequency components still left in signal x_1 have already been filtered out. Further filtering would only smooth further slightly. It is expected that the random data x_0 inhibits more uncertainty than the filtered signals. The causality measures which estimate the reduction of uncertainty should therefore detect a cause and effect relationship from x_0 to x_1 to x_2 to x_3:

The reason of the experiment using these four data samples is to see whether transfer entropy or the nearest neighbours method is better suited to detect the causality and to investigate the impact of the filter order on the causality measures.

Experiment: When applying the causality measures both prediction horizon h and time delay κ are set equal to one because of the independence of the data points of the original random time sequence. In case of a real life oscillation, the parameters in Table 7.2 would

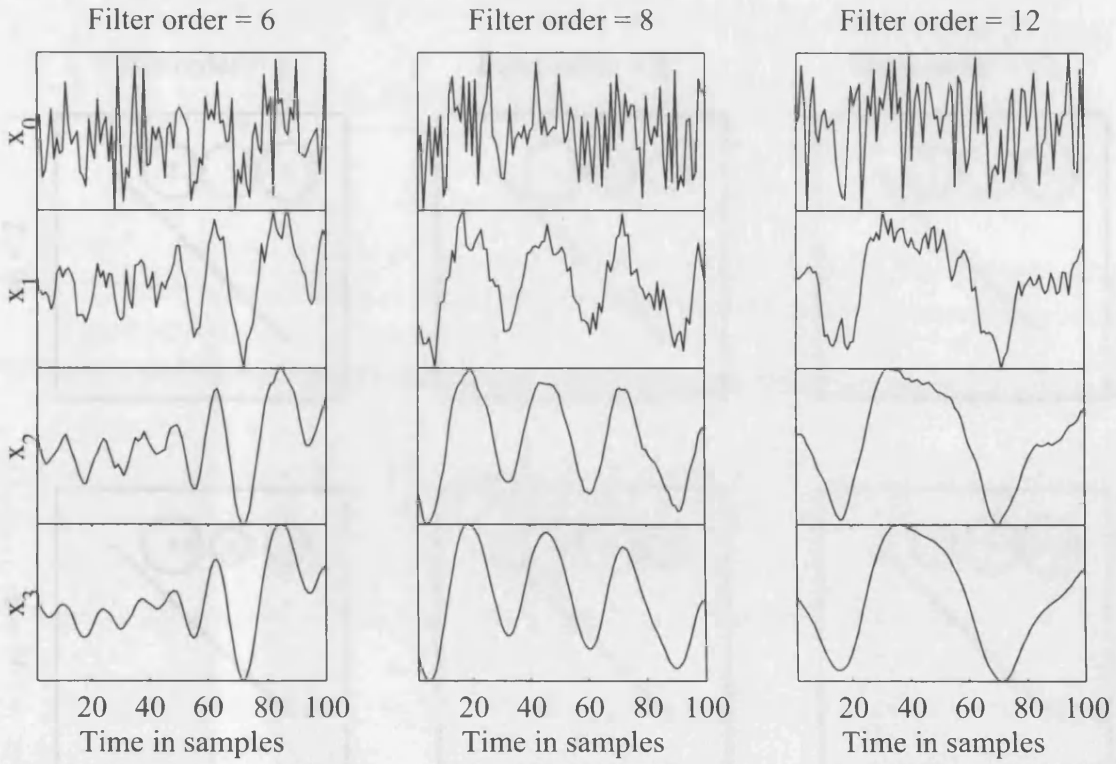


Figure 8.7: Filtered Gaussian random noise signal with varying filter order $n_f = 6, 8, 12$.

be preferred. Again, 2000 samples are used for the transfer entropy and 400 samples for the nearest neighbours method which result in approximately equal computational effort. The results of the nearest neighbours method are displayed in Figure 8.8 for all three filter orders and a variable embedding dimension of $m = 2, 3, 4$. For almost all filter orders and all embedding dimension the causal relationship between the original random noise sequence x_0 and the filtered signals is detected. However, the expected causality between x_1 and x_2, x_3 as well as x_2 and x_3 does not show up in Figure 8.8, that is, there is no line of bubbles in the second or third row in the bubble charts. Also, the significance level is much lower than for the random noise signal with a dead time. The causality measures using transfer entropy give better results as shown in Figure 8.9. The significance level is higher compared to the nearest neighbours method and additionally the dependencies between x_1 and x_2, x_3 are detected. The results do not vary in a coherent way with the filter order but exceed the threshold of three for all filter orders. The causal relationship between x_2 and x_3 does again not show up. The reason for this is probably the similarity between the time trends which are almost indistinguishable.

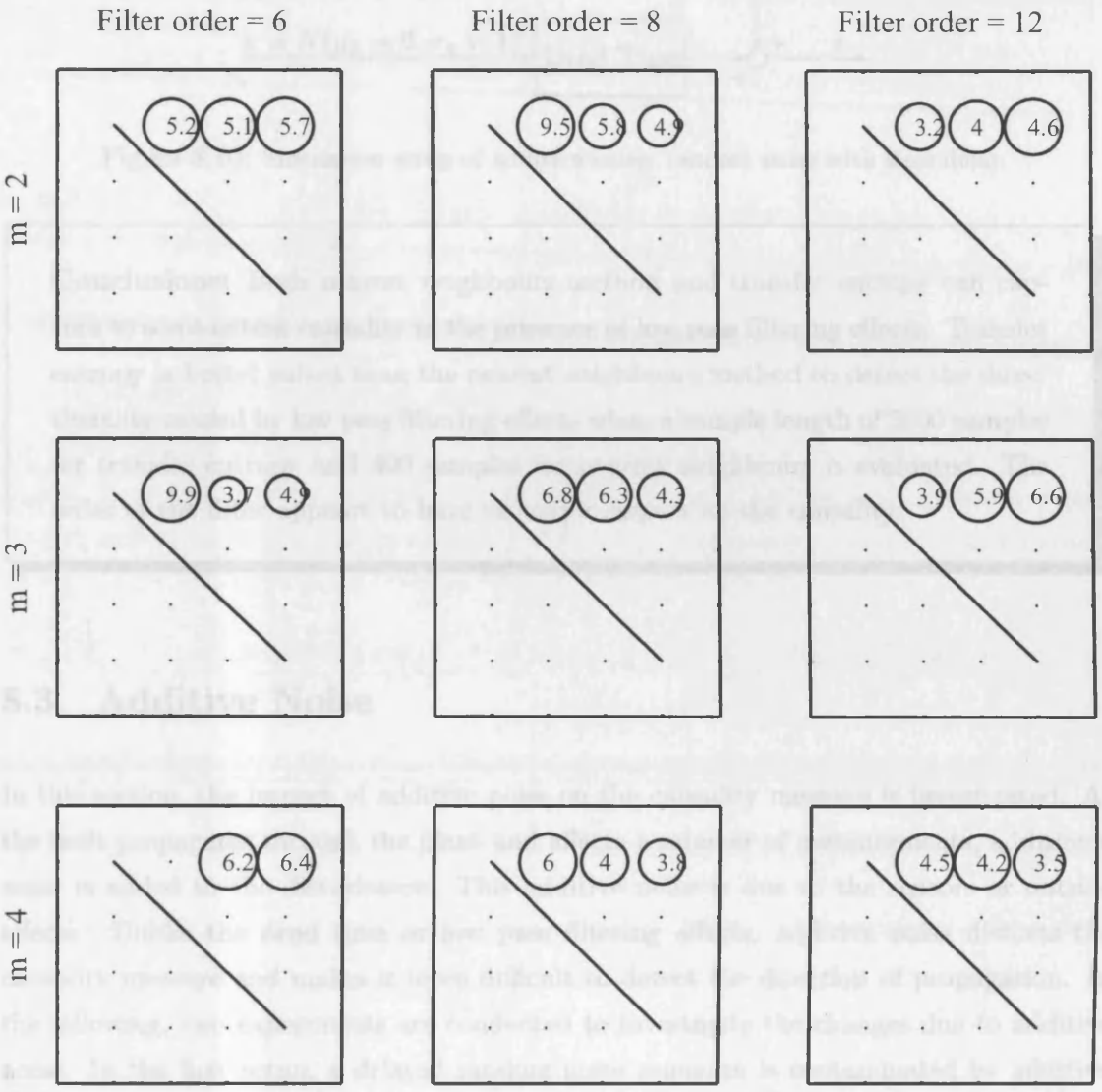


Figure 8.8: Significance level of nearest neighbours method applied to low pass filtered signal with different filter orders and varying embedding dimension; $h = \kappa = 1$.

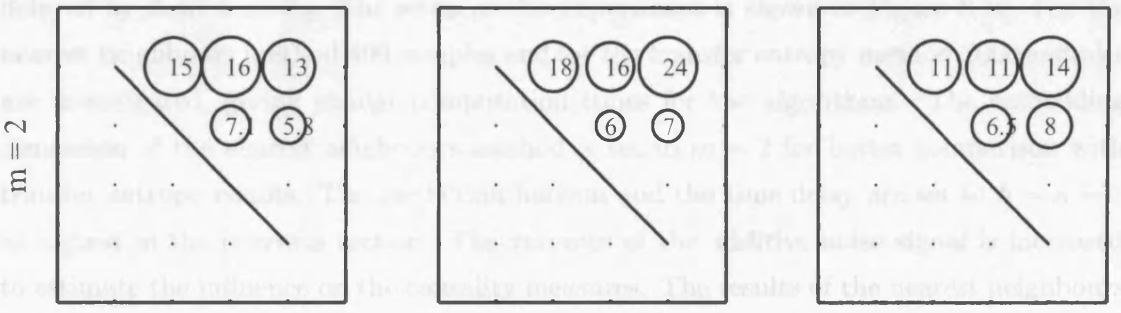


Figure 8.9: Results of transfer entropy applied to low pass filtered signal; $h = \kappa = 1$. Left plot: filter order = 6, middle plot: filter order = 8, right plot: filter order = 12.

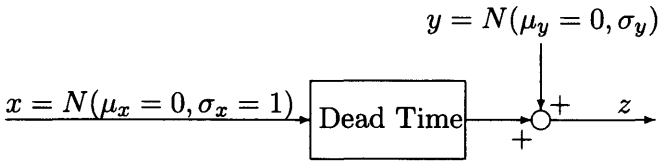


Figure 8.10: Simulation setup of additive noise: random noise with time delay.

Conclusions: Both nearest neighbours method and transfer entropy can capture to some extent causality in the presence of low pass filtering effects. Transfer entropy is better suited than the nearest neighbours method to detect the directionality caused by low pass filtering effects when a sample length of 2000 samples for transfer entropy and 400 samples for nearest neighbours is evaluated. The order of the filter appears to have no major impact on the causality.

8.3 Additive Noise

In this section, the impact of additive noise on the causality measure is investigated. As the fault propagates through the plant and affects a number of measurements, additional noise is added to the disturbance. This additive noise is due to the sensors or outside effects. Unlike the dead time or low pass filtering effects, additive noise distorts the causality measure and makes it more difficult to detect the direction of propagation. In the following, two experiments are conducted to investigate the changes due to additive noise. In the first setup, a delayed random noise sequence is contaminated by additive noise and the causality between the original noise signal and the contaminated signal is studied. In addition to the nearest neighbours and transfer entropy method the cross-correlation function is also investigated. In the second setup, the delay is replaced by a low pass filter.

Experiment (a) In the first experiment, additive noise contaminates signal previously delayed by dead time T_d. The setup of this experiment is shown in Figure 8.10. For the nearest neighbours method 400 samples and for the transfer entropy method 2000 samples are investigated, giving similar computation times for the algorithms. The embedding dimension of the nearest neighbours method is set to $m = 2$ for better comparison with transfer entropy results. The prediction horizon and the time delay are set to $h = \kappa = 1$ as argued in the previous section. The variance of the additive noise signal is increased to estimate the influence on the causality measures. The results of the nearest neighbours and transfer entropy method are then shown in Figure 8.11. The measured causality is in both cases decreasing with increased noise level. However, a causality can be detected for a additive noise variance $\sigma_y = 3$ that is three times higher than the variance of the

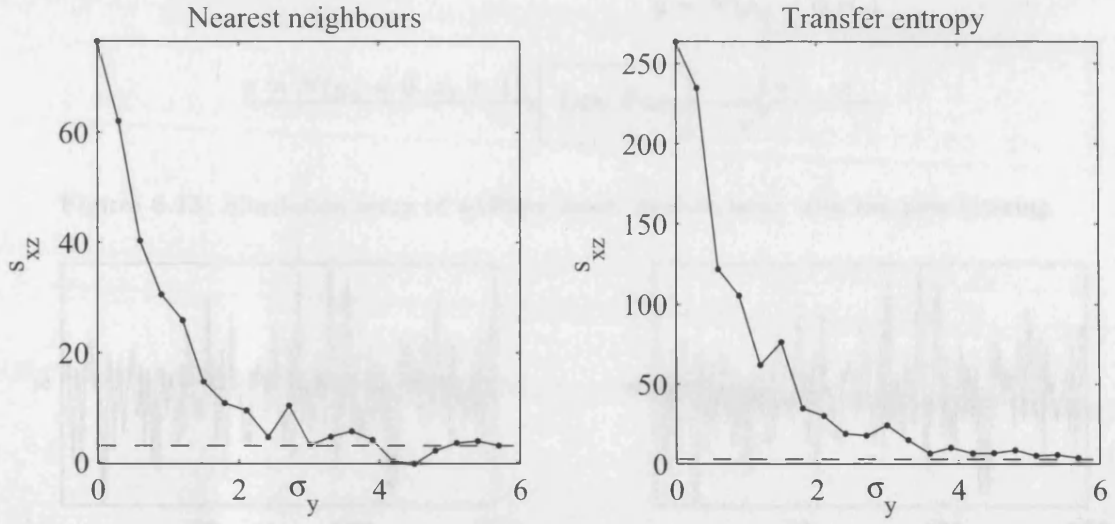


Figure 8.11: Causality analysis between random noise signal and the same delayed by dead time T_d and with additive random noise with variance σ_y: Left panel: nearest neighbours results ($h = \kappa = 1$, $m = 2$), right panel: transfer entropy results ($h = \kappa = 1$).

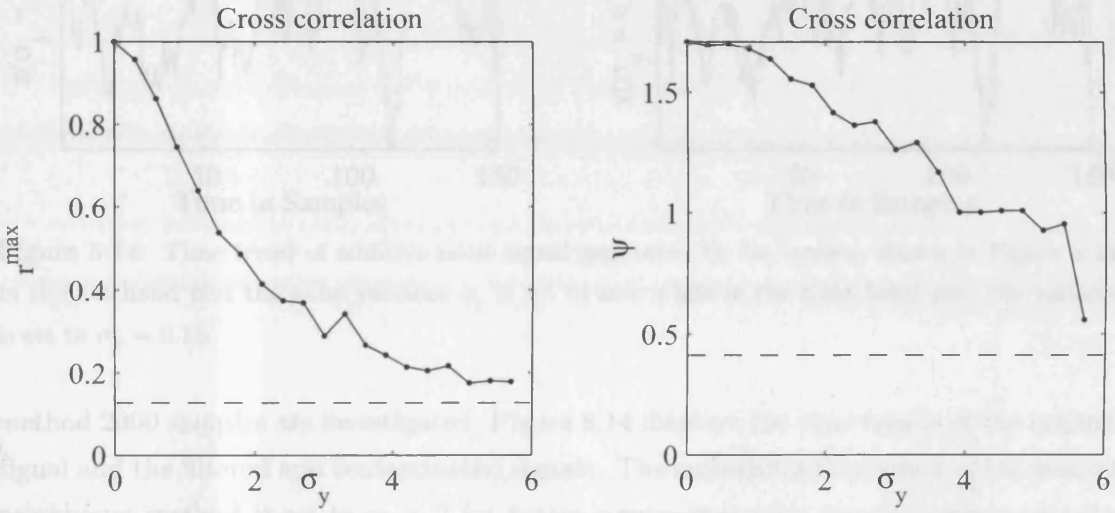


Figure 8.12: Causality analysis between random noise signal and the same delayed by dead time T_d and with additive random noise with variance σ_y: Cross-correlation results.

original signal ($\sigma_x = 1$) for the nearest neighbours method and six times higher for transfer entropy. The results of the cross-correlation function have a similar behaviour as shown in Figure 8.12. The relationship between the variance σ_y of the added noise and the cross-correlation method is shown in Figure 8.12. The time delay was detected correctly ($T_d = 10$) for all values of σ_y up to $\sigma_y = 6$. Both correlation index is shown in the left hand panel of Figure 8.12 and oscillation index ψ (right hand panel) are above the threshold r_{thresh} and ψ_{thresh}, indicated by dashed line. Thus, the correlation methods is very robust against additive noise in case of dead time.

Experiment (b) In this setup, a random noise signal is filtered with a lowpass filter before a further noise signal is added as shown in Figure 8.13. The filter order is set to $n_f = 6$. For the nearest neighbours method 400 samples and for the transfer entropy

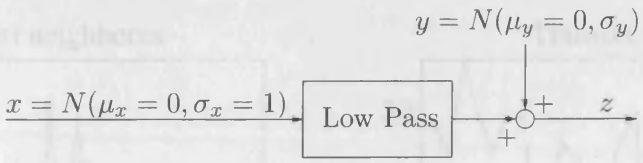


Figure 8.13: Simulation setup of additive noise: random noise with low pass filtering.

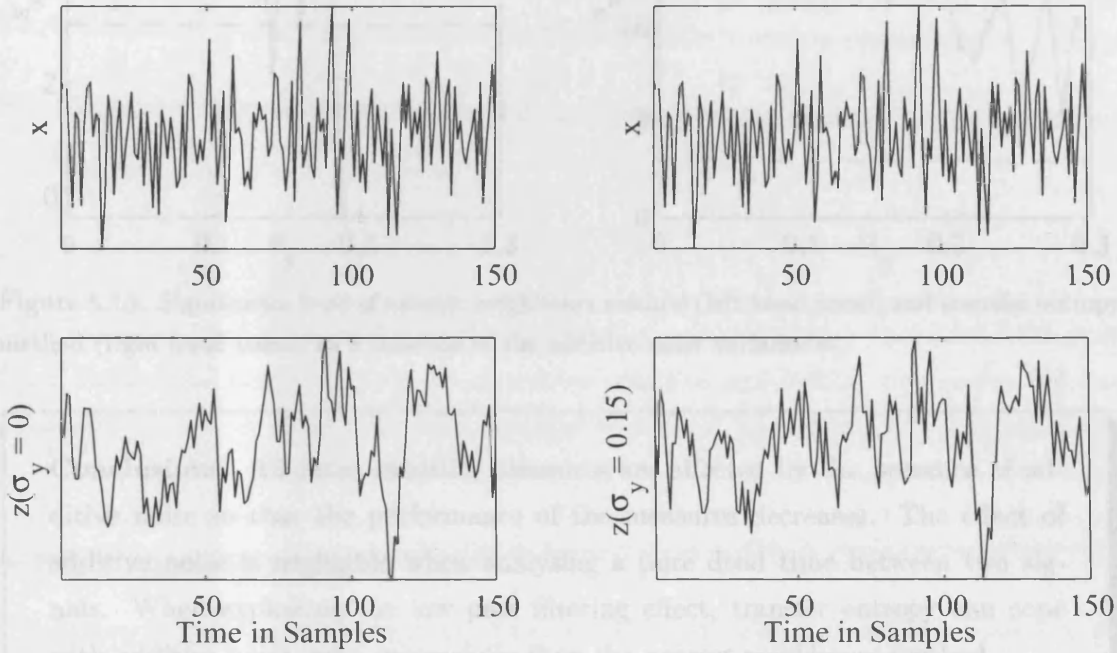


Figure 8.14: Time trend of additive noise signal generated by the system shown in Figure 8.13. In the left hand plot the noise variance σ_y is set to zero while in the right hand plot the variance is set to $\sigma_y = 0.15$.

method 2000 samples are investigated. Figure 8.14 displays the time trends of the original signal and the filtered and contaminated signals. The embedding dimension of the nearest neighbours method is set to $m = 2$ for better comparison with transfer entropy results. The prediction horizon and the time delay are set to $h = \kappa = 1$, the number of nearest neighbours is set to $K = 15$. The variance of the additive noise signal is increased to estimate the influence on the causality measures. The results of the nearest neighbours and transfer entropy methods are shown in Figure 8.15. For the nearest neighbours method, the causality measure s_{xz} between the original signal x and the output signal z only gives a correct result if the variance σ_y is lower than 0.12. The measure s_{xz} is more robust for the transfer entropy method which gives the correct direction for a variance $\sigma_y = 0.3$. In summary, the effect of additive noise is stronger when investigating the low pass filtering effect than the dead time.

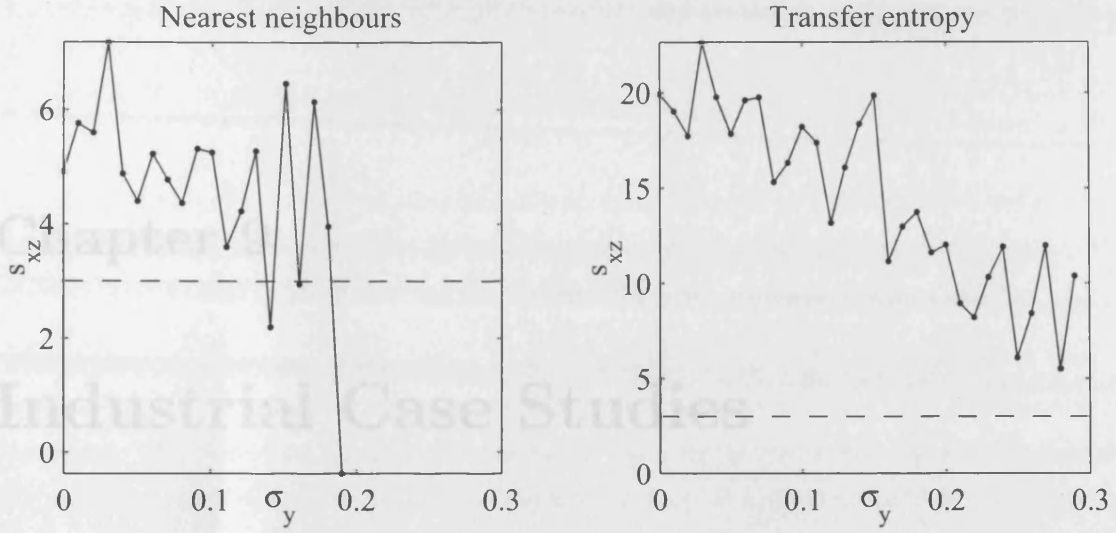
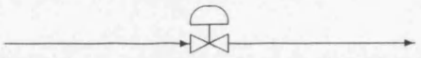


Figure 8.15: Significance level of nearest neighbours method (left hand panel) and transfer entropy method (right hand panel) as a function of the additive noise variance σ_y.

Conclusions: All three causality measures are affected by the presence of additive noise so that the performance of the measures decreases. The effect of additive noise is negligible when analysing a pure dead time between two signals. When exploiting the low pass filtering effect, transfer entropy can cope with additive noise more successfully than the nearest neighbours method.

Chapter 8 Summary

The impact of fault propagation effects on the developed causality measures has been studied in this chapter. The effects of dead time, low pass filtering and additive noise were investigated using simulated noise data. All three measures detect the causality based on dead time. Transfer entropy deals best with low pass filtering effects while cross correlation is most robust against additive noise if the causality is argued by the dead time between the two signals.



Chapter 9

Industrial Case Studies

This chapter discusses the application of the causality measures to data sets from industrial processes. Recommendations and guidelines are given for the acquisition of the data sets and interpretation of the causality results. Two new case studies are introduced exhibiting data characteristics different from the reference case study. The root cause is hypothesised and the consistency of the three methods investigated.

The application of the causality measures to simulated data in the previous chapter investigated the impact of fault propagation mechanisms on each measure of the three causality measure. The question of directionality is, however, important for plant-wide disturbances and large data sets. The selection of relevant variables and the capture of the correct time frame when the disturbance occurs is important for the conduction of the analysis. Unlike simulated data, real industrial data has a certain dynamic structure even if a time trend appears to be random.

In the first section, the acquisition of data and the general interpretation of the causality results are discussed. Section 9.2 introduces a process at Eastman Chemical Company that showed an irregular disturbance in a number of time trends while Section 9.3 describes a process at BP Chemicals with a linear oscillation. The causality measures of cross-correlation, nearest neighbours and transfer entropy are applied to the two processes.

9.1 Selection of Case Studies

Before applying the causality measures to a data set it is useful to check the data set for validity and carefully phrase the causality questions to be solved. In this section, a checklist of topics to be considered before the analysis is provided. First, the data set has to meet some requirements explained in Section 9.1.1. The directionality investigated has to be stated before applying the methods and some guidelines are given in Section 9.1.2.

Assisting information for verification of the results and measures is given in Section 9.1.3.

9.1.1 Acquisition of Data Sets

When capturing the data from acquisition system such as PI or directly from the DCS, a number of checks have to be carried out to ensure that the causality analysis gives useful results. The following requirements are necessary when preparing the analysis.

(A) Selected Process Variables A prerequisite for the causality analysis is that a number of process variables are selected that show the same features of a disturbance in the data. The detection of disturbances using data-driven methods is a solved problem for most types of disturbances. Detection can be carried out using oscillation detection or principal component analysis, see Section 3.3. The reason for analysing such reduced data sets is to limit the number of measurements. Comparing p measurements requires $p(p-1)$ directionality measures to be calculated and interpreted. It is therefore sensible to restrict the analysed measurements to those that show the presence of the disturbance to be investigated.

(B) Onset and ending of a disturbance Time periods during which the disturbance is present should be separated from time periods that appear as normal operating conditions. One reason for this is that the cause and effect relationship might only be present during the upset. Furthermore, the statistical tools require stationarity. The statistical properties usually change during the period of disturbance. The start and end of a disturbance can be detected through increased variability or limit crossings. In addition, the start and end of the disturbance should be captured which might give further conclusions towards the root cause.

(C) Process data The result of the causality analysis depends significantly on the quality of the data. Most importantly, data of the most important process variables should be available without too much data missing. In general, flow measurements are particularly interesting since the flow is in most cases the adjusted variable and the disturbance tends to propagates along the direction of flow. The quality of the data is also a requirement, that is, compression and quantisation should not be an issue. Cleaning outliers from the data is necessary for transfer entropy as the construction of the probability density function places its interval from the smallest to the largest value. Thus, the number of bins are computed incorrectly in the presence of outliers.

(D) Sampling rate The sampling rate has to be sufficiently high in order to capture the relevant features in the measurements. Measurements are usually recorded once a minute if captured with the PI historian, data from the distributed control system range between 10 and 20 second data. Subsampling might be necessary prior to the analysis when analysing low frequency features.

9.1.2 Questions on Directionality

After selecting the measurements and the time frame of the analysis a hypothesis on the direction of fault propagation can be phrased. The advantage of having a hypothesis in mind is that the interpretation of the causality measure is easier. Alternatively, the automated construction of a causal map as described in Section 4.4 can be utilised. Specific questions on directionality are for example:

- Can the directionality measures identify the correct direction of flow as given in the process schematic?
- Does directionality change if a disturbance is present compared to normal operation?
- Is there a cause and effect between controller output and process variable?
- Is there a causality between the same type of measurements, e.g. from temperature to temperature?

9.1.3 Interpretation of Results

In order to adjust the causality measures and to verify the results, additional information about the case study is helpful. If a quantitative model of the process exists, the data driven methods can be compared to the results obtained from cause and effect digraphs constructed from the model. Furthermore, the following information can be used to judge the result obtained from the causality analysis.

(A) Complete process schematics Information can be obtained by skillfully interpreting the process schematic. The process schematic contains all relevant equipment and product flows and the positions of the measurements are indicated alongside control instrumentation. The direction of flow detected with the data-driven causality measure can be confirmed from the direction of flow indicated in the process schematic.

(B) Process knowledge available Process knowledge of expert process engineers can be used to argue the information flow and confirm the root cause. Further insight into the process can help to understand the results obtained from the causality analysis.

(C) Post-analysis data set A data set captured after the disturbance has been eliminated can highlight the impact of the disturbance. Cases in which the disturbance went away after fixing the cause detected by cause and effect analysis verify the results in a constructive way.

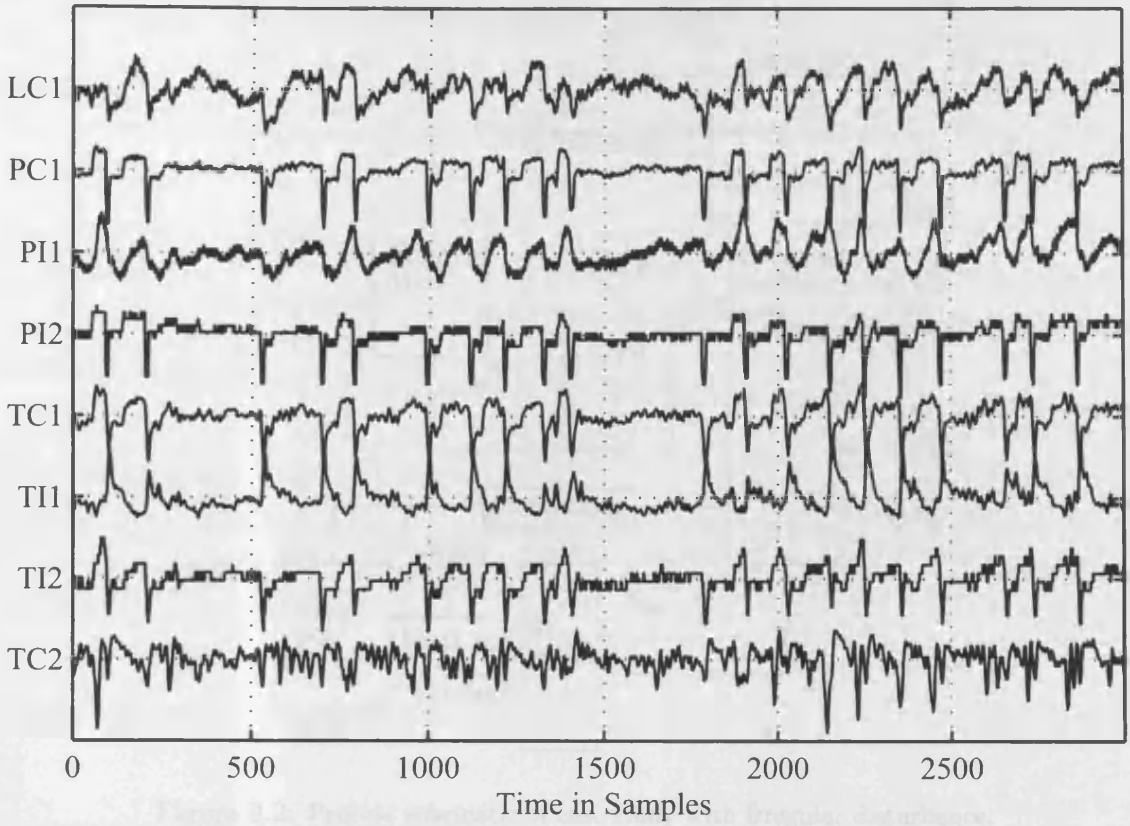


Figure 9.1: Time trends of irregular disturbance with sharp bursts.

9.2 Irregular Disturbance (Eastman Chemical Company)

This case study is selected because unlike the reference case study, the plant-wide disturbance affecting this process is of non-oscillatory nature. Irregular spikes appear in eight measurements with the peaks or valleys being clearly distinguishable (for the time trend see Figure 9.1). It is therefore expected that the dead time is a dominant effect from which cause and effect can be argued as well as the low pass filtering that changes the shape of the spikes. Since most data-driven analysis tools are mapped out for oscillatory disturbances, irregular disturbances are a challenge to these methods. In this section, the causality measures are applied to the process data to investigate the ability to analyse an irregular disturbance. The process is an acid production process at Eastman Chemical Company as shown in Figure 9.2. The disturbance was present for a period over a few weeks. The process data, schematic and insight were provided by Michelle Caveness.

The centrepiece of the process part is a reactive distillation column in which the product is processed as shown in Figure 9.2. The overhead outflow of the column is recycled in the condenser and reflux tank and fed back to the column. The pressure in the column is thereby controlled through the reflux flow and PC1. To prevent the reflux tank from overflowing an exit line outflow is supervised by LC1. The temperature in the column is critical for the chemical reaction to take place and therefore not only controlled by TC1 and TC2 but also observed by TI1 and TI2. As in all distillation processes, a pressure and

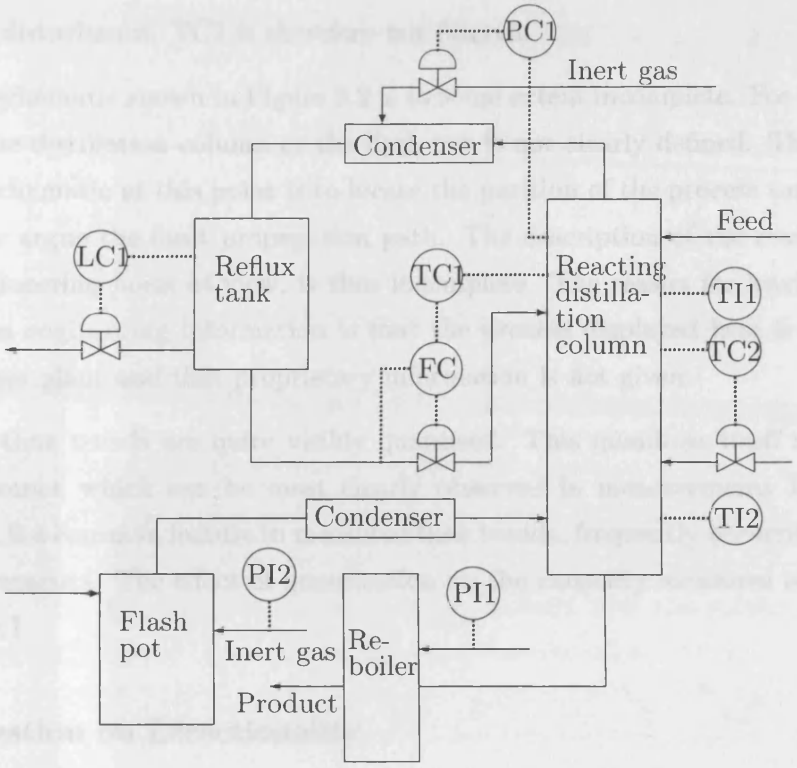


Figure 9.2: Process schematic of case study with irregular disturbance.

temperature gradient can be observed along the column with pressure and temperature being functionally related. The pressure of inflows of steam and inert gas further on in the process is monitored by PI1 and PI2 respectively.

9.2.1 Process and Data Set

In the selected part of the process a disturbance affects the measurements around the acid column in Figure 9.2. The time trend of the disturbance is shown in Figure 9.1 with 3000 samples taken at a rate of one every 10 seconds. Altogether, 8640 samples were available but only 3000 samples are displayed here for clarity. A common pattern can be seen in all time trends. Sharp spikes occur around the same time instance in the level, pressure and temperature measurements although with some interruptions and thus not necessarily regular. The fact that the spikes are part of the same disturbance can best be seen by visual inspection. The tips of the spikes line up which can be best seen when observing TC1 and TI1. Furthermore, a longer period of normal operation following sample 1500 is common to all time trends. Because of the irregular pattern Fourier analysis or autocorrelation will give no useful results. Instead, for example spectral principal component analysis [134] or, even better, time shifting principal component analysis [144] can be used to detect the disturbance. Level measurement LC1 is distorted by noise and therefore filtered with a sixth order low pass frequency filter as described in Equation 8.6. The controller temperature TC2 showed some additional noise in the same frequency

range as the disturbance. TC2 is therefore not filtered.

The process schematic shown in Figure 9.2 is to some extent incomplete. For example, the inflow into the distillation column or the flash pot is not clearly defined. The purpose of the process schematic at this point is to locate the position of the process variables which are crucial to argue the fault propagation path. The description of the reaction, from a chemical engineering point of view, is thus incomplete. The reason for leaving out some of the process engineering information is that the process displayed here is only a small part of a larger plant and that proprietary information is not given.

Some of the time trends are quite visibly quantised. This manifests itself in the square shape appearance which can be most clearly observed in measurements PI2 and TI2. Quantisation is a common feature in measured time trends, frequently occurring in temperature measurements. The effect of quantisation on the causality measures is investigated in Section 10.1.

9.2.2 Question on Directionality

The root cause of the disturbance was unknown prior to the analysis. Thus, the question of directionality is which variable affected the other variables in order to find the order of occurrence and then consequentially the root cause. If all dependencies are identified, a causal map can be constructed automatically as described in Section 4.4. The most likely root cause according to the causality analysis will be the first variable in the string of variables in the causal map.

9.2.3 Interpretation of Results

In the following, the three causality measures of cross-correlation, nearest neighbours and transfer entropy are applied to the data set. A causal map will be derived from the causality matrix and a possible explanation of root cause discussed using the process schematic in Figure 9.2 and additional expert knowledge.

Cross-Correlation Function: The method of causality using cross-correlation function was described in Chapter 5. The following table gives the sample number of the detected dead time.

x / y	LC1	PC1	PI1	PI2	TC1	TI1	TC2	TI2
LC1	-		15			3	1	11
PC1	4	-	19		4	6	4	14
PI1			-					
PI2	5	1	22	-	5	7	5	16
TC1			10		-	2		9
TI1						-		7
TC2			4			1	-	9
TI2			8					-

Only results of the detected dead time are considered that have a correlation index $r^{\max}$ larger than the threshold r_{thresh} and an oscillation index ψ larger than ψ_{thresh}. For this purpose, 3000 samples are considered so that $r^{\max} > 0.11$ after Equation 5.25 and $\psi > 0.38$ after Equation 5.26. For the generation of the causal map, non-zero entries of the matrix are considered and ordered according to the algorithm described in 4.4.

x / y	LC1	PC1	PI1	PI2	TC1	TI1	TC2	TI2
LC1	-		1			1	1	1
PC1	1	-	1		1	1	1	1
PI1			-					
PI2	1	1	1	-	1	1	1	1
TC1			1		-	1		1
TI1						-		1
TC2			1			1	-	1
TI2			1					-

It can be seen here that PI2 is preceding the seven other measurements since the values in the fourth row are all one. A further variable that causes many other variables is PC1. The sorting algorithm rearranges the order of the variables. After re-entering the detected dead times T_d, the causality matrix has the following form.

x / y	PI2	PC1	LC1	TC1	TC2	TI1	TI2	PI1
PI2	-	1	5	5	5	7	16	22
PC1		-	4	4	4	6	14	19
LC1			-		1	3	11	15
TC1				-		2	9	10
TC2					-	1	9	4
TI1						-	7	
TI2							-	8
PI1								-

This result reveals completeness and consistency. For example, PI2 not only causes all other variables but the order of occurrence can be retrieved from the dead time. The peak occurs first in PI2, then 1 sample later in PC1 and then 4 samples later in LC1. The

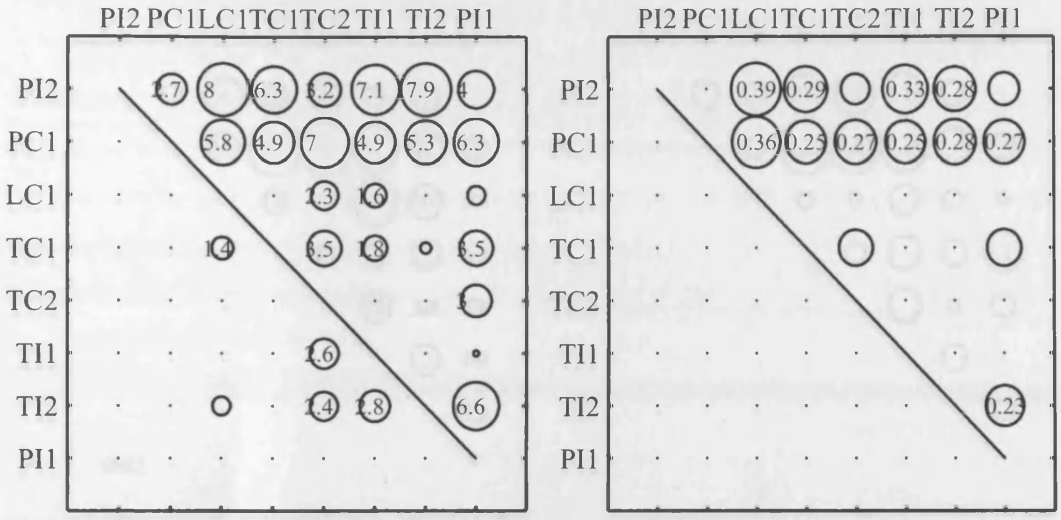
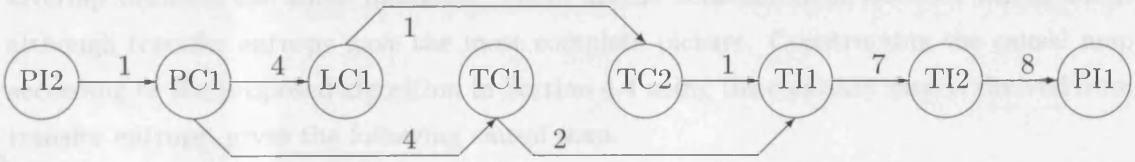
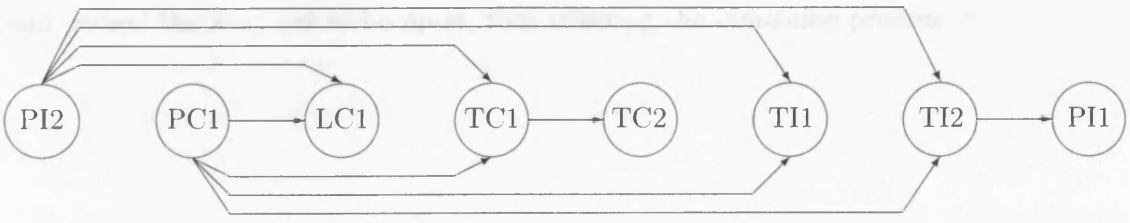


Figure 9.3: Significance level (left panel) and causality measure (right panel) of the nearest neighbours method for case study with irregular disturbance.

last variable to be affected is PI1 which occurs 22 samples after the initial disturbance in PI2. Thus, the cross-correlation function maps out a picture of cause and effect. This picture, however, is not complete. For example the dependencies between LC1 and TC1 and between TC1 and TC2 can only be concluded indirectly since no dead time between these measurements was detected. Also, the value 4 for the dependency between TC2 and PI1 does not seem to fit into the scheme. However, since this is the only exception, the value of the detected time delay is omitted here.



Nearest Neighbours: The results of the nearest neighbours method from Chapter 6 is shown in Figure 9.3. The parameters were set as given in the guidelines in Table 7.2, that is, prediction horizon and time delay were set to $h = \kappa = 4$ and the embedding dimension to $m = 3$. For this approach, 400 samples were analysed. The results show that the pressures PI2 and PC1 affect all other variables. However, unlike in the cross-correlation investigation, the nearest neighbours method does not identify whether PI2 or PC1 occurred first. Also, most other relationships between the following measurements were not detected.



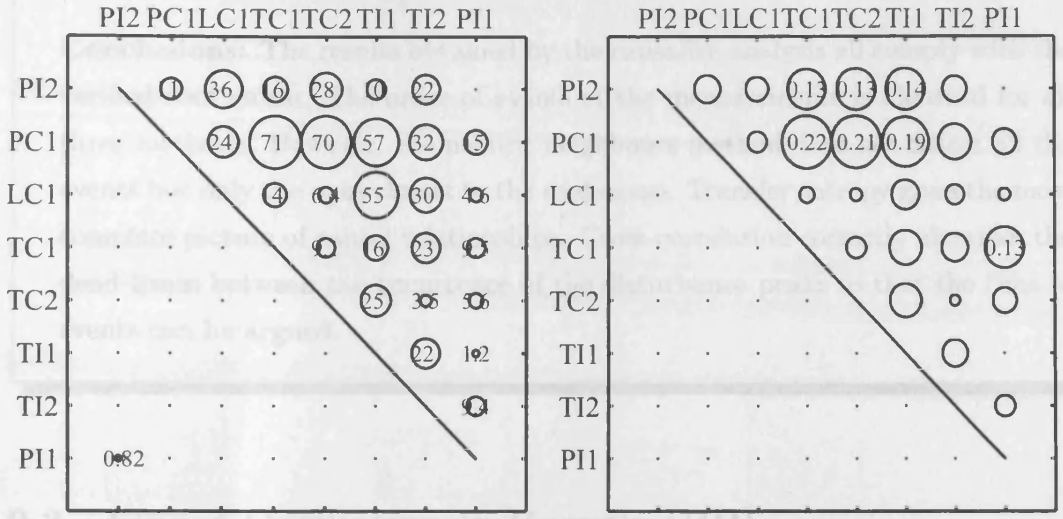


Figure 9.4: Significance level (left panel) and causality measure (right panel) of the transfer entropy method for case study with irregular disturbance.

Transfer Entropy: Figure 9.4 shows the results of transfer entropy. Again, the parameters are set according to the guidelines in Table 7.2 and in total 3000 samples were used for the analysis. The results show a picture similar to the matrix derived from the cross-correlation function. Unlike the previous methods, transfer entropy gives a complete picture and detects the dependencies that were missing previously, such as the relationship between LC1 and TC1 as well as TC1 and TC2.

Overlap between the three methods. There are no contradictions between the methods although transfer entropy gave the most complete picture. Constructing the causal map according to the proposed algorithm in Section 4.4 using the causality matrix derived from transfer entropy, gives the following causal map.

This therefore suggests that PI2 is closest to the root cause. Since PI2 is an indicator and not a control variable, the diagnosis of the root cause requires expert knowledge. Referring to the process schematic in Figure 9.2 shows that PC1 and PI2 are two pressures connected with the inert gas flow. Further investigation showed that the inert gas inflow into the condenser and the flash pot was coming from the same pipe controlled by a split pressure controller. Investigations revealed that the pressure controller had oversized split settings and caused the inert gas to be upset, thus affecting the remaining process.

Conclusions: The results obtained by the causality analysis all comply with the verified root cause. The order of events of the measurements is identical for all three methods. However, the nearest neighbours method does not detect all the events but only the ones closest to the root cause. Transfer entropy gives the most complete picture of causal relationships. Cross-correlation correctly identifies the dead times between the occurrence of the disturbance peaks so that the time of events can be argued.

9.3 Linear Oscillation in Recycle (BP)

The disturbance described in this section is included in the thesis because the nature of the disturbance is linear. In the reference case study from Section 4.2, the measurements closer to the root cause were also more nonlinear. The nonlinearity can be seen from the frequency spectrum. In this example, the oscillation that affects a number of variables is linear and without any harmonics in the frequency spectrum. Thus, the nonlinearity index provides no means for cause and effect. However, the disturbance is not a pure sinusoid but has a considerable amount of stochastic contribution. Therefore the application of the statistical causality measures derived in this work could give some information about cause and effect in the process. The three methods of cross-correlation, nearest neighbours and transfer entropy are applied and their results are compared. A hypothesis of the root cause is made. The process schematic and data are courtesy of BP Chemicals and was provided by Adrian Meaburn.

9.3.1 Process and Data Set

The process schematic of the case study is shown in Figure 9.5 and includes six control loops that were available for the analysis. The time trend of the eight loops, including controller output and controller error[1], is shown in Figure 9.6. The sampling rate is 1 per minute and 3000 samples in total were analysed. The time trend shows a regular oscillation affecting to some extent all measurements. The period of oscillation is 56 samples or 56 minutes. The oscillation can also be clearly seen in the frequency spectra in Figure 9.7. The lack of any harmonics in the frequency spectrum suggests no nonlinearity. In addition to this most prominent oscillation, level controllers LC1 and LC4 exhibit slow deviations in the lower frequencies.

The process schematic shown in Figure 9.5 is the information available from Adrian Meaburn at BP Chemicals. The information provided is not enough for understanding and

[1]The controller error is the setpoint minus the measured process variable.

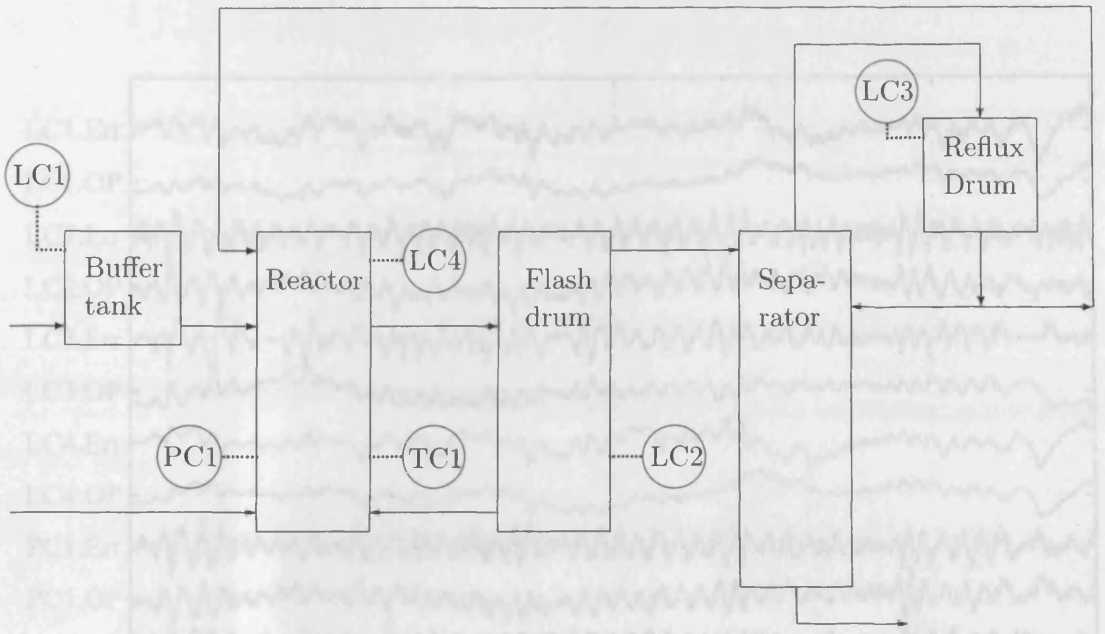


Figure 9.5: Process schematic of process with linear oscillation in a recycle.

arguing the controller schemes in place since the manipulated variables are not indicated. For the purpose of retracing the root cause by investigating the process measurements this suffices. The process schematic given here was suggested and approved by BP Chemicals. Details about the kind of the reaction cannot be disclosed.

9.3.2 Question on Directionality

The ultimate goal of the directionality analysis is to find the root cause of the 56 samples oscillation. The analysis is particularly challenging since the process includes a recycle. In the case of a recycle, there are two different kinds of disturbance propagation mechanisms:

- The disturbance originates from a single cause such as oscillatory tuning of a loop or a disturbance that enters the process through an input feed.
- The oscillation is due to the setup of the process, either by the presence of the recycle or by the interaction between two loops fighting for the same process quantity.

The automatic generation of the causal map assumes the first option and then breaks up the recycle. The first variable in the resulting chain of occurrences is the variable that is most likely to be the root cause. Therefore, no circular causality can be represented. However, the causality measure gives also insight into the second case of a inherent instability and argues the direction of propagation and the events of occurrence for the variables that cannot be straight forwardly placed in the recycle. For example, the event of occurrence between the level, pressure and temperature in the reactor in the process shown in Figure 9.5 combined with process knowledge can reveal more information about the cause of the disturbance.

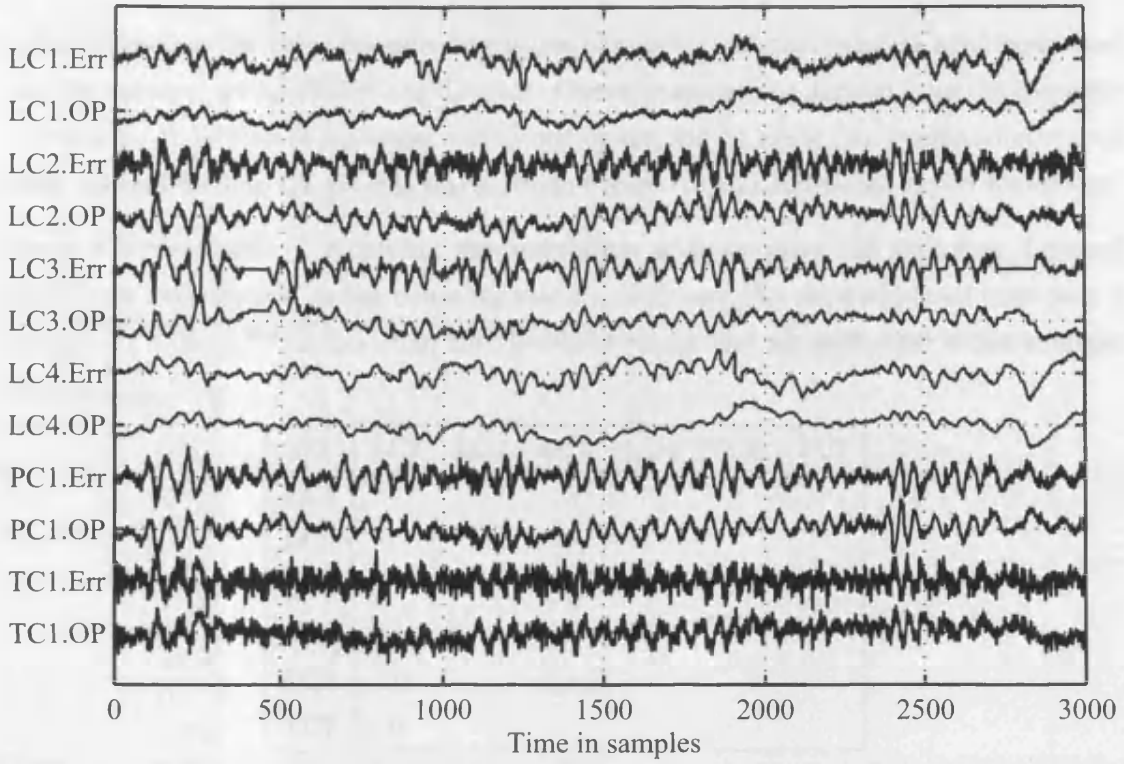


Figure 9.6: Time trends of controller errors and outputs of variables indicated in Figure 9.5.

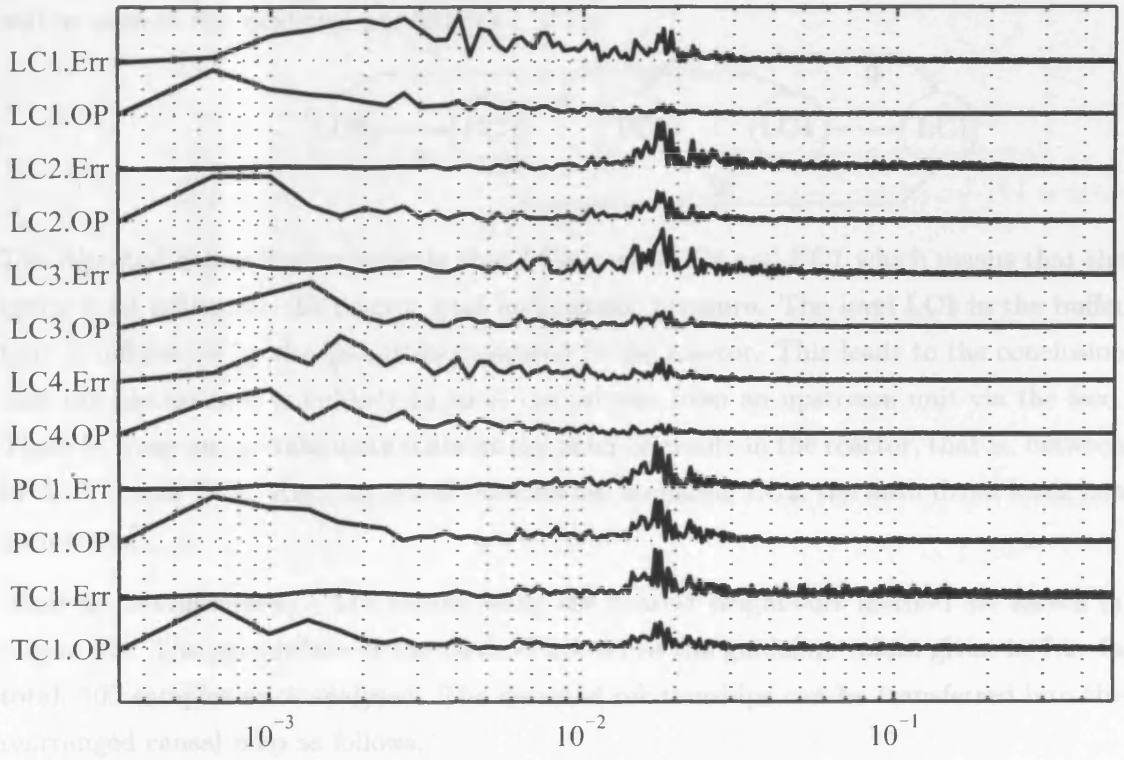


Figure 9.7: Frequency plots of controller errors and outputs of variables indicated in Figure 9.5.

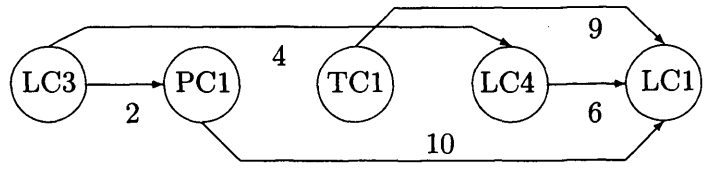
9.3.3 Interpretation of Results

In the following, the three causality measures of cross-correlation, nearest neighbours and transfer entropy are applied to the data set. Causal maps will be derived from the causality matrices for the different measures and a comparison will be made. An explanation of root cause discussed using the process schematic in Figure 9.5 and additional expert knowledge.

Cross-Correlation: Applying the correlation analysis gives the following detected dead times summarised in the causality matrix. Only samples are considered that have a correlation index $r^{\max}$ larger than the threshold r_{thresh} and an oscillation index ψ larger than ψ_{thresh}.

x/y	LC1	LC2	LC3	LC4	PC1	TC1
LC1	-					
LC2		-				
LC3			-	4	2	
LC4	6			-		
PC1	10				-	
TC1	9					

The reason why only relatively few dependencies, five out of 30 potential dependencies, are found is that the disturbance is an oscillation and that it cannot distinguished between a dead time of T_d and $T_{osc} - T_d$. After reordering the variables, the following causal map results. The order of events complies with the results obtained from all three methods as will be seen in the next two paragraphs.



The detected dependencies indicate that LC3 causes LC4 and PC1 which means that the reflux level influences the reactor level and reactor pressure. The level LC1 in the buffer tank is influenced by the quantities measured in the reactor. This leads to the conclusion that the disturbance is unlikely to enter the process from an upstream unit via the feed. There is, however, no information about the order of events in the reactor, that is, between LC4, PC1 and TC1. Also, no causal information including LC2, the flash drum level, can be inferred.

Nearest Neighbours: The results using the nearest neighbours method are shown in Figure 9.8. The parameters of the method are set to the guideline values given in 7.2. In total, 400 samples were analysed. The detected relationships can be transferred into the rearranged causal map as follows:

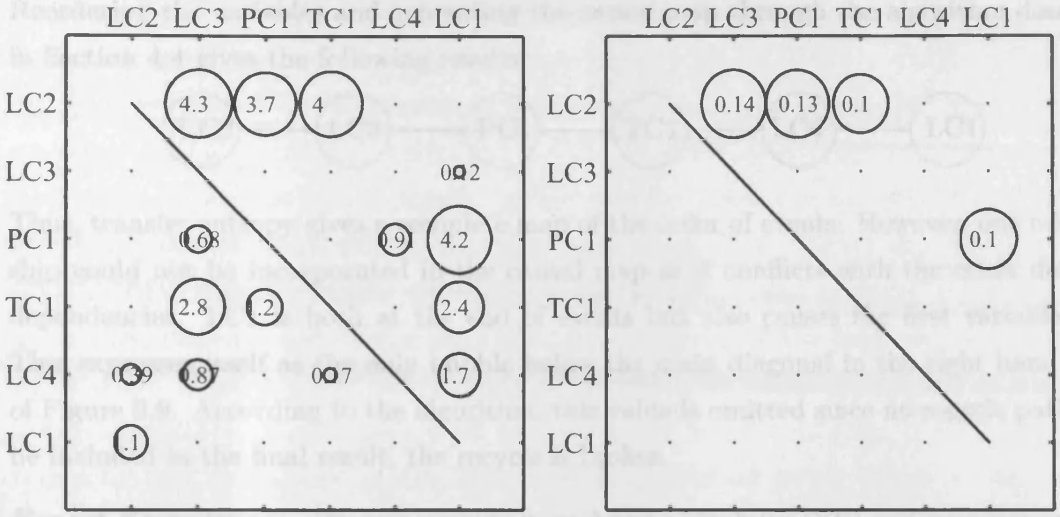


Figure 9.8: Significance level and (left panel) and causality measure (right panel) of the nearest neighbours method for case study with linear oscillation.

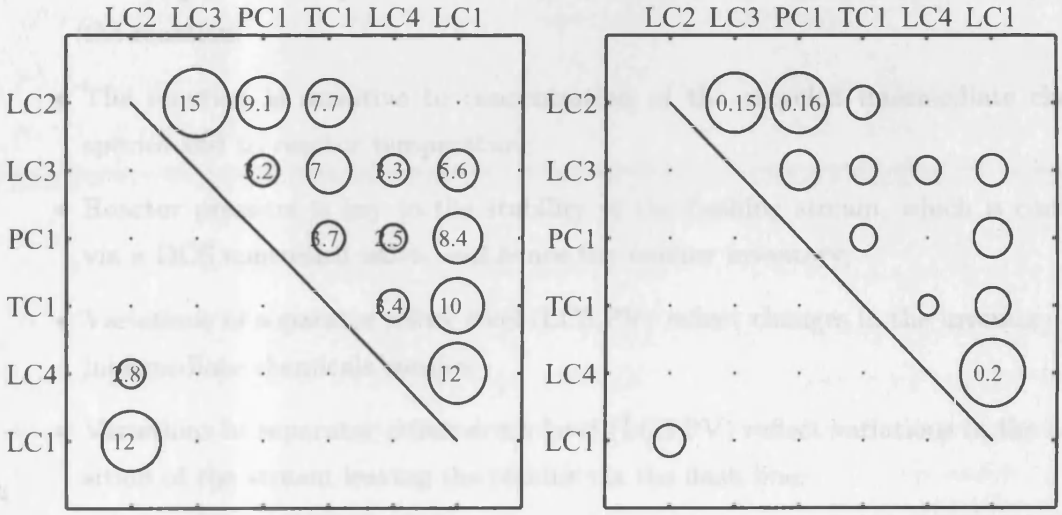
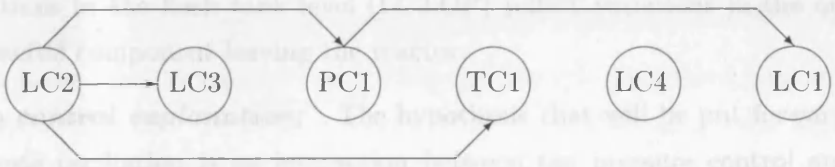


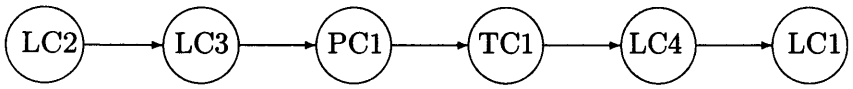
Figure 9.9: Significance level and (left panel) and causality measure (right panel) of transfer entropy for case study with linear oscillation.



Similar to the cross-correlation function method, only few dependencies are detected. Here, the level of the flash drum, LC2, influences the reflux drum level LC3 and reactor pressure and temperature (PC1 and TC1). A further relationship between PC1 and LC1 is detected. Together with the cross-correlation function, a more complete picture of cause and effect in the process evolves.

Transfer entropy: The results of the transfer entropy are shown in Figure 9.9. The standard guidelines parameter are set for the analysis and all 3000 samples were used.

Reordering the variables and generating the causal map through the algorithm described in Section 4.4 gives the following results:



Thus, transfer entropy gives a complete map of the order of events. However, one relationship could not be incorporated in the causal map as it conflicts with the other detected dependencies. LC1 is both at the end of events but also causes the first variable LC2. This expresses itself as the only bubble below the main diagonal in the right hand panel of Figure 9.9. According to the algorithm, this value is omitted since no recycle paths can be included in the final result, the recycle is broken.

Expert Knowledge: Process insights from Adrian Meaburn (BP) and information from the process schematic lead to the following comments.

- The recycle from separator to reactor contains an intermediate chemical species in the reaction;
- The reaction is sensitive to concentration of the recycled intermediate chemical species and to reactor temperature;
- Reactor pressure is key to the stability of the flashing stream, which is controlled via a DCS controlled valve, and hence the reactor inventory;
- Variations in separator reflux level (LC3.PV) reflect changes in the inventory of the intermediate chemicals species;
- Variations in separator reflux drum level (LC3.PV) reflect variations in the composition of the stream leaving the reactor via the flash line;
- Liquid recycled from the reflux drum (LC3.OP) will be cool and will disturb the reactor. Variations in flow also affect the composition in the reactor;
- Variations in the flash tank level (LC2.OP) reflect variations in the amount of an un-reacted component leaving the reactor.

Inventory control explanation: The hypothesis that will be put forward to explain the 56 minute oscillation is an interaction between the pressure control and inventory of the recycled intermediate chemical species. The coupling is taking place through the reaction. The recycling of an intermediate chemical back into the reactor means that a physical feedback loop exists. A plant with a recycle can oscillate just as a feedback control loop can oscillate if the gain is too high. Such an oscillation will be affected by process loop gain (i.e. the extent to which one physical variable influences others in the path). In this case the process loop gain would be affected by reaction rate and also by the presence or absence of buffering capacity within the recycle path.

Supporting evidence: Supporting evidence for the hypothesis is that transfer entropy testing highlighted a causal interaction chain LC2→LC3→PC1→LC4→LC1 involving flash tank level (LC2), separator reflux drum level (LC3), reactor pressure (PC1) reactor level (LC4) and feed buffer tank (LC1). All these measurements involve the inventory in the recycle and are explained by the process insights in the previous column, for instance it says that material in the reflux drum level (LC3) upsets the reactor pressure (PC1). Other evidence is that the variables participating most strongly in the oscillation are reactor pressure and temperature and the inventories of the flash tank and separator reflux drum which contain the products of the chemical reaction and give an indication of their inventory. The reactor level was not found to be oscillating overall because its behaviour was dominated by long slow deviations from set point. However there was an oscillation comprising 24% of the total signal power superimposed upon its low frequency trend. In other words, the reactor level is participating in the oscillation though not dominated by it. Transfer entropy also showed a causal path PC1→TC1→LC1 suggesting an influence from reactor pressure to temperature and feed tank level.

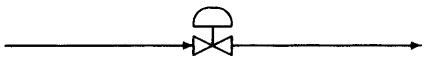
Plant testing: Step testing confirmed the interaction between the inventory in the separator reflux drum level (LC3) and the reactor level controller output (LC4.OP) and vice versa. Adrian Meaburn reported he has managed to settle the plant by making the level control in the separator reflux tank less tight thus including buffering capacity into the recycle. The effect would be to make the recycle flow more steady and thus to prevent the oscillation from feeding back and upsetting the reactor.

An advanced control solution: An advanced controller was commissioned in order to successfully stabilise the inventory controllers on the unit. This was based upon a FIR process model which showed strong interaction on the inventory controls based upon the step response data gathered on the unit. The decision to break the inventory controls was strongly influenced by the studies carried out in this report, which served to confirm operational experience on the unit.

Conclusions: In the example of a linear oscillation, cause-and-effect relationships can be detected with the methods of cross-correlation, nearest neighbours and transfer entropy. The detection is assigned to the combination of the oscillation and structured noise since a purely linear oscillation would give no indication to cause and effect. The detected relationships vary for all three methods but can all be incorporated in the same causal map. Transfer entropy gives the complete causal map by detecting all dependencies.

Chapter 9 Summary

In this chapter, the causality measures of cross-correlation, nearest neighbours and transfer entropy were applied to industrial case studies. Two new case studies were introduced and a root cause hypothesis formed from the causality analysis. The results of the three methods are consistent. In both case studies, an irregular upset and a linear oscillation, transfer entropy gave the best and most complete results.



Chapter 10

Data Quantisation and Compression

This chapter studies the impact of data compression and quantisation on the causality measure. For this purpose, the raw time trends are compressed and quantised with a varying compression factor CompDev and quantisation factor f_q. The compression algorithm investigated here is also used by the PI data historian.

The time trends that are investigated for the causality analysis are not strictly the actual process variables. Before the analysis, the time trends are measured by a sensor in the distributed control system (DCS) and then processed by the data acquisition system as described in Section 3.5.1. These steps can distort the behaviour of the investigated causality measures. The most common effects are data quantisation and data compression which will be described in the following two sections. The impact on the causality measures of cross-correlation, nearest neighbours and transfer entropy is investigated and then summarised.

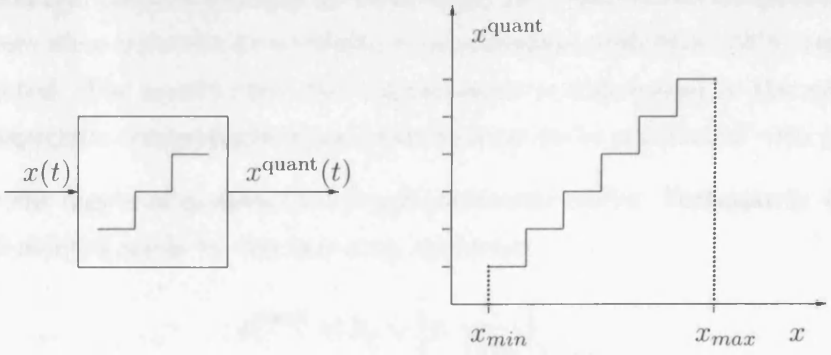


Figure 10.1: Block diagram (left) and transfer function (right) of a quantiser.

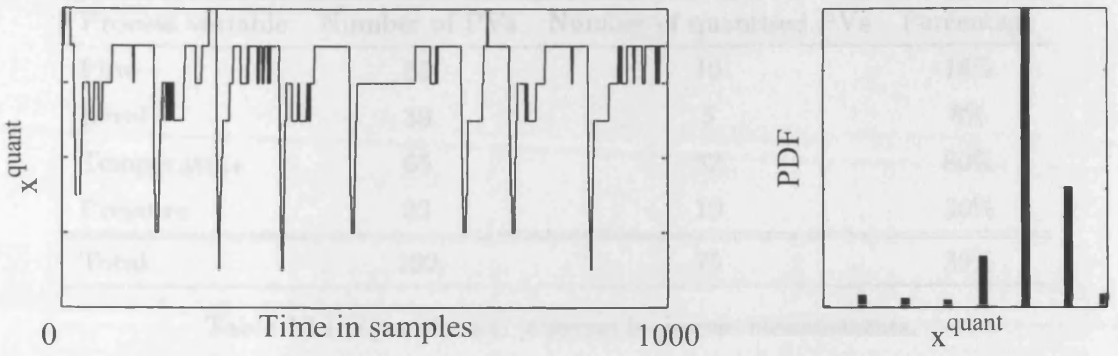


Figure 10.2: Time trend and probability density function of quantised process variable.

10.1 Data Quantisation

When converting an analogue process measurement into a discrete signal, the measurement has to be quantised. Figure 10.1 shows the symbol of a quantisation block and the functional mapping from the unquantised, continuous variable $x(t)$ onto the quantised variable $x^{\text{quant}}(t)$. The quantised variable can only assume discrete values. An example of the resulting quantised data trend is shown in the upper panel of Figure 10.2. The process variable only assumes a number of discrete amplitude values and alternates between those values. The effect of quantisation can best be observed in the probability density function (PDF), estimated by histograms[1] since the amplitude value falls into discrete amplitude bins. The amplitude bins are separated by equidistance intervals.

Quantisation occurs if the ranges of the analogue-to-digital converter, see Section 2.3.2, are set too large for the actual process conditions. Often, the ranges were set initially during plant configuration and not adjusted when the process conditions changed. A mini study using data from processes at Eastman Chemical Company and BP Chemicals was conducted for a first insight of the extent of quantisation problems in chemical processes. In total, 192 process variables were investigated taken from PI historian data: 55 flow, 39 level, 65 temperature and 33 pressure measurements. The results of visual inspection of the time trend and the histograms is given in Table 10.1. 80% of the temperature and 30% of the pressure measurements show effects of quantisation while flow (18%) and level (8%) are less affected. The results show that quantisation is widespread in the measurements taken and especially temperature measurements have to be considered with great care.

To estimate the degree of quantisation is not necessarily trivial. Fortunately, quantisation can be implemented easily by the following algorithm.

$$x_i^{\text{quant}} = f_q \, \sigma_x \left[x_i \frac{1}{f_q \, \sigma_x} \right]_{\text{round}} \tag{10.1}$$

Here σ_x the standard deviation of the time trend of x_i while $[]_{\text{round}}$ denotes rounding to the next integer value. The quantisation factor f_q gives the strength of quantisation.

[1] For PDF and histograms see Section 7.1.

Process variable	Number of PVs	Number of quantised PVs	Percentage
Flow	55	10	18%
Level	39	3	8%
Temperature	65	52	80%
Pressure	33	10	30%
Total	192	75	39%

Table 10.1: Quantisation problems in process measurements.

Large values for f_q result in fewer discrete bins and thus a higher quantisation. Thus, the quantisation factor considers the range over which the samples are most commonly spread. It does not give the actual number of bins. The number of bins is in fact not a meaningful quantity since one outlier increases the number significantly. A quantisation factor of 1 means that the majority of values will be spread over the bins 0, 1 and -1.

In the following investigations, the unquantised time trends of the reference case study in Section 4.2 will be quantised using equation 10.1. The causality analysis is carried out using the cross-correlation, nearest neighbours and transfer entropy methods. This approach gives the causality measures as a function of quantisation factor f_q and thus the impact of quantisation on the analysis. The quantisation error is a function of the quantisation factor and can be expressed as follows:

$$\epsilon^{\text{quant}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_i^{\text{quant}}\right)^2} \tag{10.2}$$

The quantised time trends of the reference case study with a quantisation factor of $f_q = 1$ and $f_q = 2$ are shown in Figure 10.3. The quantisation error as a function of the quantisation factor f_q for all variables from the reference case study is shown in Figure 10.4. Up to a quantisation factor $f_q = 0.8$, the error shows a similar behaviour. For higher values of f_q, TI5, LC1 and TI7 show higher errors than all other process variables. The reason for this could be that those variables are similar to a sine wave and thus are more exposed to the quantisation than the sharp peaks of signals TI1 to TI4.

Impact of Quantisation on Cross-Correlation: To investigate the impact of quantisation on the cross-correlation causality measure from Chapter 5, the original signals are quantised by a varying factor f_q. shown in Figure 10.5. The original unquantised time trend are successively artificially quantised as described in Equation 10.1. The time trends of the reference case are taken and the five dependencies, between TI1 to TI5 and LC1 and TI7, are investigated. The plots in the left column of Figure 10.5 show the detected dead time T_d. The values as detected in Table 5.2 were TI1→TI2: 2 samples, TI2→TI3: 6 samples, TI3→TI4: 5 samples, TI4→TI5: 11 samples and LC1→TI7: 8 samples. All signal dead times are correctly identified with a variation of ±1 samples up to a quantisation factor of approximately 2.9. The correlation index $r^{\max}$ is shown in the middle

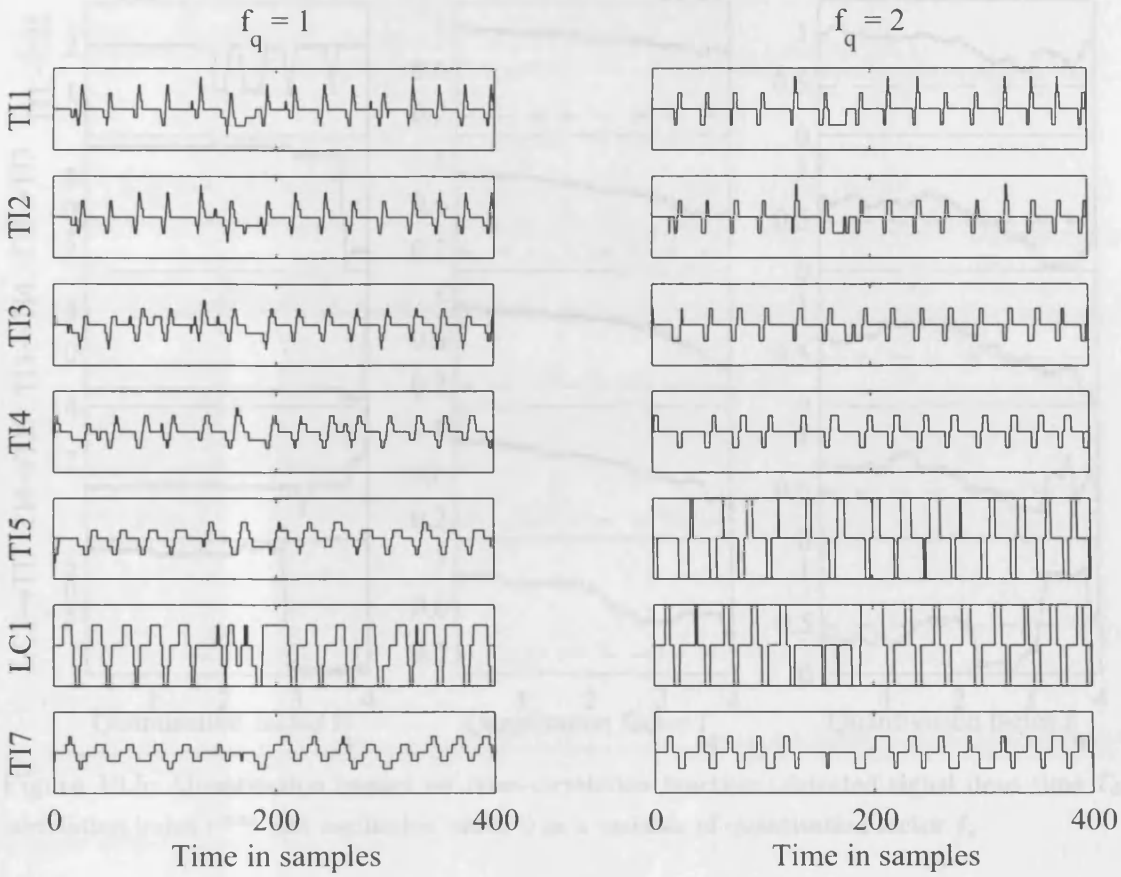


Figure 10.3: Quantised time trends of the reference case study with two quantisation factors: $f_q = 1$ (left panels) and $f_q = 2$ (right panels).

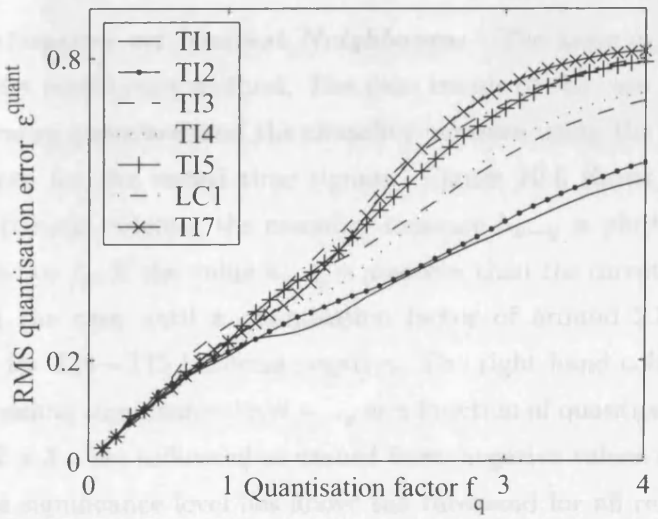


Figure 10.4: Quantisation error as function of quantisation factor f_q for time trends of reference case study.

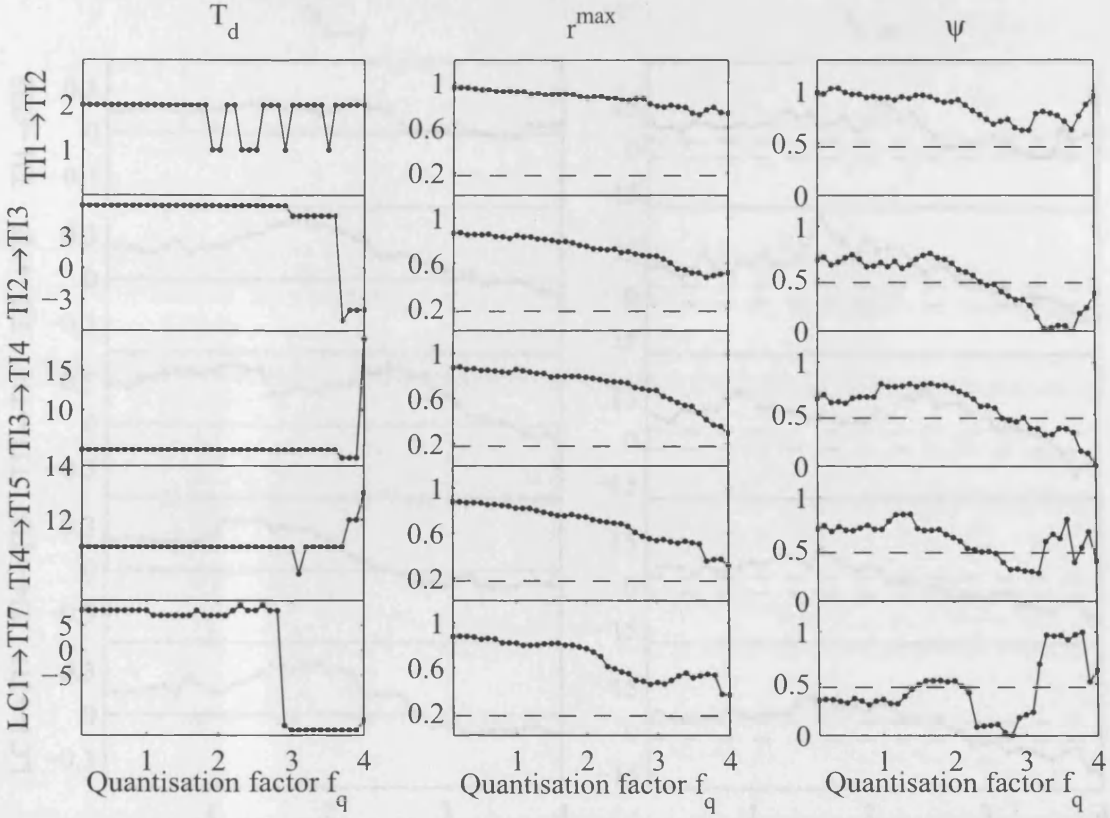


Figure 10.5: Quantisation impact on cross-correlation function: detected signal dead time T_d, correlation index $r^{\max}$ and oscillation index ψ as a variable of quantisation factor f_q.

column of Figure 10.5 and the oscillation index in the right hand column. The correlation coefficient decreases with increasing f_q but not below the critical threshold r_{thresh} that is indicated by dashed lines. The oscillation index ψ falls below the threshold ψ_{thresh} for a quantisation factor larger than 2 for the relationships TI2→TI3, TI3→TI4 and TI4→TI5. This supports the detection of the correct signal dead time by T_d for these relationships.

Impact of Quantisation on Nearest Neighbours: The same experimental setup is made for the nearest neighbours method. The time trends of the case study from Section 4.2 are gradually more quantised and the causality measure using the nearest neighbours method is computed for the varied time signals. Figure 10.6 shows the results of this analysis. In the left hand column, the causality measure $h_{x\to y}$ is plotted as a function of the quantisation factor f_q. If the value $h_{x\to y}$ is positive, then the directionality is detected correctly. This is the case until a quantisation factor of around 2.5 above which the causality measure for TI4→TI5 becomes negative. The right hand column of Figure 10.6 shows the corresponding significance level $s_{x\to y}$ as a function of quantisation factor f_q. The threshold values of $\pm$ 3 σ are indicated as dashed lines, negative values indicating a inverse directionality. The significance level lies above the threshold for all relationships up to a quantisation factor of 2.1 below which the significance level between LC1 and TI7 drops below 3. The quantisation significance level leads to false negative values for very high quantisation errors because a different structure of the underlying system is introduced

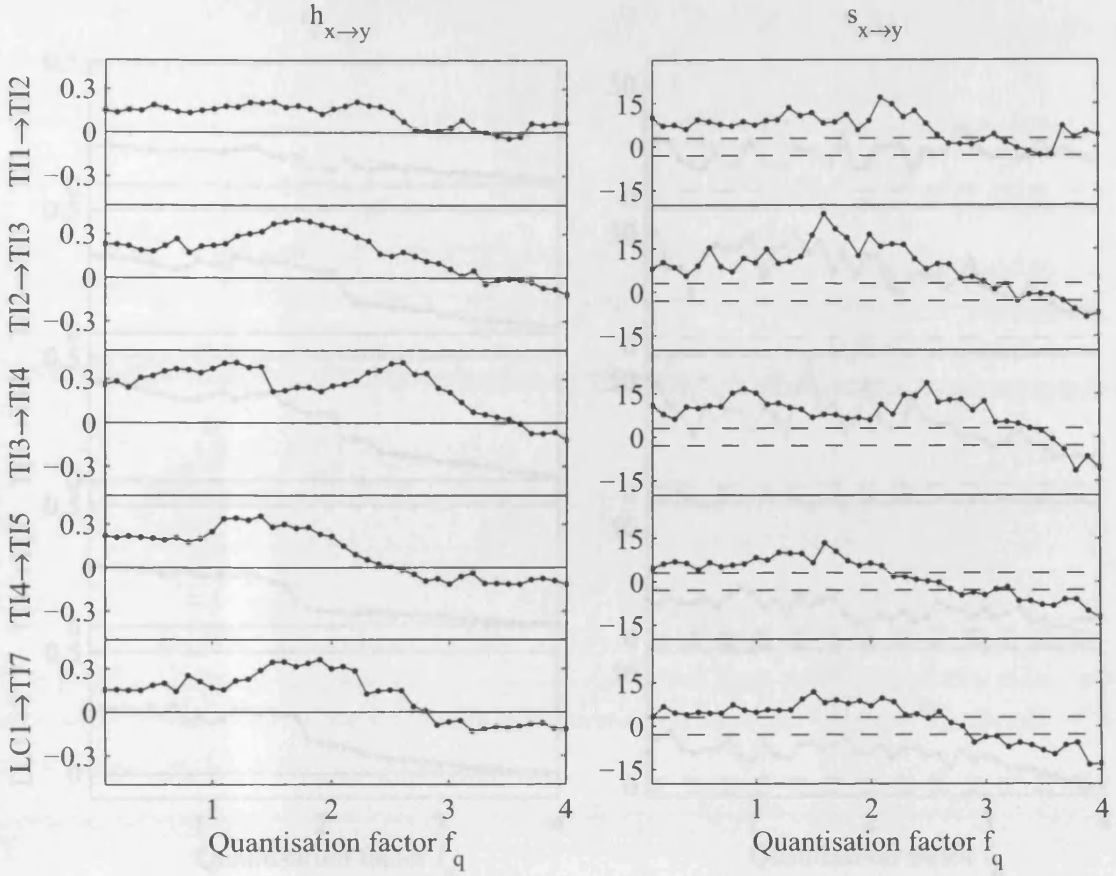


Figure 10.6: Quantisation impact on nearest neighbours causality measure $h_{x\to y}$ and significance level $s_{x\to y}$ as function of compression factor f_q.

by strong quantisation. For example, a square wave might be introduced through high quantisation.

Impact of Quantisation on Transfer Entropy: The impact of quantisation on transfer entropy is shown in Figure 10.7. The left hand column shows the dependency of the transfer entropy difference $t_{x\to y}$ on the quantisation factor f_q. The entropy measure $t_{x\to y}$ is positive even for high quantisation errors and therefore correctly detects the directionality. The significance level $s_{x\to y}$ as a function of the quantisation factor is shown in the right hand column of Figure 10.7. Like the transfer entropy measure, the significance level is also above the detection threshold of 3 for all values of f_q. Only the significance level $s_{\text{LC1}\to\text{TI7}}$ falls below the threshold for a quantisation error larger than 3.8. No direction is detected for this case, unlike the nearest neighbours method for which incorrect directions were detected. Thus, transfer entropy performs better when exposed to quantisation than causality measures using cross-correlation or nearest neighbours.

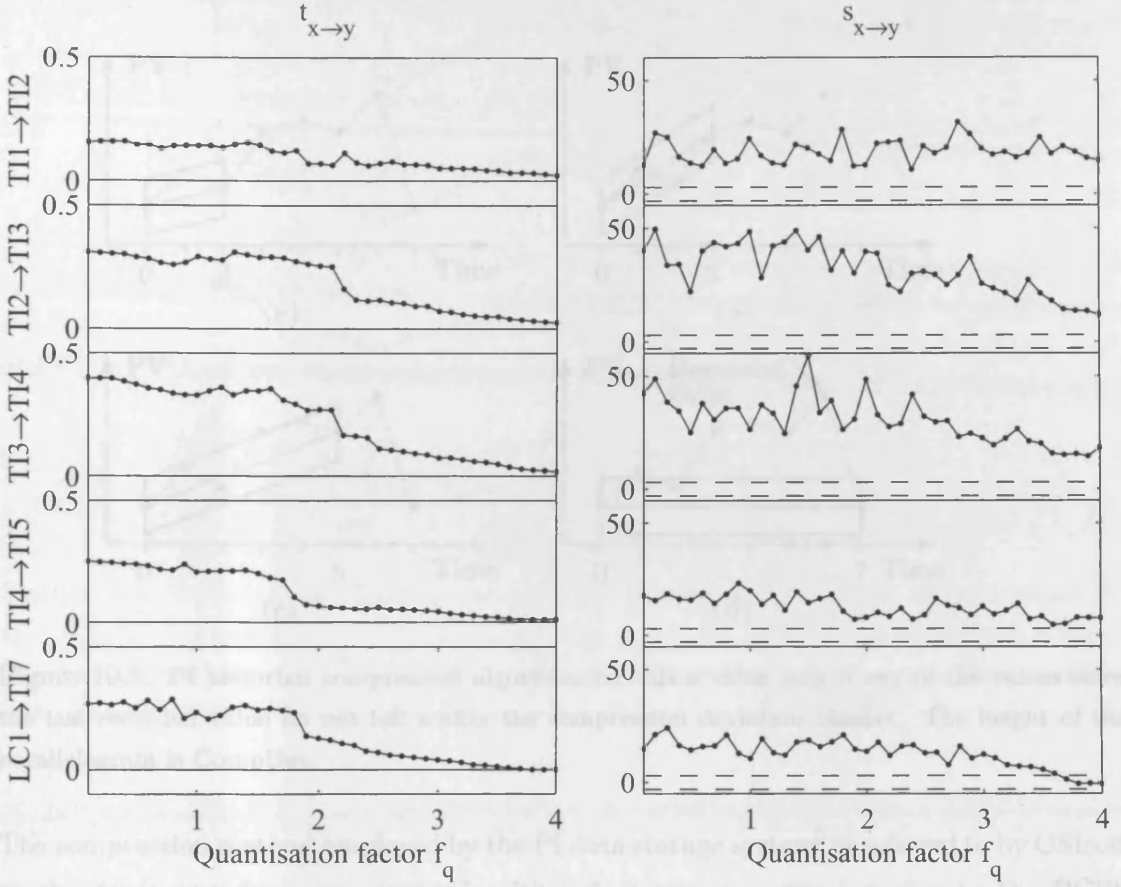


Figure 10.7: Quantisation impact on transfer entropy causality measure $t_{x\to y}$ and significance level $s_{x\to y}$ as function of compression factor f_q.

Conclusions: Quantisation affects the performance of all three causality measures. A quantisation factor up to $f_q = 2$ can still be dealt with by all the causality measures. For the case that $f_q > 3$, the nearest neighbours and cross-correlation method gave incorrect results while transfer entropy detected no causality instead of an incorrect direction. Thus, transfer entropy is the most robust method to quantisation compared to the other two causality measures.

10.2 Data Compression

This section describes the impact of data compression on the causality measures. Process data is often compressed for data storage purposes. Compression involves the storage of only a fraction of the data points. The compressed data is then restored for conducting fault diagnosis and applying data-driven methods. A number of compression algorithms are used in the process industries and SCADA systems. The swinging door compression algorithm [9] and the Box Car Back Slope (BCBS) [94] algorithm are most frequently used.

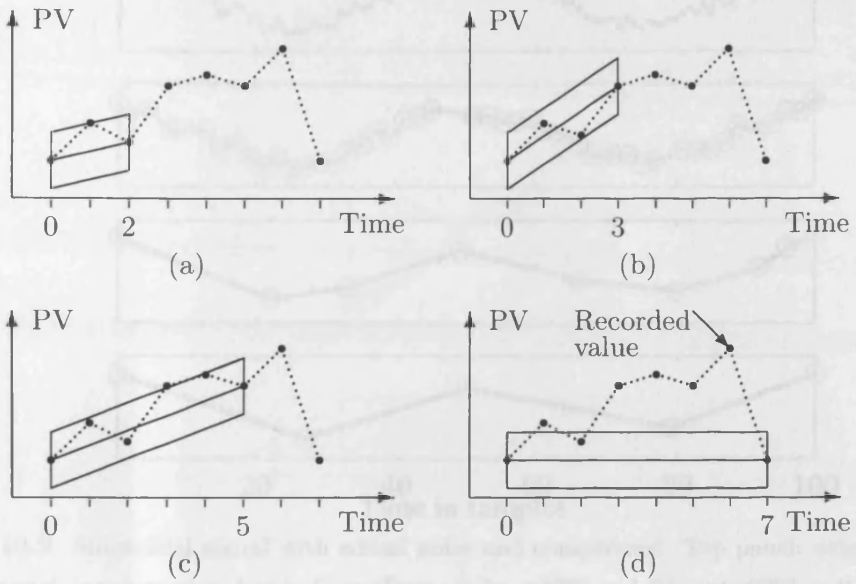


Figure 10.8: PI historian compression algorithm records a value only if any of the values since the last recorded value do not fall within the compression deviation blanket. The height of the parallelogram is CompDev.

The compression method employed by the PI data storage system[2] is referred to by OSIsoft as the "swinging door compression", although it uses a method similar to the BCBS method. In the following, the PI algorithm will be explained in detail. Singhal and Seborg [121] investigate the PI compression algorithm effect on pattern matching techniques. The impact of swinging door compression on statistical properties, the nonlinearity index (see Section 6.2.3) and the Harris index (Section 3.1.3) has been investigated by Thornhill et al. [138]. The authors recommend caution when dealing with compressed data and provide an estimate of the compression factor from process data. The PI historian compression algorithm will be described in the next section and the impact of this algorithm on the causality measures will be discussed thereafter.

PI Historian Compression Algorithm: The PI data historian compression algorithm is displayed in Figure 10.8 and can be described as follows. All values that fall on a line connecting values that are recorded in the archive [146]. When a new value is received the previous value is recorded only if any of the values since the last recorded value do not fall within the compression deviation blanket. The deviation blanket is a parallelogram extending between the last recorded value and the new value with a width equal to twice the compression deviation specification. When the current measured value has exceeded the error bound, defined by the recording limit, the value at the *previous* time step is recorded and the algorithm repeated. The discussed procedure is illustrated in Figure 10.8. In the first three figures (i) to (iii), all previous values fall in the deviation blanket.

[2]http://www.osisoft.com; Information on the PI data historian by OSIsoft, March 2005.

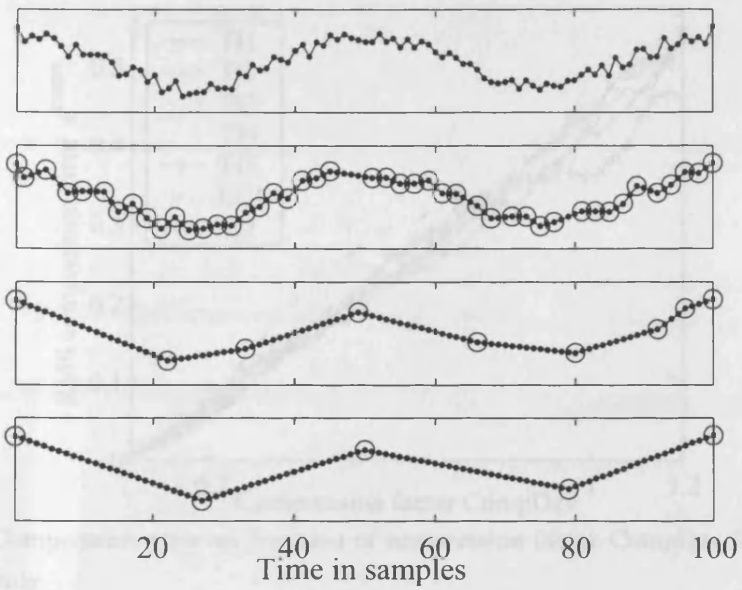


Figure 10.9: Sinusoidal signal with added noise and compressed. Top panel: original time series; second panel: compression factor CompDev= 0.2σ, $r^{\text{comp}} = 1.33$ and $\epsilon^{\text{comp}} = 0.02$; third panel: compression factor ComDev= 0.8σ, $r^{\text{comp}} = 10$ and $\epsilon^{\text{comp}} = 0.2$; bottom panel: compression factor CompDev= σ, $r^{\text{comp}} = 20$ and $\epsilon^{\text{comp}} = 0.24$.

In figure (iv) this is not the case and thus a new value, the previous, is recorded. The method is called "swinging door compression" since one side of the deviation blanket is fixed while the other side swings according to the data points.

The data historian then stores the recorded values and their time indices. The time trend is reconstructed by extrapolating the missing values through linear segments. Obviously, an error is made by reconstructing the signal. If both the original and the restored time trend are available, the average compression error $\bar{\epsilon}$ can be calculated as:

$$\epsilon^{\text{comp}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_i^{\text{restored}}\right)^2}. \tag{10.3}$$

The compression ratio r^{comp} is derived from the number of samples of the compressed time series N_{original} and the number of samples of the original time series N:

$$r^{\text{comp}} = \frac{N}{N_{\text{original}}}. \tag{10.4}$$

An important parameter in the compression procedure is the compression deviation. It can be adjusted to specify the compression error and the compression ratio. Additionally, a minimum and maximum compression time can be implemented in the algorithm. A new event is *not* recorded if the time series since the last recorded event is less than the minimum compression time for the point. This can be expressed in mathematical terms for the new event x_{i+L} which is not recorded if:

$$x_j \in x_i + j\frac{x_{i+L} - x_i}{L}\left[-\text{CompDev};\ +\text{CompDev}\right],\ \text{for all } i < j < i + L, \tag{10.5}$$

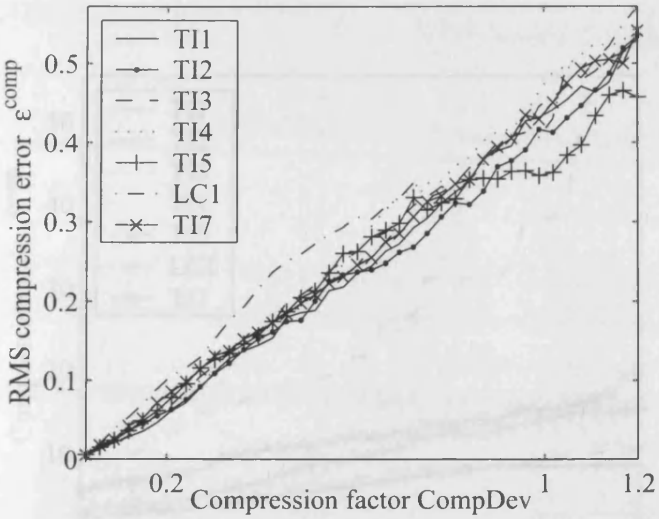


Figure 10.10: Compression error as function of compression factor CompDev for time trends of reference case study.

$$L > \text{CompMin}, \tag{10.6}$$

$$L < \text{CompMax}. \tag{10.7}$$

Here, x_i is the last recorded value, CompDev is the height of the deviation blanket and CompMin and CompMax are the minimum and maximum recording time. If any of the three requirements in Equations 10.5 to 10.7 is not fulfilled, then the value is recorded.

The parameters to be adjusted are hence CompDev, CompMin and CompMax. The parameters are set separately for every control loop in the industrial process. This is often done only once for each loop for the initial setup of the PI system. However, process changes might require changes of the compression parameters and omitting these changes results in poor adjustments of the compression parameter. The deviation blanket parameter CompDev will be referred to as the compression factor in the following sections. The threshold parameters CompMin and CompMax are ignored in the following investigations. CompDev is measured in engineering units in the PI system. For the purpose of investigating the impact of compression on the causality measures, all signals will be normalised, that is, adjusted to zero mean and unit variance. In the following, the compression factor CompDev is therefore a pure number and not in engineering units.

Figure 10.9 shows an example of a sinusoidal signal with added noise that is compressed with varied compression deviations. For a compression factor ComDev of 0.2 times σ, the standard deviation of the signal, as shown in the second panel, the compression ratio 1.33. The difference in the compression error for a compression factor of 0.8 and 1.0 is very small, however, the compression error is increased by a factor two ($r^{\text{comp}} = 5$ for 0.8 and $r^{\text{comp}} = 10$ for 1.0). However, if one is interested in the high frequency part of the signal, the highly compressed signals cannot be used for analysis because only the sinusoidal signal is represented in the time trend.

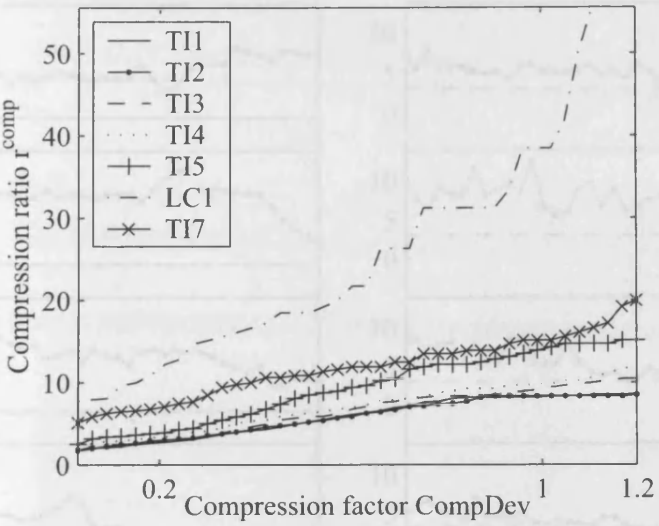


Figure 10.11: Compression ratio as function of compression factor CompDev for time trends of reference case study.

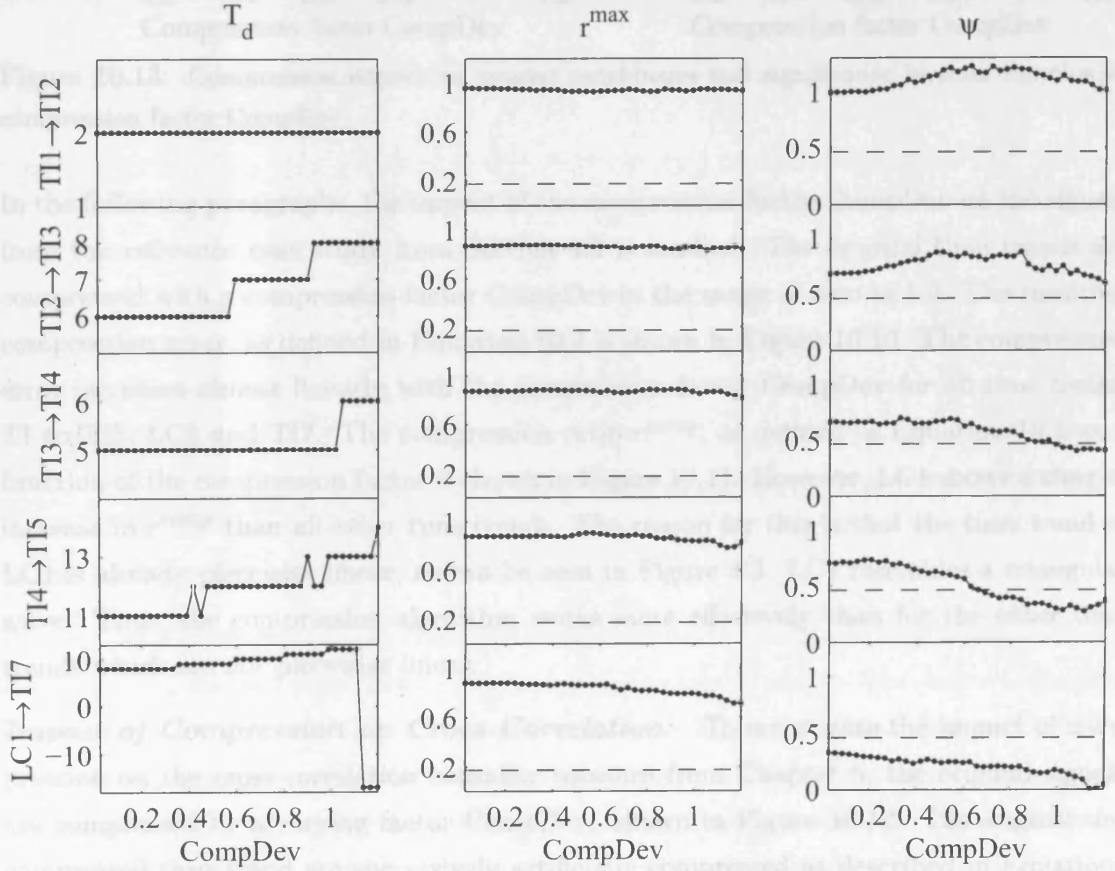


Figure 10.12: Compression impact on cross-correlation function: detected signal dead times and correlation coefficient as a variable of compression factor CompDev.

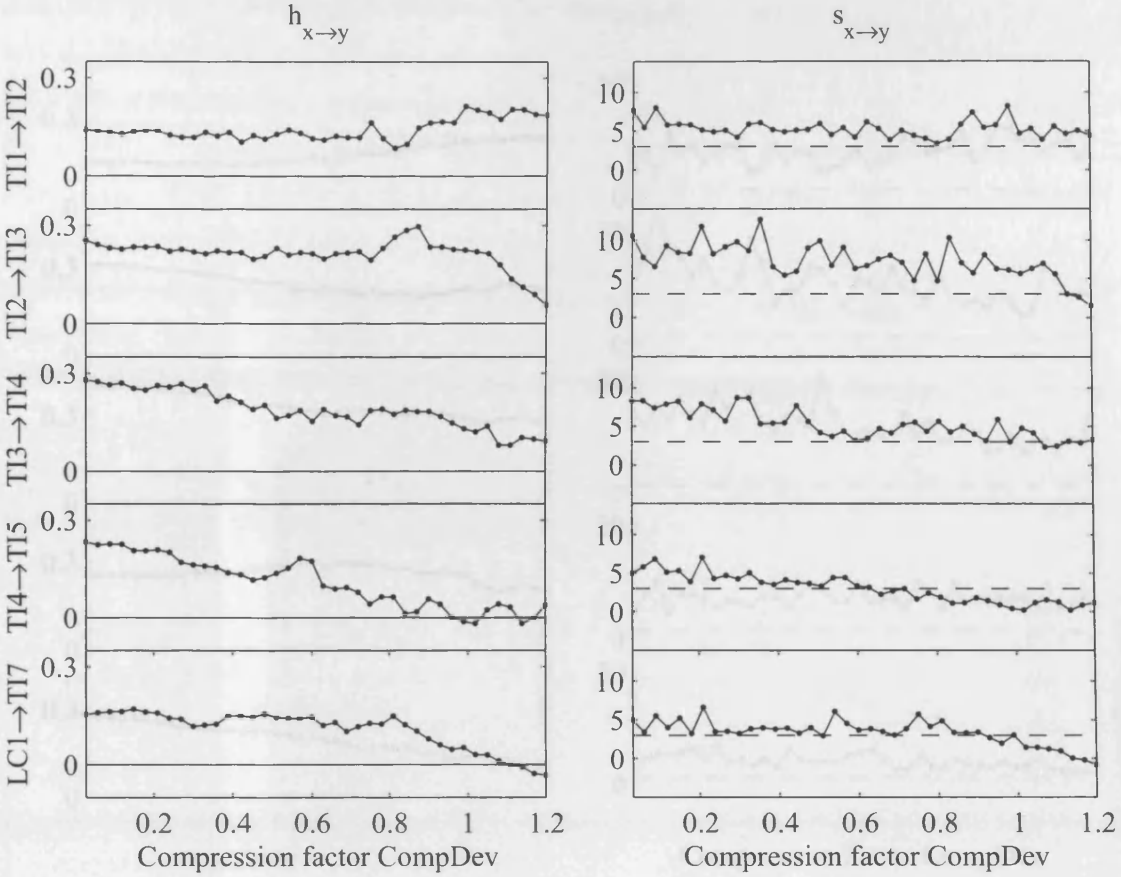


Figure 10.13: Compression impact on nearest neighbours and significance level as function of compression factor CompDev.

In the following paragraphs, the impact of the compression factor CompDev on the signals from the reference case study from Section 4.2 is studied. The original time trends are compressed with a compression factor CompDev in the range of zero to 1.2. The resulting compression error, as defined in Equation 10.3 is shown in Figure 10.10. The compression error increases almost linearly with the compression factor CompDev for all time trends TI to TI5, LC1 and TI7. The compression ratio r^{comp}, as defined in Equation 10.4 as a function of the compression factor is shown in Figure 10.11. However, LC1 shows a steeper increase in r^{comp} than all other time trends. The reason for this is that the time trend of LC1 is already piecewise linear, as can be seen in Figure 4.3. LC1 resembles a triangular wave. Thus, the compression algorithm works more effectively than for the other time trends which are not piecewise linear.

Impact of Compression on Cross-Correlation: To investigate the impact of compression on the cross-correlation causality measure from Chapter 5, the original signals are compressed by a varying factor CompDev. shown in Figure 10.12. The original uncompressed time trend are successively artificially compressed as described in Equations 10.5 to 10.7. The time trends of the reference case are taken and the five dependencies, between TI1 to TI5 and LC1 and TI7, are investigated. The plots in the left column of Figure 10.12 show the detected signal dead time T_d. The values as detected in Table 5.2

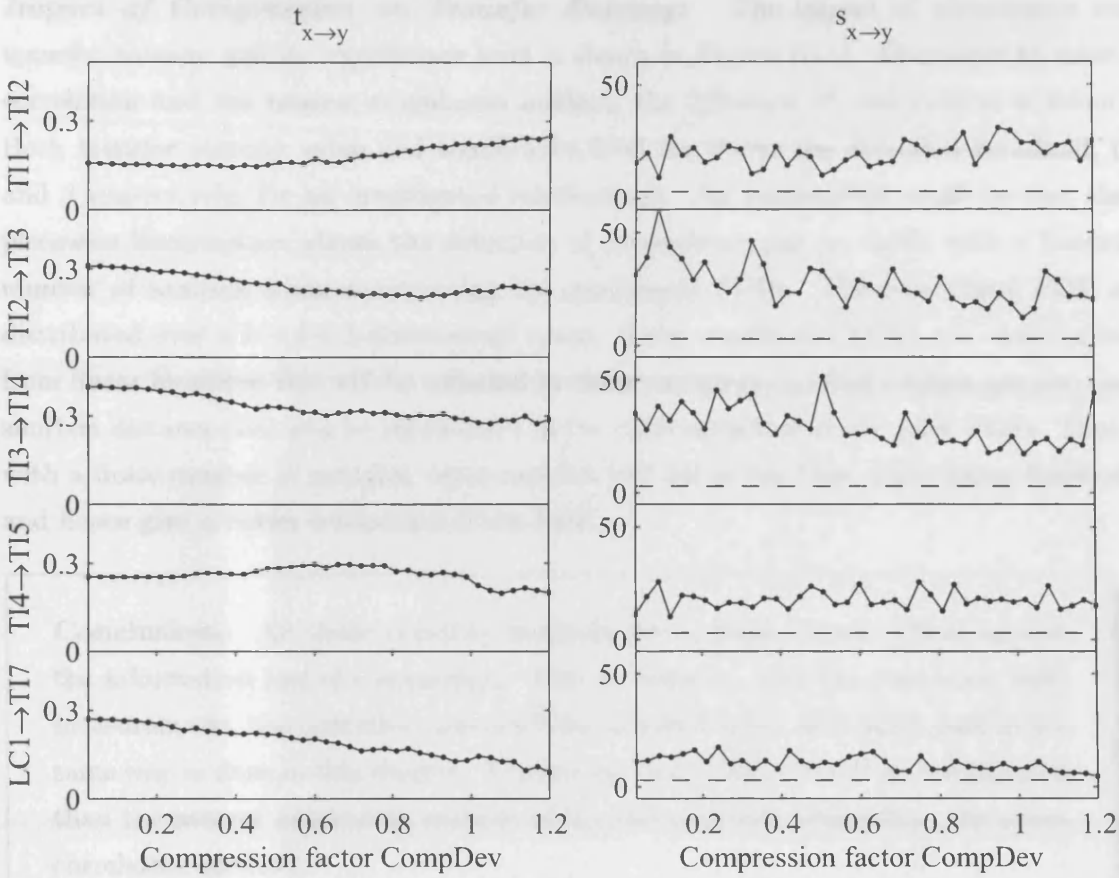


Figure 10.14: Compression impact on transfer entropy and significance level as function of compression factor CompDev.

were TI1→TI2: 2 samples, TI2→TI3: 6 samples, TI3→TI4: 5 samples, TI4→TI5: 11 samples and LC1→TI7: 8 samples. All signal dead times are correctly identified with a variation of ±1 samples up to a compression factor of approximately CompDev= 0.6. The correlation index $r^{\max}$ is shown in the middle column of Figure 10.12 and the oscillation index in the right hand column. The correlation coefficient is almost constant with increasing CompDev and well above the critical threshold r_{thresh}, indicated by dashed lines. The oscillation index ψ falls below the threshold ψ_{thresh} for a compression factor larger than 0.6 for the relationships TI3→TI4. This is supports the detection of the correct signal dead time T_d for these relationships.

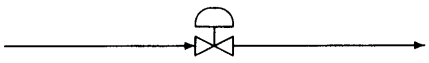
Impact of Compression on Nearest Neighbours: The nearest neighbours causality measure $h_{x\to y}$ and the significance level $s_{x\to y}$ as a function of compression factor CompDev is shown in Figure 10.13. For the relationship between TI1→TI2, TI2→TI3 and TI3→TI4 the causality measure is almost independent of the compression. This might be explained by the periodicity of the time series which is more similar to the linearly compressed signal. It is speculated that the nearest neighbours are found more easily if the signal is periodic and piecewise linear. For the relationships TI4→TI5 direction of propagation is not detected accurately for a compression factor larger than 0.6. The significance level for LC1→TI7 falls also below the 3σ threshold above a compression factor of around 0.85.

Impact of Compression on Transfer Entropy: The impact of compression on transfer entropy and its significance level is shown in Figure 10.14. Compared to cross-correlation and the nearest neighbours method, the influence of compression is minor. Both transfer entropy value and significance level are above the detection threshold, 0 and 3 respectively, for all investigated relationships. An explanation could be that the piecewise linearisation allows the detection of dependency just as easily with a limited number of samples when constructing the conditional PDFs. The conditional PDF is distributed over a $k + l + 1$-dimensional space. If the conditional PDFs are constructed from linear functions this will be reflected by linear curves in the PDFs which are also the shortest distance that can be represented in the cubic structure of the joint PDFs. Thus, with a finite number of samples, more samples will fall in the bins of the linear function and hence give a better estimation of the PDF.

Conclusions: All three causality methods are to some extent robust against the information loss of compression. This is, however, only the case when both measurements, the potential cause and the potential effect, are compressed in the same way as done in this chapter. Transfer entropy is more robust to compression than the nearest neighbours method which still performs better than the cross-correlation method.

Chapter 10 Summary

In this chapter, quantisation and compression was formulated and the impact on the data-driven causality measures was investigated. The Box Car Back Slope algorithm used in the PI data historian has been studied. Quantisation and compression affect the performance of all three measures. Transfer entropy is the most robust measure for both compression and quantisation.



Part IV

Conclusions

Chapter 11

Summary of Data-Driven Causality Measures

This chapter explains when to use which causality method. It also gives a comparison to other data-driven methods by discussing the characteristics from Section 3.1.2 on process monitoring systems. A decision tree is developed to decide which method to use depending on the data characteristics. Furthermore, a summary of the steps to conduct the causal analysis is provided by a causal analysis toolbox.

In the first section of this chapter, the results obtained by applying the developed causality measures of cross-correlation, nearest neighbours and transfer entropy to simulated and process data are summarised and discussed. Recommendations for the choice of method based on the data characteristics are given and a decision tree is developed. In Section 11.2, the benchmark criteria by Venkatasubramanian et al. [141] that were introduced in Section 3.2 are evaluated and discussed. In the last section of this chapter, all facilities that are needed to conduct the causality analysis are comprised in a causal analysis toolbox including fault detection, data pre-processing, the causality analysis and the automatic generation of a causal map.

11.1 Finding the Best Method for a Data Set

In Part II, three causality measures were introduced. The first approach was a simple cross-correlation method that finds the dead time between two process measurements and argues cause and effect by establishing in which measurement the disturbance occurred first. The other two methods, based on the nearest neighbours concept and transfer entropy, find out whether past value of one measurement are good predictors of the future value of a second measurement rather than vice versa. To accomplish this task, transfer entropy estimates probability density functions while the nearest neighbours method compares the Euclidean

distances of the predicted future values. All three methods are statistical methods that can investigate deterministic as well as random time sequences.

The question now is, which method to apply for which disturbances? One solution is to simply apply all three methods and compare the results for consistency. However, the computational effort is high, in particular for the transfer entropy and nearest neighbours method, so that it is not desirable to carry out all three calculations. The choice of the method depends foremost on the data characteristics. Thus, in the following section the impact of the data characteristics on the causality measures are discussed and tests for detecting these characteristics are devised. This discussion together with the conclusions from the previous chapters are used to derive a decision tree for deciding which method to use. The decision tree is given in the Section 11.1.3.

11.1.1 Data Characteristics

The two causality methods of transfer entropy and nearest neighbours work on the principle of predictability. Thus, any structure in a time trend that can predict the future value of a second time trend can be used to argue cause and effect. The structure can be either deterministic or stochastic, linear or nonlinear. If the disturbance is an oscillation then the surrogates that are used to establish a significance level are particulary important since they describe nonlinearity. It must be kept in mind that the causality measure using cross-correlation can only be applied if a dead time exists between two variables. The following paragraphs will reflect on the data characteristics and their implication for the causality measures. Tests for detecting nonlinearity, determinism and oscillations are available so that the knowledge of the nature of the data can be used to select the most promising measure of causality.

Stochastic and Deterministic Signals: All methods investigated here are statistical tools for random signals. However, deterministic signals can also be investigated when containing a certain degree of uncertainty. Deterministic signals contain information that cannot be investigated with statistical measures. In particular, transfer entropy uses probability density functions that omit the deterministic structure of the signal and thus loses information inherent in the time sequence although a certain degree of determinism is still sustained by investigating the relationship between past and future values. The same is the case for the nearest neighbours method. A differentiation can be made between structured and unstructured random signals. For example, a structured relationship exists if one value relates to the next value via a quadratic function. This will not be picked up by the linear method of cross-correlation. However, both transfer entropy and nearest neighbours will capture the relationship in either the probability density function or by mapping the nonlinear relationship in a higher dimension phase portrait. Tests of determinism were described in Section 6.2.2 and include the delta-epsilon method by Kaplan

[60] or the delay vector variance method by Gautama et al. [31]. Both methods are based on the nearest neighbours principle.

Oscillations: The causality methods can all investigate both oscillatory and non oscillatory signals. Oscillatory signals show the same pattern with a periodic repetition. They are deterministic since they can be described through sums of sinusoidal functions with a varying period of oscillation. A difficulty with using the cross-correlation method for oscillatory signals is that the dead time, from which cause-and-effect relationships are argued, is ambiguous. A dead time T_d can also be interpreted as a delay of $T_{osc}/2 - T_d$ because of the symmetry of an oscillation. A deterministic oscillation that does not change its shape as it travels from one measurement to the next cannot be detected with any of the methods as since directionality in both directions, A causes B to the same extent as B causes A, will be observed. Tests whether a disturbance is of oscillatory nature are given by Hägglund [41] as well as Thornhill and Hägglund [132]. An oscillation index of periodicity is derived that gives an indication of the amount of oscillation present in a signal.

Linear and Nonlinear Signals: Nonlinearity in the context of time series has two meanings. The two types of nonlinearity are as follows:

- Type 1: The signal originates from a nonlinear process;
- Type 2: There is a nonlinear functional relationship between two signals.

A signal originating from a nonlinear process (type 1) has harmonics in its frequency spectrum and shows phase coherence between these harmonics. Test for detecting type 1 nonlinearity were discussed in Section 6.2.3 and include an approach based on finding the nearest neighbours by Thornhill [137] and an approach based on higher order statistics by Choudhury et al. [19]. Although a linear tool in the sense of type 2 nonlinearity, the cross-correlation also works for nonlinear oscillations of type 1. The nearest neighbours approach is best suited for the type 1 nonlinear oscillation as it exploits the phase coherence. Type 1 nonlinearity of a stochastic signal does not manifests itself in harmonics in the frequency spectrum. A test for nonlinearity for time series of an unknown system is described by Mizuta et al. [83]. It is defined as the excess output variance from a system that cannot be described by a linear system. This nonlinearity can be captured both by transfer entropy and the nearest neighbours method. The cross-correlation approach, however, fails in this situation.

11.1.2 Results from Application

In this section, the conclusions of the previous chapters are compiled to give an overview of the best method for the various data characteristics. The conclusions are summarised in Table 11.1. These conclusions combined with the discussion from on linearity, oscillations

	Effect studied	Method	Sec.	Data analysed
A	Dead time	CCF	8.1	Simulated random noise
B	Low pass filtering	TE	8.2	Simulated random noise
C	Additive noise	TE, CCF	8.3	Simulated random noise
D	Nonlinear oscillation	NN	4.2	Reference case study
E	Irregular disturbance with linear structure	TE	9.2	Eastman case study
F	Linear oscillation and structured noise	TE	9.3	BP case study
G	Quantisation	TE, NN	10.1	Reference case study
H	Compression	TE	10.2	Reference case study

Table 11.1: Summary and best method of causality methods.

and randomness will then be used in the next section to derive a flowchart that allows the selection of the best method when the data characteristics are known.

(A) Dead Time The three methods of cross-correlation, nearest neighbours and transfer entropy can all detect causality through the presence of dead time. Cross-correlation function gives the best result since it also provides the length of the dead time. Furthermore, the following observations can be noticed.

- If dead time T_d is known then the prediction horizon h for transfer entropy and the nearest neighbours method should be $h = T_d$;
- Larger embedding dimensions m in the nearest neighbours method have more chances to capture the causality but will result in a lower significance level;
- A prediction horizon h lower than dead time T_d can still detect the causality in both transfer entropy and nearest neighbours method but if h is larger than T_d then the causality is missed;
- A lower time delay κ in the nearest neighbours method has a higher probability of finding the dependency if embedding dimension m is large enough.

(B) Low Pass Filter Both nearest neighbours method and transfer entropy can capture causality to some extent in the presence of low pass filtering effects. Transfer entropy is better suited than the nearest neighbours method to detect the directionality caused by low pass filtering effects. The order of the filter appears to have no major impact on the causality.

(C) Additive Noise All three causality measures are affected by the presence of additive noise so that the performance of the measures decreases. The effect of additive noise is negligible when analysing a pure dead time between two signals. When exploiting the low pass filtering effect, transfer entropy can cope with additive noise more successfully than the nearest neighbours method.

(D) Nonlinear Oscillation In the reference case study, the causal relationship between the variables with different amounts of nonlinearity can be argued. The cross-correlation function detects the most important relationships but also gives the correct dead time between the events. The most complete causal map can be obtained using the nearest neighbours method. The reason for this is that the phase coupling which occurs in nonlinear signals is best utilized with this approach.

(E) Irregular Disturbance with Linear Structure The results obtained by the causality analysis all comply with the verified root cause. The order of events of the measurements is identical for all three methods. However, the nearest neighbours method does not detect all the events but only the ones closest to the root cause. Transfer entropy gives the most complete picture of causal relationships. Cross-correlation correctly identifies the dead times between the occurrence of the disturbance peaks so that the time of events can be argued.

(F) Linear Oscillation and Structured Noise In the example of a linear oscillation, cause-and-effect relationships can be detected with the methods of cross-correlation, nearest neighbours and transfer entropy. The detection is assigned to the combination of the oscillation and structured noise since a purely linear oscillation would give no indication to cause and effect. The detected relationships vary for all three methods but can all be incorporated in the same causal map. Transfer entropy gives the complete causal map by detecting all dependencies.

(G) Quantisation Quantisation affects the performance of all three causality measures. A quantisation factor up to $f_q = 2$ can still be dealt with by all the causality measures. For the case that $f_q > 3$, the nearest neighbours and cross-correlation method gave incorrect results while transfer entropy detected no causality instead of an incorrect direction. Thus, transfer entropy is the most robust method to quantisation compared to the other two causality measures.

(H) Compression All three causality methods are to some extent robust against the information loss of compression. This is, however, only the case when both measurements, the potential cause and the potential effect, are compressed similarly. Transfer entropy is more robust to compression than the nearest neighbours method which still performs better than the cross-correlation method.

11.1.3 Decision Tree for the Selection of Causality Method

The decision tree shown in Figure 11.1 comprises the conclusions from case studies in Table 11.1 and the discussion of Section 11.1.1. First, the duration of the disturbance has to be ascertained. If the disturbance only occurs for a short time or if only a short data sample is available then the nearest neighbours method is the best method since it only requires around 400 samples. Short data sets are common in batch processes to which

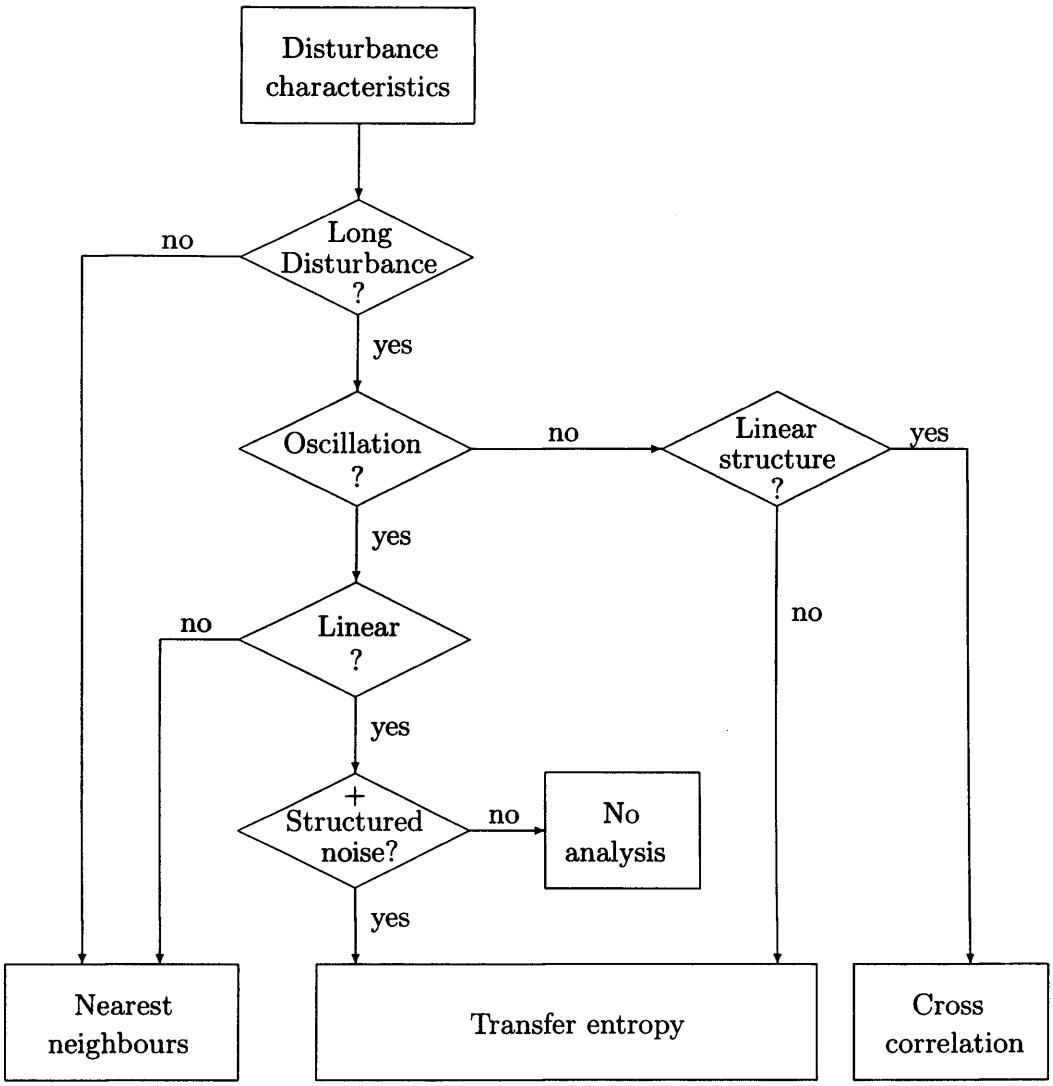


Figure 11.1: Decision tree for deciding on the analysis method depending on the data characteristics.

this method could also be applied. If the data set is long, around 2000 samples or more, then further question about the nature of the disturbance can be asked. As a next step, a check whether the disturbance is of oscillatory nature is conducted. If the disturbance is not oscillatory, then a nonlinearity test can be applied. In case of a linear disturbance together with a dead time between the measurements, the method of cross-correlation is the best choice since it also provides information about the size of the dead time. If the disturbance is nonlinear, transfer entropy can capture any nonlinear dependencies.

If the disturbance is detected to be an oscillation, the question arises whether the oscillation is linear or nonlinear. In case of a nonlinear oscillation, the nearest neighbours approach should give good results as it captures the phase coupling of the nonlinearity. In case of a linear oscillation, it has to be ascertained whether the oscillation contains additional structured noise. For a purely sinusoidal oscillation no statement can be made towards any cause-and-effect relationship. Transfer entropy gives good results for the analysis of the combination of oscillation and structured noise.

Characteristic	CCF	NN	TE	Eval.
Fast diagnosis	(No)	No	No	- -
Isolability	No	No	No	-
Robustness	Medium	High	High	+
Novelty Identification	(Yes)	Yes	Yes	+
Adaptability	Yes	Yes	Yes	+
Explanation facility	Yes	Yes	Yes	++
Multiple faults	No	No	No	-
Classification error	No	No	No	-
Modelling requirements	None	None	None	++
Storage requirements	Medium	Medium	High	-
Computational effort	Medium	High	High	- -

Table 11.2: Evaluation of benchmark criteria for the causality measures developed in this work (CCF: cross-correlation function, Chapter 5; NN: nearest neighbours method, Chapter 6; TE: transfer entropy, Chapter 7).

11.2 Evaluation of Benchmark Criteria

In Section 3.1.2, the benchmark criteria by Venkatasubramanian et al. [141] were listed. These criteria are desirable for any diagnostic systems, model-based or data-driven to allow a generic comparison. The different properties, features and expenses are summarised in Table 11.2. The last column of Table 11.2 gives an approximate rating of the performance of the methods in comparison to other diagnostic approaches. The most dominant advantage of the causality measures are the explanation facility that is not featured by most fault diagnosis methods. Furthermore, the lack of model requirements is a strong benefit since most processes in the chemical industry have no existing model and the derivation of a model would show to be tedious and time consuming. The most important disadvantages are the length of the diagnosis and the computational effort which is extremely high for the nearest neighbours and transfer entropy method. Thus, online fast fault diagnosis is not possible. A more detailed discussion of the benchmark criteria follows below. An advantage that is not included in the list by Venkatasubramanian et al. is the non-invasiveness of the methods. No changes, such as step tests, have to be made to the process.

Fast Detection and Diagnosis: Fast detection and diagnosis concerns the early analysis of faults. For all three causality methods a data set has to be captured for a certain time length and is then analysed. The length of the data set varies for the method. The nearest neighbours method requires about 400 data points (see Section 6.3.7) while the transfer entropy method requires for the probability density function estimation more than 2000 samples (see Section 7.5.3). The estimation of the cross-correlation function requires

more than a few hundred samples (> 300) [97]. The fast diagnosis of a disturbance is of less importance when conducting it off-line. Disturbances that affect the performance of a process but still allow a certain quality of the product are often persistent for a couple of weeks or months. If this is the case, the early diagnosis is not of high priority to the process engineer. Thus, the methods are for plant audit purposes rather than for online use.

Isolability: Isolability is the ability of the diagnostic tool to differentiate between different faults, that is, the tool not only detects a fault but also places it into a fault category. The causality methods are not able to differentiate between fault classes. Only some deductions can be made by looking at the fault propagation pattern that might vary for different faults.

Robustness: A number of experiments concerning the robustness of the methods were carried out in the previous chapters. In Section 8.3, the impact of additive noise was investigated. All three methods were significantly robust to uncorrelated noise. Quantisation and compression were investigated in Sections 10.1 and 10.2 and showed that the variable can be both quantised and compressed and still give some useful results.

Novelty Identification: This is the ability to decide if a detected fault is due to a known fault class or to an unknown fault class. Since no fault classes are formed for the propagation analysis, this is not quite applicable to the causality measure. However, both known and unknown faults can be identified with the methods of nearest neighbours and transfer entropy as long as they have some structure from which dependency can be deduced. The cross-correlation method can only deal with a new fault if it causes dead time between two measurements.

Adaptability: The ability to adapt to process operating changes due to changing environmental conditions is a property of all three measures. The reason for this is that the data is normalised before analysis so that a change of the operating conditions does not have an impact on the results. In case of setpoint changes during the investigated time, the controller error can be used for analysis rather than the process variable.

Explanation Facility: The main purpose of investigating the cause-and-effect relationships is to find the fault of origin and root cause as well as propagation to the current situation. This is therefore the main advantage of the methods developed in this work. In most cases, process insight is still required to interpret the results, however the results focus attention onto the area of the process where the fault is located.

Multiple Faults: None of the causality measures is able to deal with several faults present in the same measurements. The reason for this is that the question if A influence B or B influences A is a yes or no decision and does not deal with multiple answers. However, this question has not been addressed in this work. A way around this problem might be to conduct the analysis with different sets of parameters, say a high value for

the prediction horizon for a slow disturbance and a low value for the prediction horizon for a fast disturbance.

Classification Error: The classification error is an a-priori estimate of the significance level. Since no fault classes are given, this property cannot be addressed by the causality measures.

Modelling Requirements: The data-driven approaches pursued here do no require any model but rather the opposite is the case: the causality analysis gives a causal map that acts as a qualitative model. This is a major advantage of the pursued approach in comparison to model-based approaches.

Storage Requirements: In the case of off-line analysis, the storage requirement is linked to the fast diagnosis ability. Since a large number of data points is required to conduct the analysis there is a need for storage. However, fewer data points have to be stored for the transfer entropy method than for the nearest neighbours method. The reason for this is that transfer entropy estimates the probability density functions. The sample contributes to the PDF and can then consequentially be discarded. Thus, only the values of the PDF have to be stored. The nearest neighbours method, on the other hand, compares each data sample with all other data samples so that all samples have to be kept until the analysis is complete.

Computational Effort: The computational effort for the nearest neighbours and transfer entropy methods is significant. It is discussed for the former method in Section 6.3.2 and increases approximately with square of the number of samples. The computational effort for the transfer entropy method is given in Section 7.4.3 and increases approximately proportional with the number of samples. The high computational cost is also due to the calculation of the significance level which requires at the computation of the measures for the surrogates to achieve a significant comparison. The cross-correlation method has a comparatively low computational cost.

11.3 Causal Analysis Toolbox

In this thesis, several aspects and stages for the causality analysis for fault diagnosis have been discussed. For implementation in a process monitoring software tool, these stages have to be carried out one after the other. A summary for analysis is shown in Figure 11.2. In a first step, the presence of a fault is detected and a selection of measurements is made that show the same disturbance. A large number of methods have been developed for this purpose which were reviewed in Section 3.3. Further pre-processing normalises the data trends and removes unwanted frequency components through filtering. The following data subset is used to select the best method, as described in Section 11.1, and to select the parameters required for the analysis. For the nearest neighbours method, the parameter

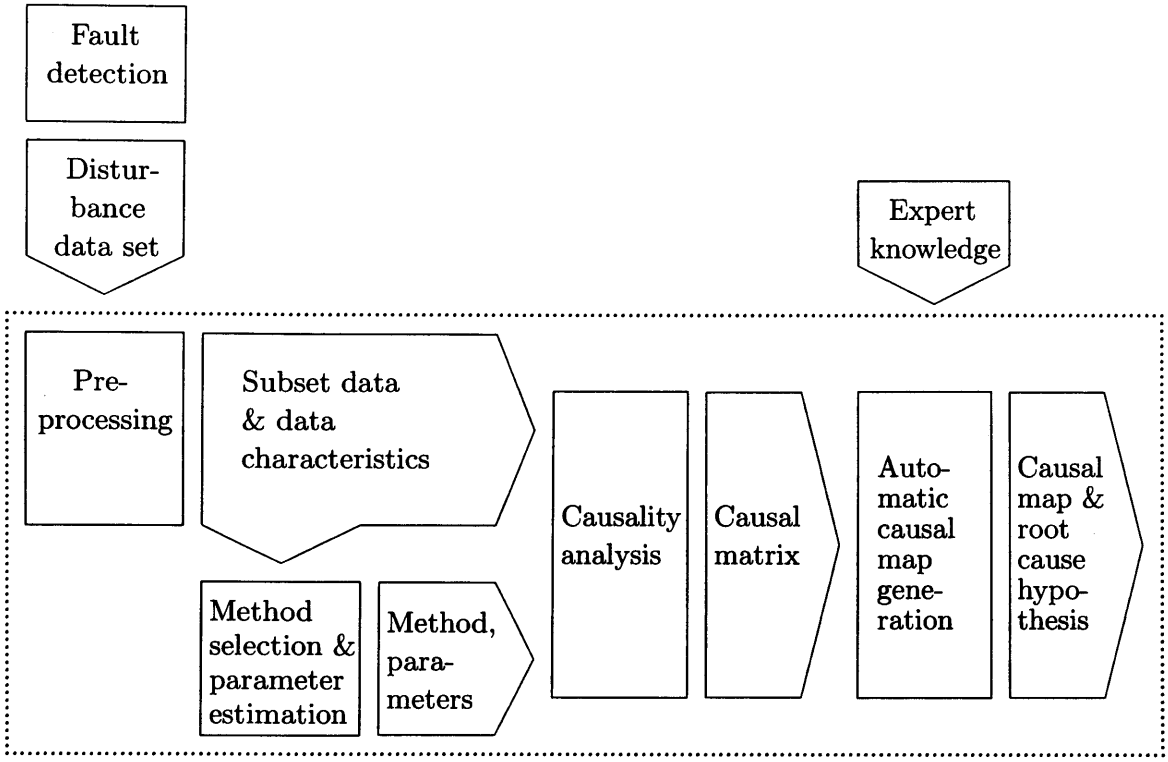
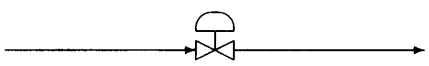


Figure 11.2: Flow chart of causal analysis toolbox.

estimation was described in Section 6.3.7 and for the transfer entropy method in Section 7.5. No parameters have to be adjusted for the cross-correlation method. The causality analysis is then conducted with the selected method (see Chapters 5, 6 and 7). The result of the analysis is the causality matrix that gives the relationship for all combinations of the p variables, thus it is of dimension pxp. The automatic causal map generation algorithm, described in Section 4.4, then generates a causal map from which the fault propagation path and a root cause hypothesis is deduced. In addition, expert knowledge is used to interpret the results and verify the hypothesis.

Chapter 11 Summary

In this chapter, characteristics, that is, randomness, oscillation and nonlinearity, of a data set have been discussed. The results of the industrial case studies and simulated effects from Part III have been summarised. Recommendations are drawn from these results and were incorporated into a decision tree. The benchmark criteria by Venkatasubramanian et al. [141] have been discussed for the developed methods. Finally, all elements of the causal analysis have been placed into context in a causal analysis toolbox.



Chapter 12

Discussions

Plant-wide disturbances in a chemical process affect a large number of measurements and decrease the performance of the process. Faults and disturbances can also put the safety of the process at risk or affect the quality of the product. A number of methods have been developed in the past decade to detect plant-wide disturbances at an early stage. Once the disturbance is detected, a further issue arises, that is, the diagnosis and elimination of the disturbance. The root cause analysis of a plant-wide disturbance is not trivial since a large number of measurements are affected.

The purpose of this work was to develop data-driven approaches to isolate the root cause of a plant-wide disturbance using historical process data. The objective of finding a data-driven method instead of a model-based method was motivated by the fact that a model exists for only about 5% of all chemical processes, while historical process data is readily available in all processes. The desired explanatory methodology should be designed to carry out the following tasks:

- Find the origin of the fault (root cause);
- Identify cause and effect relationships;
- Explain the propagation from the root cause to the locations at which the fault is observed (current situation).

These three objectives can be achieved by developing a qualitative model in the form of a causal map. In a causal map, all measurements are represented by nodes and the detected relationships by arcs pointing from one node to a second node. The root cause is most likely to be the node with arcs pointing from it but no arcs pointing to it. The propagation to the current situation can be explained if at least one arc points towards each node.

Causality relationships are deduced from fault propagation effects. These effects are dead time and low pass filtering. Cause and effect is argued from the presence of a dead time observed in two time trends. The rationality behind this argument is that a feature that

occurs at a first variable and then after a dead time T_d at a second variable is caused at a point that is closer to the first variable than the second. The second effect of low pass filtering occurs because most process equipment acts as a low pass filter. High frequency components are removed as the disturbance travels along the process equipment. Thus, the measurement with the "spikier"-looking time trend is thought to be closer to the origin of the disturbance.

Review of Methods

In this work, three methods for identifying cause and effect relationships have been proposed: cross-correlation, nearest neighbours and transfer entropy. The measures give a causality matrix when investigating the relationships between all combinations of process variables. An automated algorithm has been proposed in Section 4.4 to construct a causal map from a causality matrix and thus find the root cause and explain the propagation to the current situation.

The first method uses the *cross-correlation function* and was introduced in Chapter 5. In case of dead time between two measurements, the cross-correlation function of two time trends has a maximum at the time delay that is equal to the dead time. Since all cross-correlation functions have a maximum, two indices are introduced to measure the significance of the detected dead time. The correlation index establishes whether the correlation between the two time trends is high enough, that is, whether the two trends are similar enough. The oscillation index checks if there is a large difference between the minimum and maximum of the cross-correlation function. If the difference is too small, neither the time lag of the minimum or maximum are identified as the signal dead time. Cross-correlation works best in case of dead time between two measurements and an irregular disturbance. An advantage of cross-correlation is that the result gives also the detected dead time. It is also the method that requires the least computational effort. Since cross-correlation is a linear statistic, nonlinear transfer functions between two measurements cannot be analysed.

Chapter 6 describes a causality method using *nearest neighbours*. The method is based on the concept of predictability. Embedded vectors are constructed to represent the phase space. The nearest neighbours of the embedded vectors are found and future images are assigned. If the image of B associated with an embedded vector of A is similar to the images of B that are associated with the nearest neighbours of the embedded vector of A, then A is a good predictor of B. Causality of A to B is inferred by comparing if A is a better predictor of B than B is of A. This statistical approach is a very recent branch and a number of alternative algorithms have been developed by a number of authors. These alternative variations are compared in this work using industrial process data and the best method is selected for future use. Parameters to adjust are, for the construction of the embedded vectors, prediction horizon h, time delay κ and embedding dimension m. A significance level was established using surrogate data points to establish whether the

resulting causality is significant enough to be recorded. The nearest neighbours method uses dead time and low pass filtering effects to argue cause and effect but also a structure in the data that is difficult to measure with other statistical approaches. The advantage of the nearest neighbours method is that short time trends can be investigated. It performs best for nonlinear oscillations as shown in the reference case study. For other data sets, it did not find as many dependencies as the other two causality measures of cross-correlation and transfer entropy.

Transfer entropy has been recently proposed by Schreiber [113] to measure information transfer. Causality is argued by finding out whether A transfers more information to B than B to A, see Chapter 7. The basis for transfer entropy are transition probabilities that measure the probability of the future value of B having a certain value if the past values of A had their certain values. Thus, a similar concept of predictability as in the nearest neighbours method establishes cause and effect. The transition probabilities are estimated using the Kernel estimation. In this work, transfer entropy was proposed for fault analysis and a significance level established using surrogate time trends. The parameters to adjust are, similar to the nearest neighbours method, the prediction horizon h and the time delay κ. Transfer entropy gave the best results for the industrial case studies of an irregular disturbance and a linear oscillation in Chapter 9 as well as for the simulated data of low pass filtering and additive noise effects. On the other hand, a disadvantage of transfer entropy is that a large number of data points is required to estimate the transition probabilities.

Limitations and Implications

At a first glance, interpreting the results from statistical methods gives the impression of reading tea leaves because no mathematical proofs for stability or robustness exist and the result is made with a confidence level smaller than 100%. Model-based approaches have the advantages of giving an exact and certain result. However, since the nature of the data in chemical process like any real life data is stochastic, statistical tools are best suited for describing and analysing the disturbance. The structure that is inherent in the real process data is also difficult to simulate.

The causality analysis, however, will encounter difficulties if more than one fault is present in the same data trends. Multiple faults have not been studied in this work but it is expected that the more dominant disturbance will result in a cause and effect relationship. It is, however, difficult to predict which disturbance will be more dominant than the other. Interpretation of the results will prove to be difficult.

A further difficulty is the case of bidirectional coupling. If A influences B but B also influences A, then no directionality can be established. This is the case for the relationship between the process variable (PV) and the controller output (OP) in a control loop. Variations in the PV will cause the controller to react and thus affect the OP. The OP at the same time as an impact on the PV. No causality can be detected in this case with

the methods discussed here. An example of the interaction between process variable and controller output is included in Appendix D.

The algorithm for the automatic generation of the causal map in Section 4.4 arranges all process variables in a row in the order of their occurrence. Detected causalities that do not fit into this optimised scheme are discarded. This procedure does not allow the incorporation of recycle paths. For example, the situation might be that A causes B, B causes C and C causes A. This situation cannot be represented by a line of events. The automated algorithm from Section 4.4 works well if the disturbance is caused at a specific point in the recycle. It is in fact desired to cut the recycle at a point and thus identify the origin of the disturbance. The disturbance, on the other hand, might be caused by the recycle itself, that is, a structural disturbance. In this case, the automated algorithm cannot explain the situation and gives a misleading indication of the root cause.

Recommendations

Guidelines and recommendations for the selection of the best method and selecting appropriate variables were incorporated in a decision tree in Section 11.1.3. A further recommendation concerns the *number of process variables* to be used in the causality analysis. In the case studies discussed in this thesis, a maximum number of ten process variables were investigated in the reference case study. The industrial case studies in Chapter 9 analysed eight and six variables. If the number of variables p is very large (>12) then the number of causality measures that have to be calculated increases with $p(p-1)$ and the computational effort will also increase with $p(p-1)$. It is recommended to limit the number of investigated variables to less than twelve due to computational bounds. Otherwise, the algorithm is capable of dealing with larger number of process variables if the user is prepared to wait for the results.

Future Research

For the last few years, commercial tools have been emerging which are dedicated to the analysis of plant-wide disturbances using historical process data, such as the plant-wide disturbance analysis (PDA) tool by ABB/University College London as described in Section 3.5. The causality measures developed in this work will find their place in these tools rather than control loop assessment tools that investigate single loops and variables.

Some open issues and extensions to the causality measures suggested for future research at this point include:

- The ability to investigate multiple faults, for example by filtering the signal and conducting the causality analysis for the filtered signals with different filter cut-off frequencies;
- The incorporation of model-based approaches, such as expert systems and fault trees, to verify and extend cause and effect graphs by upstream/downstream argumenta-

tion;

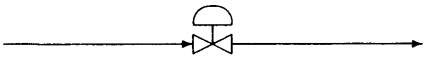
- The combination of causality measures with single indices that argue the root cause, such as the nonlinearity index (Section 6.2.3) or variability of the signal;
- The indication of directions in the causal map (signed digraph): is an "effect variable" increased or decreased as a result of a "cause variable"? A plus sign is attached to the vector in case if the cause variable is increased and a negative sign if it is decreased;
- The inclusion of low sample frequency laboratory data or data with other sampling frequencies. At present only time trends with the same sample frequencies can be included. More insight can be gained by incorporating lab data that is taken over longer intervals.

Chapter 12 Summary

This chapter has reviewed and given a critical evaluation of the work on data-driven causality measures presented in the thesis and highlighted direction for future research. The research in this work has led to data-driven measures that use historical data to argue cause and effect between process measurements in case of plant-wide disturbances. The use of causal maps has been discussed as a graphical representation for the process control engineer. The contributions of this thesis to chemical process monitoring and its main achievements are:

- **Three alternative methods for establishing cause and effect relationships (cross-correlation function, nearest neighbours method and transfer entropy);**
- **An algorithm for the automatic generation of a causal map from the three alternative causality methods;**
- **The introduction of three new industrial case studies with plant-wide disturbances affecting a large number of process variables and the successful application of the causality measures to the case studies.**

Critical evaluation has highlighted some suggestions for future work. The main criticism is the inability of the measures to deal with multiple faults affecting the same process variables and the inability of the causal map algorithm, not of the causality measures, to incorporate recycles. The main criticism will be overcome in future by frequency filtering to deal with multiple faults and by the incorporation of model-based approaches such as expert knowledge to argue the presence of a recycle. These approaches will lead to a robust and comprehensive solution to fault isolation of plant-wide disturbances.



Part V

Appendix

Appendix A

Correlation Derivations

A.1 Mean Value and Variance of Correlation Coefficient

The correlation coefficient is estimated from the N samples with index i of two variables x and y as follows:

$$\begin{aligned} r &= \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}} \\ &= \frac{1}{N} \frac{\sum_{i=1}^{N} x_i y_i}{\sigma_x \sigma_y} \end{aligned} . \tag{A.1}$$

Time sequences x and y have zero mean. The null hypothesis is defined such that x and y are uncorrelated.

$$\mathrm{H}_0 : \quad x,\, y \quad \text{uncorrelated} \tag{A.2}$$

Assuming H_0 gives the following estimate of the mean of correlation coefficient r.

$$\begin{aligned} \mu_r = \mathrm{E}\{r\} &= \mathrm{E}\{\frac{1}{N\sigma_x\sigma_y} \sum_{i=1}^{N} x_i y_i\} \\ &= \frac{1}{N\sigma_x\sigma_y} N \mathrm{E}\{xy\} \\ &= 0 \end{aligned} \tag{A.3}$$

The expectation value $\mathrm{E}\{xy\}$ is zero since x and y are assumed to be uncorrelated. The variance of the distribution of r can be estimated accordingly.

$$\begin{aligned} \sigma_r^2 &= \mathrm{E}\{r^2\} - \mu_r^2 \\ &= \mathrm{E}\{\frac{1}{N^2\sigma_x^2\sigma_y^2} \left(\sum_{i=1}^{N} x_i y_i\right)^2\} \\ &= \frac{1}{N^2\sigma_x^2\sigma_y^2} \mathrm{E}\{\left(\sum_{i=1}^{N} x_i y_i\right)^2\} \\ &= \frac{1}{N^2\sigma_x^2\sigma_y^2} \mathrm{E}\{\sum_{i=1}^{N} x_i^2 y_i^2\} \\ &= \frac{1}{N^2\sigma_x^2\sigma_y^2} N\, \mathrm{E}\{\sum_{i=1}^{N} x^2 y^2\} \\ &= \frac{1}{N\sigma_x^2\sigma_y^2} \sigma_x^2 \sigma_y^2 \\ &= \frac{1}{N} \end{aligned} \tag{A.4}$$

Thus, the variance of the correlation coefficient r is $\sigma_r = \frac{1}{\sqrt{N}}$. For a 3σ test, the fact that r can be positive and negative (two tailed) must be taken into account. Thus, x and y

are correlated with a probability of 99.74% if the correlation coefficient lies outside the interval defined by

$$|r| \geq \frac{3}{\sqrt{N}}. \tag{A.5}$$

A.2 Autocorrelation Function of Sine Wave

The formula of the autocorrelation function of a sine wave is used in Equation 5.12, Section 5.2. The derivation of the formula is as follows.

Consider the continuous form of the autocorrelation function (ACF) from for example [35]:

$$\phi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)x(t-\tau)\mathrm{d}t. \tag{A.6}$$

The function $x(t)$ is limited here to a finite interval since an infinite sequence would have infinite power and the ACF would hence also be infinite. The sine function is therefore defined on the interval from 0 to T_p as:

$$x(t) = \begin{cases} \sin t, & \text{for } 0 \leq t \leq T_p, \\ 0 & \text{elsewhere} \end{cases} \tag{A.7}$$

Using the symmetry property of the ACF ($\phi_{xx}(\tau) = \phi_{xx}(-\tau)$) and the fact that the function is limited to an interval gives the following form for the ACF of the sine wave.

$$\phi_{xx}(\tau) = \begin{cases} \frac{1}{T_p}\int_{\tau}^{T_p} \sin t \sin(t-\tau)\mathrm{d}t & 0 \leq \tau \leq T_p \\ \phi_{xx}(-\tau) & -T_p \leq \tau \leq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{A.8}$$

The integral is calculated from the boundaries τ to T_p since this is the overlap of the non-zero parts of the functions $\sin t$ and $\sin(t-\tau)$. Only the case for $0 \leq \tau \leq T_p$ has to be calculated while the other cases can be deduced from it:

$$\phi_{xx}(\tau) = \frac{1}{T_p}\int_{\tau}^{T_p} \sin t \sin(t-\tau)\mathrm{d}t \tag{A.9}$$

Using the equality: $\sin a \sin b = \frac{1}{2}[\cos(a-b) - \cos(a+b)]$, see [10].

$$\begin{aligned} \phi_{xx}(\tau) &= \tfrac{1}{2T_p}[\textstyle\int_{\tau}^{T_p} \cos\tau - \cos(2t-\tau)\mathrm{d}t] \\ &= \tfrac{1}{2T_p}[\textstyle\int_{\tau}^{T_p} \cos\tau\mathrm{d}t - \int_{\tau}^{T_p} \cos(2t-\tau)\mathrm{d}t] \end{aligned} \tag{A.10}$$

For the second integral, the substitution method [10] can be applied as follows: $t = \varphi(z) = \frac{z}{2} + \frac{\tau}{2}$; $\frac{\mathrm{d}t}{\mathrm{d}z} = \frac{\mathrm{d}\varphi}{\mathrm{d}z} = \frac{1}{2}$; $z = \psi(t) = 2t - \tau$ and hence for the integral boundaries: $\psi(\tau) = \tau$; $\psi(T_p) = 2T_p - \tau$.

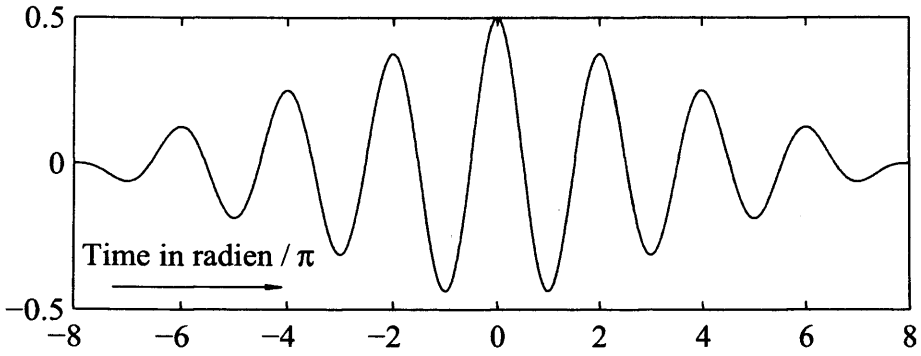


Figure A.1: Autocorrelation function of a sine wave of length $T_p = 8\pi$.

$$\begin{aligned} \phi_{xx}(\tau) &= \frac{1}{2T_p}\left[\int_{\tau}^{T_p} \cos\tau \mathrm{d}t - \int_{\tau}^{2T_p-\tau} \cos z \mathrm{d}z\right] \\ &= \frac{1}{2T_p}\left[\cos\tau[t]_{\tau}^{T_p} - \tfrac{1}{2}[\sin z]_{\tau}^{2T_p-\tau}\right] \\ &= \frac{1}{2T_p}\left[(T_p-\tau)\cos\tau - \tfrac{1}{2}\sin(2T_p-\tau) + \tfrac{1}{2}\sin\tau\right] \end{aligned} \tag{A.11}$$

Replacing Equation A.8 by the derived expression now yields:

$$\phi_{xx}(\tau) = \begin{cases} \frac{1}{2T_p}\left[(T_p-\tau)\cos\tau - \frac{1}{2}\sin(2T_p-\tau) + \frac{1}{2}\sin\tau\right] & 0 \le \tau \le T_p \\ \phi_{xx}(-\tau) & -T_p \le \tau \le 0 \\ 0 & \text{elsewhere} \end{cases} \tag{A.12}$$

which is then used in Equation 5.12, Section 5.2. This autocorrelation function of a sine wave is shown in Figure A.1 for a time sequence of length T_p.

A.3 Cross-correlation Function of Two Delayed Sine Waves

The cross-correlation function (CCF) of a sine wave and a delayed form of the sine wave by half the oscillation period are given in Section 5.2. The analytical description of the CCF is shown in this section. Consider the continuous form of the cross-correlation function (CCF) from for example [35]:

$$\phi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)y(t-\tau)\mathrm{d}t. \tag{A.13}$$

The function $x(t)$ is here a function defined within an interval:

$$x(t) = \sin t, \text{ where } 0 \le t \le T_p. \tag{A.14}$$

and the second function $y(t)$ is a delayed version of $x(t)$:

$$y(t) = \sin(t-\tau_1), \text{ where } 0 \le t \le T_p \text{ and } -T_p \le \tau_1 \le T_p. \tag{A.15}$$

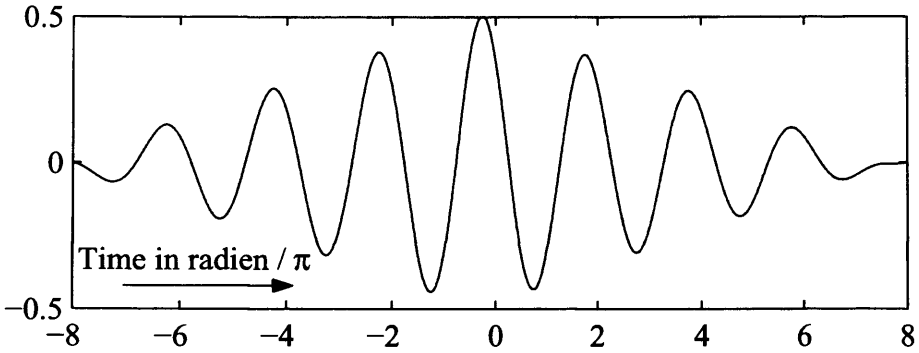


Figure A.2: Cross-correlation function of a sine wave and a sine wave delayed by $\tau_1 = \pi/4$, both signals of length $T_p = 8\pi$.

First, the CCF is calculated for interval $0 \leq \tau \leq T_p$. The CCF can be expressed by the following equation.

$$\phi_{xy}(\tau) = \frac{1}{T_p} \int_{\tau}^{T_p} \sin t \sin(t - \tau - \tau_1) \mathrm{d}t \tag{A.16}$$

Using the equality: $\sin a \sin b = \frac{1}{2}[\cos(a-b) - \cos(a+b)]$, see [10], yields.

$$\begin{aligned} \phi_{xy}(\tau) &= \tfrac{1}{2T_p}[\textstyle\int_{\tau}^{T_p} \cos(\tau + \tau_1) - \cos(2t - \tau - \tau_1)\mathrm{d}t] \\ &= \tfrac{1}{2T_p}[\textstyle\int_{\tau}^{T_p} \cos(\tau + \tau_1)\mathrm{d}t - \int_{\tau}^{T_p} \cos(2t - \tau - \tau_1)\mathrm{d}t] \end{aligned} \tag{A.17}$$

For the second integral, the substitution method [10] can be applied as follows: $t = \varphi(z) = \frac{z}{2} + \frac{\tau+\tau_1}{2}$; $\frac{\mathrm{d}t}{\mathrm{d}z} = \frac{\mathrm{d}\varphi}{\mathrm{d}z} = \frac{1}{2}$; $z = \psi(t) = 2t - \tau - \tau_1$ and hence for the integral boundaries: $\psi(\tau) = \tau - \tau_1$; $\psi(T_p) = 2T_p - \tau - \tau_1$.

$$\begin{aligned} \phi_{xy}(\tau) &= \tfrac{1}{2T_p}\left[\textstyle\int_{\tau}^{T_p} \cos(\tau + \tau_1)\mathrm{d}t - \int_{\tau-\tau_1}^{2T_p-\tau-\tau_1} \cos z \mathrm{d}z\right] \\ &= \tfrac{1}{2T_p}\left[\cos(\tau + \tau_1)[t]_{\tau}^{T_p} - \tfrac{1}{2}[\sin z]_{\tau-\tau_1}^{2T_p-\tau-\tau_1}\right] \\ &= \tfrac{1}{2T_p}\left[(T_p - \tau)\cos(\tau + \tau_1) - \tfrac{1}{2}\sin(2T_p - \tau - \tau_1) + \tfrac{1}{2}\sin(\tau - \tau_1)\right] \end{aligned} \tag{A.18}$$

For negative values of τ, $-T_p \leq \tau \leq 0$, the integral boundaries are 0 to $T_p + \tau$. The transformed boundaries are then $\psi(0) = -\tau - \tau_1$ and $\psi(T_p + \tau) = 2T_p + \tau - \tau_1$. The summarised CCF is then as follows:

$$\phi_{xy}(\tau) = \begin{cases} \frac{1}{2T_p}\left[(T_p - \tau)\cos(\tau + \tau_1) - \frac{1}{2}\sin(2T_p - \tau - \tau_1) + \frac{1}{2}\sin(\tau - \tau_1)\right] & 0 \leq \tau \leq T_p \\ \frac{1}{2T_p}\left[(T_p + \tau)\cos(\tau + \tau_1) - \frac{1}{2}\sin(2T_p + \tau - \tau_1) - \frac{1}{2}\sin(\tau + \tau_1)\right] & -T_p \leq \tau \leq 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{A.19}$$

The function $\phi_{xy}(\tau)$ is plotted for a time delay of $\tau_1 = \pi/4$ in Figure A.2.

Appendix B

Nearest Neighbours Derivation

B.1 Mean Value of Self-Predictability

In this section, the expected value of the self-predictability as required in Section 6.3.2 is derived for the case that the signal is uniform distributed random noise. The self-predictability is defined, after Equation 6.12 for an embedding dimension of $m = 1$ is:

$$D_i(X|X) = \frac{1}{K}\sum j = 1^K |x_i - x_{r_{i,j}}| \tag{B.1}$$

where x_i is the ith sample and $x_{r_{i,j}}$ are the nearest neighbours of x_i, Since X is a random function, the nearest neighbours indices $r_{i,j}$ can be regarded as independent of i, that is, $x_{r_{i,j}} = x_j$. The probability distribution of the sum of two independent variables with zero mean and symmetric probability distributions each equals to the probability distribution of the difference of the two independent variables and is:

$$F(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{-y+z} f(x,y)\mathrm{d}x\mathrm{d}y, \quad z = x \pm y \tag{B.2}$$

where $f(x,y)$ is the joint probability density function of x and y. For uniform probability density functions (PDF) with zero mean and standard variance, the joint PDF is:

$$f(x,y) = \begin{cases} \frac{1}{12}, & -\sqrt{3} \le x \le +\sqrt{3}, \quad -\sqrt{3} \le y \le +\sqrt{3} \\ 0, & \text{elsewhere} \end{cases} \tag{B.3}$$

Using this in Equation B.2 yields for all negative values of z

$$\begin{aligned} F(z) &= \int_{-\sqrt{3}}^{\sqrt{3}+z}\int_{-\sqrt{3}}^{y-z}\frac{1}{12}\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{12}\int_{-\sqrt{3}}^{\sqrt{3}+z} -y + z + \sqrt{3}\mathrm{d}y \\ &= \frac{1}{24}[z^2 + 4\sqrt{3}z + 6]. \end{aligned} \tag{B.4}$$

Thus, the PDF of $z = x \pm y$ results in

$$f(z) = \frac{\mathrm{d}F(z)}{\mathrm{d}z} = \frac{1}{12}z + \frac{1}{2\sqrt{3}}, \tag{B.5}$$

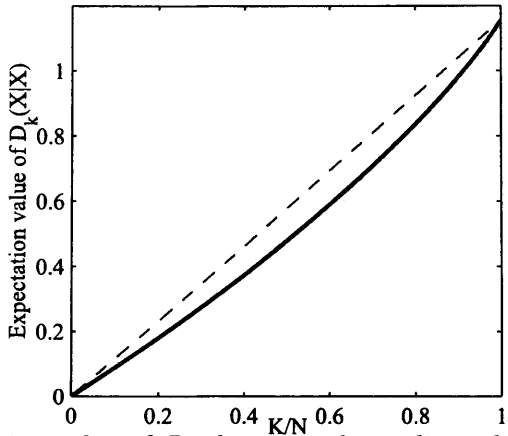


Figure B.1: The expectation value of D_i shows an almost linear behaviour on the number of nearest neighbours, K. N is the number of samples from which for which the self predictability should be estimated.

for $-2\sqrt{3} \leq z \leq 0$. Using the symmetric property of the PDFs of x and y gives the PDF for all z. The absolute value of $z' = |z|$ is positive with the same PDF shape as the PDF for z.

$$f(z') = \begin{cases} \frac{1}{\sqrt{3}} - \frac{1}{6}z' & 0 \leq z' \leq 2\sqrt{3} \\ 0 & \text{elsewhere} \end{cases} \tag{B.6}$$

The probability distribution is then

$$F(z') = \begin{cases} 0 & z < 0 \\ z'(\frac{1}{\sqrt{3}} - \frac{1}{6}z') & 0 \leq z \leq \sqrt{12} \\ 1 & z > \sqrt{12} \end{cases} \tag{B.7}$$

Let $\hat{K}$ be the percentage of K nearest neighbours compared to the number of samples from which the PDF is constructed: $\hat{K} = \frac{K}{N}$, such that $0 \leq \hat{K} \leq 1$. The probability distribution of $F(z') = \hat{K}$ now gives the value z'_1 that is the boundary for the PDF of the K nearest neighbours. The expected value of D_i will lie between zero and that boundary. Solving $F(z'_1) = \hat{K}$ for z'_1 results in

$$z'_1 = \sqrt{12}(1 - \sqrt{1 - \hat{K}}). \tag{B.8}$$

The expected value of D_i can now be calculated from the adjusted PDF of the K smallest values of all $|x_i - x_j|$ in Equation B.1. The adjusted PDF is

$$f(z') = \begin{cases} \frac{1}{\hat{K}}\left[\frac{1}{\sqrt{3}} - \frac{1}{6}z'\right] & 0 \leq z' \leq z'_1 \\ 0 & \text{elsewhere} \end{cases} \tag{B.9}$$

The expectation value or mean of D_i is then:

$$\begin{aligned} \mathrm{E}\{D_i\} = \mathrm{E}\{z'\} &= \int_{-\infty}^{\infty} z' f(z') \mathrm{d}z' \\ &= \int_0^{z'_1} z' \frac{1}{\hat{K}}\left[\frac{1}{\sqrt{3}} - \frac{1}{6}z'\right] \mathrm{d}z' \end{aligned} \tag{B.10}$$

and with $z_1' = \sqrt{12}(1 - \sqrt{1 - \hat{K}})$

$$\mathrm{E}\{D_i\} = \frac{4}{\sqrt{3}\hat{K}}[\sqrt{1 - \hat{K}} - 1] + 2\sqrt{3}[1 - \frac{2}{3}\sqrt{1 - \hat{K}}]. \tag{B.11}$$

The expected value for D_i if $\hat{K} = 1$ is equal to the expected value of $|x_i - x_j|$. The expected value of $|x_i - x_j|$ is $\int z' f(z')\mathrm{d}z' = \frac{2}{\sqrt{3}}$ and thus equals $D_i, \hat{K} = 1$. Figure B.1 shows the function of Equation B.11.

Appendix C

Entropy Derivation

The relationship between entropy, conditional entropy and mutual information is required to show that conditional entropy cannot be used as a causality measure. The definition of entropy is after Equation 7.24

$$G(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\mathrm{d}x. \tag{C.1}$$

where $p(x)$ is the probability density function of x. Mutual information was defined in Equation 7.26 as

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\mathrm{d}y\mathrm{d}x \tag{C.2}$$

where $p(x,y)$ is the joint probability density function of x and y. Conditional entropy is defined as

$$G(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x,y)}{p(y)}\mathrm{d}y\mathrm{d}x. \tag{C.3}$$

The sum of mutual entropy and conditional entropy can be modified as follows. For clarity, the integral boundaries from $-\infty$ to ∞ are omitted.

$$\begin{aligned}
I(X,Y)+G(X|Y) &= \int\int p(x,y)\log\tfrac{p(x,y)}{p(x)p(y)}\mathrm{d}y\mathrm{d}x - \int\int p(x,y)\log\tfrac{p(x,y)}{p(y)}\mathrm{d}y\mathrm{d}x \\
&= \int\int p(x,y)\log\tfrac{p(x,y)}{p(x)p(y)} - p(x,y)\log\tfrac{p(x,y)}{p(y)}\mathrm{d}y\mathrm{d}x \\
&= \int\int p(x,y)\left[\log\tfrac{p(x,y)}{p(x)p(y)} - \log\tfrac{p(x,y)}{p(y)}\right]\mathrm{d}y\mathrm{d}x \\
&= \int\int p(x,y)\left[\log\tfrac{p(x,y)p(y)}{p(x,y)p(x)p(y)}\right]\mathrm{d}y\mathrm{d}x \\
&= -\int\int p(x,y)\log p(x)\mathrm{d}y\mathrm{d}x \\
&= -\int\log p(x)\int p(x,y)\mathrm{d}y\mathrm{d}x
\end{aligned} \tag{C.4}$$

The integral over y can be dealt with separately and is the marginal density function, that is, $\int p(x,y)\mathrm{d}y = p(x)$. Thus,

$$I(X,Y)+G(X|Y) = \int p(x)\log p(x)\mathrm{d}x = G(X), \tag{C.5}$$

which is the relationship between mutual information, conditional entropy and entropy.

Appendix D

Two Interacting Controllers

In this chapter, the interaction between process variable and controller output is investigated using benchmark data sets from the literature. The interaction between controller output and process variable is of interest since it may give insight into the performance of the control loop. The purpose of analysing benchmark data is also to allow a comparison to other tools and methodologies. The first benchmark data investigated here is a data set from two competing control loops of an industrial pulp and paper process, a concentration and a flow controlled loop, courtesy of Alexander Horch [46]. The main purpose of this case study is to investigate the relationship between cause and effect of process variable and controller output. The second benchmark data comes from a valve stiction simulator by Shoukat Choudhury [19, 20].

Horch investigated in his PhD thesis the detection of static friction (stiction) of control valves by analysing the cross-correlation function [46]. He found that a phase shift between controller output and process variable of a quarter of the oscillation period is an indicator for stiction. This method was evaluated using an industrial case study from a paper mill. In this example, the relationship between PV and OP of two coupled oscillating loops are examined. Further advancement in this area were made by Choudhury [19] in his PhD thesis. Instead of analysing stiction from industrial data, the stiction is modelled so that the amount of stiction can be varied in a measurable way. In the following, the causality analysis will be applied to both industrial and simulated data.

D.1 Industrial Data

In this section, the data from the industrial case study was provided by Horch to investigate the causal relationship between controller output and process variable in feedback loops. This particular relationship is curious since the PV influences the OP and the OP the PV. In a sense, this circular influence is the same problem as a number of process measurements in a recycle path discussed in Section 9.3. However, the question of causality is a larger

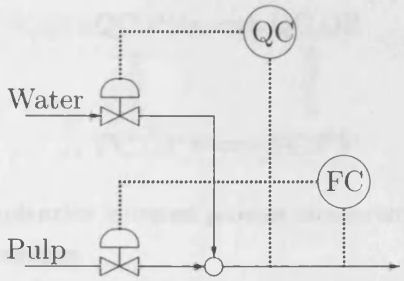


Figure D.1: Two interacting controllers act on the water and pulp inflow of a mixing process.

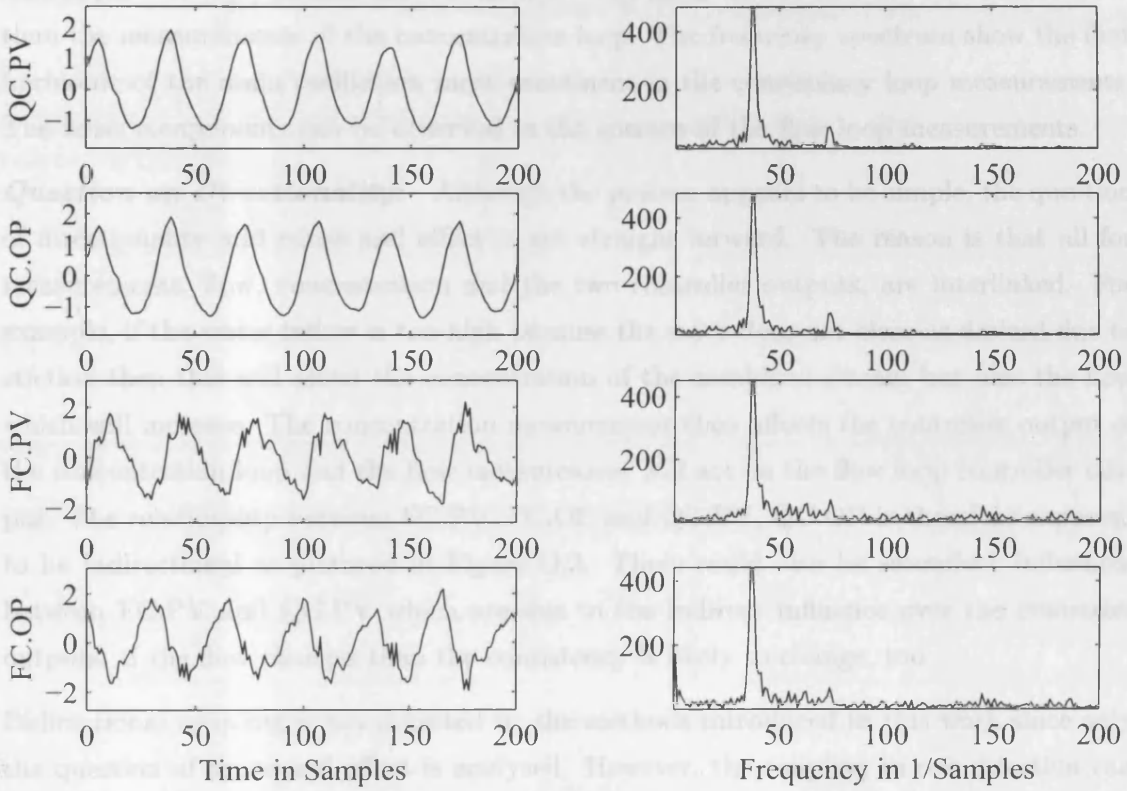


Figure D.2: Time trend and frequency spectra of process measurements from Figure D.1 when stiction in the consistency control valve occurs causing a limit cycle.

challenge as the dynamics of time delay, filtering and accumulative noise act in a shorter time frame.

Process and Data Set: In the process investigated in [46] paper pulp is diluted with water to achieve a desired consistency. The process schematic of this setup is shown in Figure D.1. The water inflow is controlled by the consistency measurement while the pulp inflow is controlled by the flow measurement. The consistency control valve shows a hysteresis caused by stiction in the valve while the flow control valve is operating as desired. The hysteresis causes a limit cycle oscillation. Due to the physical interaction between two controllers the oscillation can be seen in both measurements.

Time trend and frequency spectrum of the process with the limit cycle oscillation are shown in Figure D.2. Altogether 1196 samples were available courtesy of A. Horch. Both OP

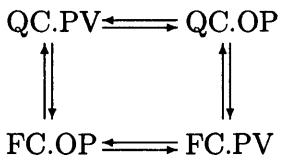


Figure D.3: Expected dependencies between process measurements of two interacting controllers argued from process understanding.

and PV from the concentration and flow loop show the same oscillation period of around 32 samples. The measurements of the flow loop looks noisier and thus more uncertain than the measurements of the concentration loop. The frequency spectrum show the first harmonic of the main oscillation most prominent in the consistency loop measurements. The noise components can be observed in the spectra of the flow loop measurements.

Question on Directionality: Although the process appears to be simple, the question of directionality and cause and effect is not straight forward. The reason is that all for measurements, flow, concentration and the two controller outputs, are interlinked. For example, if the water inflow is too high because the valve does not close as desired due to stiction then this will affect the concentration of the combined stream but also the flow which will increase. The concentration measurement then affects the controller output of the concentration loop and the flow measurement will act on the flow loop controller output. The relationship between FC.PV, FC.OP and QC.PV, QC.OP is therefore expected to be bidirectional as pictured in Figure D.3. There could also be secondary influences between FC.PV and QC.PV which are due to the indirect influence over the controller outputs: if the flow changes than the consistency is likely to change, too.

Bidirectional coupling is not detected by the methods introduced in this work since only the question of cause and effect is analysed. However, the coupling in one direction can be measured by transfer entropy and the nearest neighbours method if not the combined measure is investigated but the single measures. That is, for transfer entropy Equation 7.32 is not calculated $t_{x\to y} = T(Y|X) - T(X|Y)$ but rather the two entropy measures $T(Y|X)$ and $T(X|Y)$. The same is the case for the nearest neighbours method.

Interpretation of Results: Since the time signal of the investigated disturbance is strongly periodic, finding the time delay does not give a conclusion about the root cause. This is because a time delay from x to y κ_{xy} could also be interpreted as a time delay from y to x of $\kappa_{yx} = 2\pi - \kappa_{xy}$. This can easily detected if the time delay detected with the cross-correlation method between x and y plus the time delay between y and x is half the period of oscillation. When examining the following table and remembering that the oscillation period is 32 samples, this can be confirmed for all relationships between FC and QC. The following Table gives the detected dead times T_d between the four process measurements QC.PV, QC.OP, FC.PV and FC.OP.

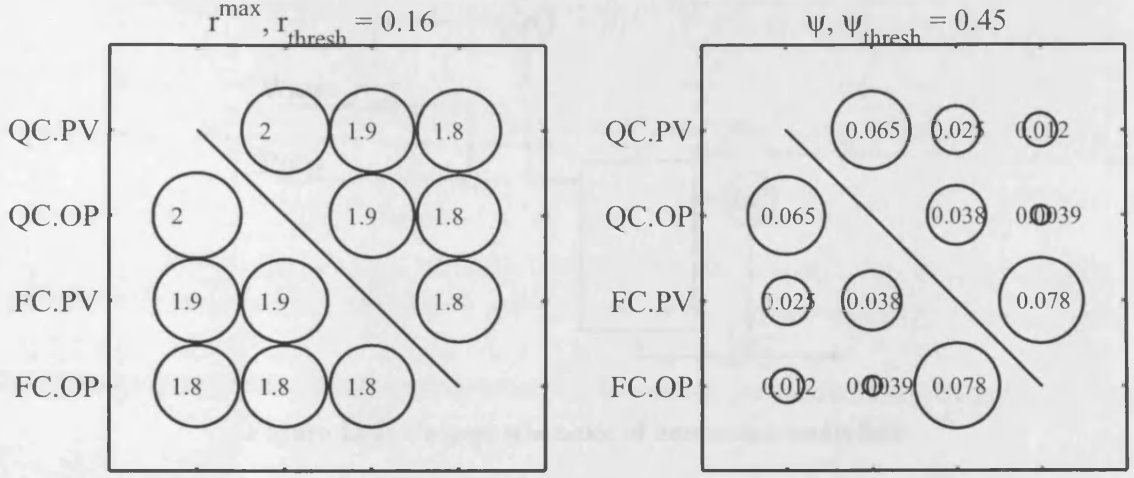


Figure D.4: Correlation and oscillation index for detected dead times of simulated data for two interacting controllers.

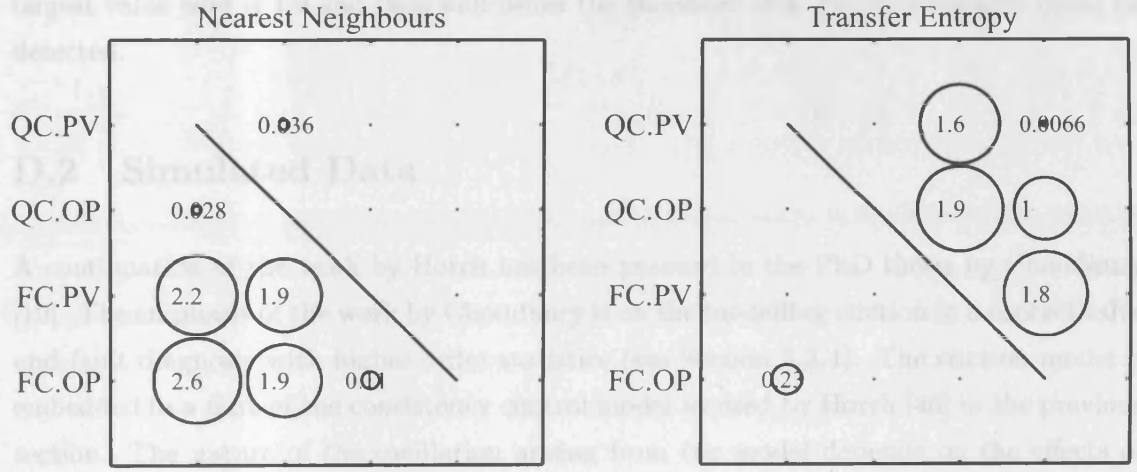


Figure D.5: Significance level of causality analysis on oscillating signals where $m = 4$, $h = \kappa = 1$.

x / y	QC.PV	QC.OP	FC.PV	FC.OP
QC.PV	0	1	6	-8
QC.OP	-1	0	6	-8
FC.PV	-6	-6	0	1
FC.OP	8	8	-1	0

Figure D.4 shows the corresponding correlation and oscillation indices. The correlation index in the left hand side of Figure D.4 is near to the maximum ($\max\{r^{\max}\} = 2$), and thus well above the threshold r_{thresh}. The oscillation index ψ in the right hand plot, however, is below the threshold ψ_{thresh} for all relationships between the variables. This means that there is only a small difference between the maximum and minimum of the cross-correlation function. The dead times are therefore not detected.

The results of the nearest neighbours and transfer entropy method are shown in Figure D.5. The left hand plot shows the bubble chart of the significance level using the nearest neighbours method. The largest significance level is for the causality measurement between

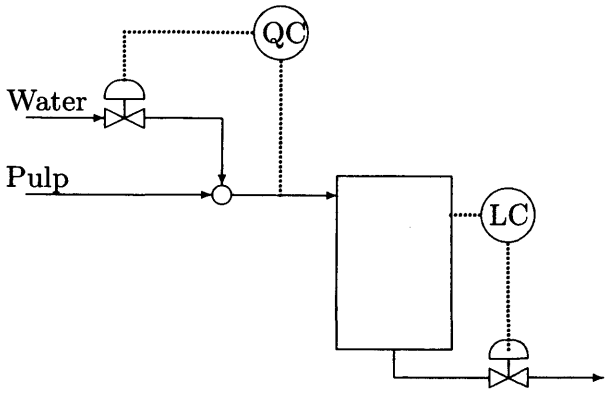


Figure D.6: Process schematic of interacting controllers.

FC.OP and QC.PV and has a value of 2.6 which is below the threshold value of 3. The right hand plot shows the bubble chart of the significance level using transfer entropy. The largest value here is 1.9 and thus well below the threshold of 3. No directionality could be detected.

D.2 Simulated Data

A continuation of the work by Horch has been pursued in the PhD thesis by Choudhury [19]. The emphasis of the work by Choudhury is on the modelling stiction in a control valve and fault diagnosis with higher order statistics (see Section 3.3.4). The stiction model is embedded in a part of the consistency control model as used by Horch [46] in the previous section. The nature of the oscillation arising from the model depends on the effects of stiction incorporated in the model. These effects are hysteresis, deadband and slipjump [20]. In the following, the process will be introduced and questions on directionality phrased. Results from the application of the causality measure will be discussed.

Process and Data Set: In the work by Choudhury [19, 20], process, controller and stiction are modelled. The representation of the processes as a process schematic is shown in Figure D.6. Similar to the process used by Horch [46], water is added to a pulp flow. The water inflow is controlled by the consistency. Instead of a flow controller for the pulp flow, the mixture is collected in a tank and any disturbances in the inflow are buffered by the tank which is controlled by a level. The model of the tank can be found in the earlier example in Section 2.3. The transfer function of the process there was identified as $1/As$ with A being the area of the tank. The process comprising the consistency control valve and the mixing is modelled of a low pass filter with time constant 1/10 and a time delay of 10 samples. Both level and consistency controllers are PI controllers. The stiction model by Choudhury allows settings for the undershoot, deadband and slipjump, for details see [20]. The settings here are for the undershoot SU=3, for the deadband SD=2 and for the slipjump J=1.

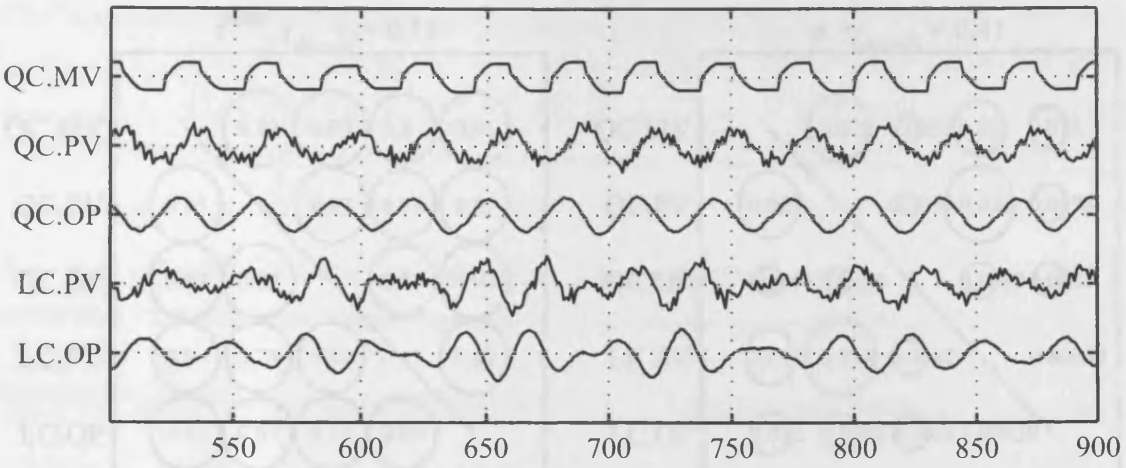


Figure D.7: Time trend of two interacting controllers.

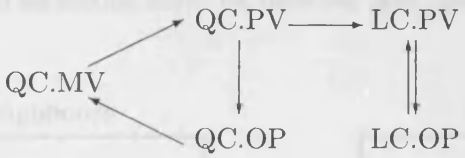


Figure D.8: Expected dependencies between process measurements of concentration and level loop.

The time trends of the process variables, controller outputs and the manipulated variables of the consistency valve with stiction are shown in Figure D.7. The nonlinearity in the signal can be most seen most prominently in the manipulated variable QC.MV which looks the least like a sinusoidal oscillation. Noise components are present only in the process variables since the PI controllers act as low pass filters and eliminate high frequency components. The level variable and output show an irregular signature on top of the oscillation frequency inherent in the other time trends.

Question on Directionality: Similar to the causality analysis of the consistency and flow control loops in the previous section, the questions on directionality is not simple. The potential cause and effect relationships are shown in Figure D.8. In the consistency loop, the process variable affects the controller output which affects the manipulated variable which in turn affects the process variable. Thus, a mini recycle is formed posing the question of the origin of the disturbance. In the model, the consistency is connected to the level via a time delay. Level and the controller output in the level influence one another in a bidirectional fashion. The task when applying the causality measures to the data shown in Figure D.7 is to confirm any of these directionalities.

Interpretation of Results: All three causality measures were applied to the simulated data set. The dead times resulting from the cross-correlation function method are listed in the following table

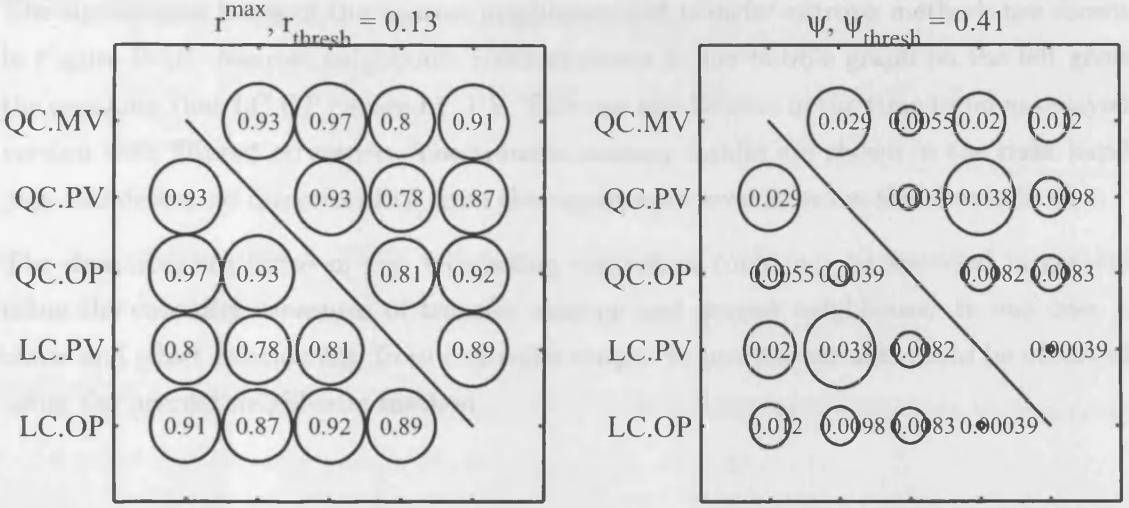


Figure D.9: Correlation and oscillation index for detected dead times of simulated data for two interacting controllers.

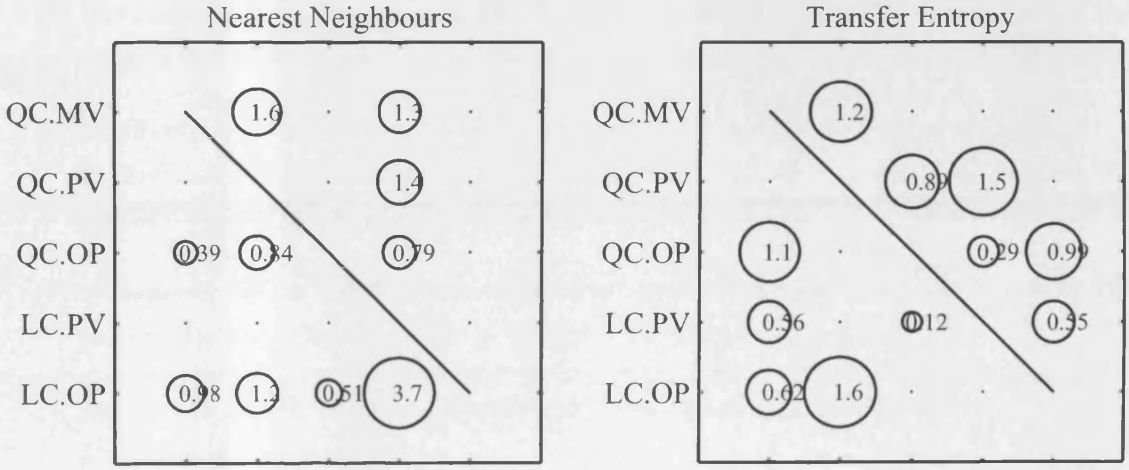


Figure D.10: Results nearest neighbours and transfer entropy, $h = \kappa = 4$, $m = 3$.

x/y	QC.MV	QC.PV	QC.OP	LC.PV	LC.OP
QC.MV	x	5	-3	23	15
QC.PV	-5	x	7	18	10
QC.OP	3	-7	x	11	18
LC.PV	-23	-18	-11	x	6
LC.OP	-15	-10	-18	-6	x

The correlation and oscillation indices are shown in bubble charts in Figure D.9. Like in the industrial data, all correlation $r^{\max}$ indices exceed the threshold r_{thresh} while all oscillation indices ψ are far below the threshold ψ_{thresh}. Thus, no dead time or hence directionality is detected between any of the process measurements. However, if looking at the expected directions in Figure D.8, the time delay match for the relationships between QC.MV, QC.PV and QC.OP. A time delay of five samples is detected between QC.MV and QC.PV, of seven samples between QC.PV and QC.OP and of three samples between QC.OP and QC.MV.

The significance levels of the nearest neighbours and transfer entropy methods are shown in Figure D.10. Nearest neighbours method shown in the bubble graph on the left gives the causality that LC.OP causes LC.PV. This can also be seen in the time trend as delayed version with filtered structure. The transfer entropy results are shown in the right hand plot and detect no directionality since the significance level is below the threshold of 3.

The directionality between two interacting controllers could not be specified in general using the causality measures of transfer entropy and nearest neighbours. In one case, a cause and effect relationship from controller output to process variable could be observed using the nearest neighbours method.

References

[1] Anderson, T.W., 1984. *An Introduction to Multivariate Statistical Analysis.* Chichester: Wiley.

[2] Arnhold, J., Grassberger, P., Lehnertz, K. and Elger, C.E., 1999. A robust method for detecting interdependences: Application to intracranially recorded EEG. *Physica D*, 134, 419-430.

[3] Avlonitis, S.A., Pappas, M., Moutesidis, K., Avlonitis, D., Kouroumbas, K. and Vlachakis, N., 2004. PC based SCADA system and additional safety measures for small desalination plants. *Desalination*, 165, 165-176.

[4] Becraft, W.R. and Lee. P.L., 1993. An integrated neural-network expert-system approach for fault-diagnosis. *Computers and Chemical Engineering*, 17, 1001-1014.

[5] Bhattacharya, J., Petsche, H. and Pereda, E., 2001. Long-range synchrony in the gamma band - role in the music perception. *Journal of Neuroscience*, 21, 6329-6337.

[6] Bhattacharya, J., Pereda, E. and Petsche, H., 2003. Effective detection of coupling in short and noisy bivariate data. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 33, 85-95.

[7] Bialkowski, W.L., 1993. Dream vs. reality: modern process control in the pulp and paper industry. *Control Systems*, Whistler 29 Sept.-1 Oct. 1992.

[8] Bonvin, D., 1998. Optimal operation of batch reactors - a personal view. *Journal of Process Control*, 8, 355-372.

[9] Bristol, E.H., 1990. Swinging door trending: adaptive trend rcording. *ISA National Conference Proceedings*, 1990, 749-753.

[10] Bronštein, I.N. and Semendjajev, K.A., 1977. *Taschenbuch der Mathematik für Ingenieure und Studenten.* Moskow: FIZMATLIT.

[11] Cao, L. and Mees, A., 2000. Deterministic structure in multichannel physiological data. *International Journal of Bifourcation and Chaos*, 10, 2767-2780.

[12] Casdagli, M., 1989. Nonlinear predition of chaotic time series. *Physica D*, 35, 335-354. Reprint in [87].

[13] Čenys, A., Lasiene, G. and Pyragas, K., 1991. Estimation of interrelation between chaotic observables. *Physica D*, 52, 332-337.

[14] Chiang, L.H., 2001. *Fault Detection and Diagnosis for Large-Scale Systems.* Ph.D. thesis, University of Illinois at Urbana-Champaign.

[15] Chiang, L.H., Russell, E.H. and Braatz, R.D., 2001. *Fault detection and diagnosis in industrial systems.* London: Springer.

[16] Chiang, L.H. and Braatz, R.D., 2003. Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemometrics and Intelligent Laboratory Systems*, 65, 159-178.

[17] Choi, S.W., Lee, C., Lee, J.-M., Park, J.H. and Lee, I.-B., 2005. Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems*, 75, 55-67.

[18] Choudhury, M.A.A.S., Shah, S.L. and Thornhill, N.F., 2003. Diagnosis of poor control loop performance using higher order statistics. *Automatica*, 40, 1719-1728.

[19] Choudhury, M.A.A.S, 2004. *Detection and Diagnosis of Control Loop Nonlinearities, Valve Stiction and Data Compression.* PhD Thesis, University of Alberta.

[20] Choudhury, M.A.A.S., Thornhill, N.F. and Shah, S.L., 2005. Modelling valve stiction. *Control Engineering Practice*, 13, 641-658.

[21] Desborough, L. and Miller, R., 2001. Increasing customer value of industrial control performance monitoring - Honeywell's experience. *Proceedings of the CPC6 Conference*, Tucson, AZ, 172-192.

[22] Desforges, M.J., Jacob, P.J. and Cooper, J.E., 1998. Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the Institution of Mechanical Engineers - Part C*, 212, 687-703.

[23] Downs, J.J. and Vogel, E.F., 1991. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 17, 245-255.

[24] Duda, R.O., Hart, P.E. and Stork, D.G., 2001. *Pattern Classification.* New York: Wiley.

[25] Ellis, G, 2004. *Control Systems Design Guide.* Oxford: Elsevier.

[26] Ender, D.B., 1993. Process control performance: not as good as you think. *Control Engineering Forum*, September Issues.

[27] Farmer, J.D. and Sidorowich, J., 1987. Predicting chaotic time series. *Physical Review Letters*, 59, 545-548. Reprint in [87].

[28] Feldmann, U. and Bhattacharya, J., 2004. Predictability improvement as an asymmetrical measure of interdependence in bivariate time series. *International Journal of Bifurcation and Chaos*, 14, 505-514.

[29] Forbes, M.G., Guay, M. and Forbes, J.F., 2003. Control design for first-order processes: shaping the probability density of the process state. *Journal of Process Control*, 14, 399-410.

[30] Frank, P.M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results. *Automatica*, 26, 459-474.

[31] Gautama, T., Mandic, D.P. and Van Hulle, M.M., 2003. Indications of nonlinear structures in brain electrical activity. *Physical Review E*, 67, 046204.

[32] Gautama, T., Mandic, D.P. and Van Hulle, M.M., 2003. A differential entropy based method for determining the optimal embedding parameters of a signal. *IEEE ICASSP* 2003, 29-31.

[33] Gautama, T., Mandic, D.P. and Van Hulle, M.M., 2004. The delay vector variance method for detecting determinism and nonlinearity in time series. *Physica D*, 190, 167-176.

[34] Gertler, J.J., 2000. *Structured parity equations in fault detection and isolation.* In [90].

[35] Girod, B., Rabenstein, R. and Stenger, A., 2001. *Signals and Systems.* Chichester: Wiley.

[36] Glass, L. and Mackey, M.C., 1979. Pathological conditions resulting from instabilities in physiological control systems. *Ann. N.Y. Academical Science*, 316, 214-235.

[37] Goodridge, C.L., Pecora, L.M., Carroll, T.L. and Rachford, F.J., 2001. Detecting functional relationships between simultaneous time series. *Physical Review E*, 64, 026221.

[38] Grassberger, P. and Procaccia, I., 1983. Characterization of strange attractors, *Physical Review Letters*, 50, 346-349.

[39] Gray, R.M., 1990. *Entropy and Information Theory.* London: Springer.

[40] Gregersen, L. and Jørgensen, S.B., 1999. Supervision of fed-batch fermentations. *Chemical Engineering Journal*, 75, 69-76.

[41] Hägglund, T., 1995. A control-loop performance monitor. *Control Engineering Practice*, 3, 1543-1551.

[42] Harris, T.J., 1989. Assessment of control loop performance. *The Canadian Journal of Chemical Engineering*, 67, 856-861.

[43] Harris, T.J., 2004. Statistical properties of quadratic-type performance indices. *Journal of Process Control*, 14, 899-914.

[44] Himmelblau, D.M., 1978. *Fault detection and diagnosis in chemical and petrochemical processes.* Amsterdam: Elsevier Press.

[45] Horch, A., 1999. A simple method for detection of stiction in process control loops. *Control Engineering Practice*, 7, 1221-1231.

[46] Horch, A., 2000. *Condition Monitoring of Control Loops.* PhD Thesis, Royal Institute of Technology, Stockholm, Sweden.

[47] Horch, A. and Isaksson, A.J., 2001. Detection of valve stiction in integrating processes. In: *European Control Conference*, Seminário de Vilar, Porto, Portugal, September 2001.

[48] Horch, A., Hegre, V., Hilmen, K., Melbo, H., Benabbas, L., Pistikopoulos, S. and Thornhill, N.F., 2005. Computer-aided plant auditing made possible by successful university cooperation. *ABB Review*, 2.

[49] Huang, B. and Shah, S.L., 1999. *Control Loop Performance Assessment: Theory and Applications.* London: Springer.

[50] Ignova, M., Montague, G.A., Ward, A.C. and Glassey, J., 1998. Fermentation seed quality analysis with self-organising neural networks. *Biotechnology and Bioengineering*, 64, 82-91.

[51] Ihara, S., 1993. *Information theory for continuous systmes.* Singapore: World Scientific.

[52] Iri, M.K, Aoki, E. O'Shima, E. and Matsuyama, H., 1979. An algorithm for diagnosis of system failures in the chemical process. *Computers and Chemical Engineering*, 3, 489-493.

[53] Isermann, R. and Ballé, P., 1997. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5, 709-719.

[54] Jamsä-Jounela, S.L., Vermasouri, M., Enden, P. and Haavisto, S., 2002. A process monitoring system based on the Kohonen self-organising maps. *Control Engineering Practice*, 11, 83-92.

[55] Kaiser, A. and Schreiber, T., 2002. Information transfer in continuous processes. *Physica D*, 166, 43-62.

[56] Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I. and Ohno, H., 2003. Monitoring independent components for fault detection. *AIChE Journal*, 49, 969-976.

[57] Kantz, H. and Schreiber, T., 1997. *Nonlinear Time Series Analysis.* Cambridge: Cambridge University Press.

[58] Kaplan, D.T. and Glass, L., 1992. Direct test for determinism in a time series. *Physical Review Letters*, 68, 427-430. Reprint in [87].

[59] Kaplan, D.T. and Glass, L., 1993. Coarse-grained embeddings of time series: random walks, Gaussian random processes, and deterministic chaos. *Physica D*, 64, 431-454.

[60] Kaplan, D.T., 1994. Exceptional events as evidence for determinism. *Physica D*, 73, 38-48.

[61] Karpenko, M. and Sepheri, N., 2002. Neural network classifiers applied to condition monitoring of a pneumatic process valve actuator. *Engineering Applications of Artificial Intelligence*, 15, 273-283.

[62] Kennel, M.B., Brown, R. and Abarbanel, H.D.I, 1992. Determining minimum embedding dimension using a geometrical construction. *Physical Review A*, 45, 3403-3411.

[63] Kocarev, L. and Parlitz, U., 1996. Generalized synchronization, predictability and equivalence of unidirectional coupled dynamical systems. *Physical Review Letter*, 76, 1816-1819.

[64] Kocarev, L., Parlitz, U. and Brown, R., 2000. Robust synchronization of chaotic systems. *Physical Review E*, 61, 3716-3720.

[65] Komulainen, T., Sourander, M. and Jamsä Jounela, S.L., 2004. An online application of dynamic PLS to a dearomatization process. *Computers and Chemical Engineering*, 28, 2611-2619.

[66] Kourti, T. and MacGregor, J.F., 1995. Process analysis, monitoring and diagnosis using multivariate projection methods - A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 28, 3-21.

[67] Kourti, T. and MacGregor, J.F., 1996. Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, 28, 409-428.

[68] Kourti, T., 2002. Process analysis and abnormal situation detection: from theory to practice. *IEEE Control Systems Magazine*, 22, 10-25.

[69] Kresta, J.V., MacGregor, J.F. and Marlin, T.E., 1991. Multivariable statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, 69, 35-47.

[70] Kullback, S., 1959. *Information Theory and Statistics.* New York: Wiley.

[71] Lenk, J.D., 1980. *Handbook of Controls and Instrumentation.* New Jersey: Prentice-Hall.

[72] Le Van Quyen, M., Dama, C., Baulac, M, Martenierie, J. and Varela, F.J., 1998. *Brain Research*, 792, 24-40.

[73] Le Van Quyen, M., Martinerie, J., Adam, C. and Varela, F.J., 1999. Nonlinear analysis of interictal EEG map the brain interdependencies in human focal epilepsy. *Physica D*, 127, 250-266.

[74] Li, R.F. and Wang, X.Z., 2002. Dimension reduction of process dynamic trends using independent component analysis. *Computers and Chemical Engineering*, 26, 467-473.

[75] Lipták, B.G., 2000. *Instrument Engineers Handbook, Volume 1 (Process Measurement)*. Philadelphia: Chilton Books.

[76] Luyben, W.L. and Luyben, M.L., 1997. *Essentials of process control.* Singapore: McGraw Hill.

[77] MacGregor, J.F. and Kourti, T., 1995. Statistical process control of multivariate processes. *Control Engineering Practice*, 3, 403-414.

[78] Marschinski, R. and Kantz, H., 2002. Analysing the information flow between financial time series. *The European Physical Journal B*, 30, 275-281.

[79] Martin, E.B., Morris, J. and Lane, S., 2002. Monitoring process manufacturing performance. *IEEE Control Systems Magazine*, Oct, 26-39.

[80] Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V., 2003. A sytematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Industrial Engineering Chemistry Research*, 42, 4789-4810.

[81] Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V., 2003. A sytematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. *Industrial Engineering Chemistry Research*, 42, 4811-4827.

[82] Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V., 2004. Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Engineering Applications of Artificial Intelligence*, 17, 501-518.

[83] Mizuta, H. Jibu, M. and Yana, K., 2000. Adaptive estimation of the degree of system nonlinearity. *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 1-4 October 2000, Lake Louise.

[84] Montague, G.A., Gent, C., Morris, A.J., Buttress, J., 1996. Industrial reactor modelling with artificial neural networks. *Transactions of the Institute of Measuremnt and Control*, 18, 118-124.

[85] Nam, D.S., Han, C., Jeong, C.W. and Yoon, E.S., 1996. Automatic construction of extended sympton-fault associations from signed digraph. *Computers and Chemical Engineering*, 20, S605-S610.

[86] Netoff, T.I., Pecora, L.M. and Schiff, S.J., 2004. Analytical coupling detection in the presence of noise and nonlinearity. *Physical Review E*, 69, 017201.

[87] Ott, E. Sauer, T. and Yorke, J.A. (Eds.), 1994. *Coping with Chaos - Analysis of Chaotic Data and the Exploitation of Chaotic Systems.* New York: Wiley Series.

[88] Page, E.W., 1954. Continuous inspection schemes. *Biometrika*, 41, 100-114.

[89] Palus, M. and Stefanovska, A., 2003. Direction of coupling from phases of interacting oscillators: an information theoretic approach. *Physical Review E*, 67, 055201.

[90] Patton, R.J., Frank, P.M. and Clark, R.N., 2000. *Issues of Fault Diagnosis for Dynamic Systems.* London: Springer.

[91] Patton, R.J., 2005. A benchmark study approach to fault diagnosis of industrial process control systems. *IEE Seminar on Control Loop Assessment and Diagnosis*, 16 June, London, UK.

[92] Paulonis, M. A. and Cox, J.W., 2003. A practical approach for large-scale controller performance assessment, diagnosis and improvement. *Journal of Process Control*, 13, 155-168.

[93] Pecora, L.M., Carroll, T.L. and Heagy, J.F., 1995. Statistics for mathematical properties of maps between time series embeddings. *Physical Review E*, 53, 3420-3436.

[94] Pettersson, J. and Gutman, P.-O., 2003. Automatic tuning of the window size in the Box Car Back Slope data compression algorithm. *Journal of Process Control*, 14, 431-329.

[95] Pinder, A.C. and Godfrey, G., 1993. *Food process monitoring.* Glasgow: Chapman & Hall.

[96] Piovoso, M.J. and Hoo, K.A., 2002. Mutivariate statistics for process control. *IEEE Control Systems Magazine*, 22, 8-9.

[97] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 2002. *Numerical Recipes.* Cambridge: Cambridge University Press.

[98] Qin, S.J., 1998. Control performance monitoring - a review and assessment. *Computers and Chemical Engineering*, 23, 173-186.

[99] Qin, S.J., Valle, S. and Piovoso, M.J., 2001. On unifying multiblock analysis with application to decentralized process monitoring. *Journal of Chemometrics*, 15, 715-742.

[100] Quian Quiroga, R., Arnhold, J. and Grassberger, P., 2000. Learning driver-response relationships from synchronization patterns. *Physical Review E*, 61, 5142-5148.

[101] Raich, A.C. and Çinar, A., 1995. Multivariate statistical methods for monitoring continuous processes: assessment of discrimination power of disturbance models and diagnosis of multiple disturbances. *Chemometrics and Intelligent Laboratory Systems*, 30, 37-48.

[102] Rengaswamy, R., Hägglund, T., Venkatasubramanian, V., 2001. A qualitative shape analysis formalism for monitoring control loop performance. *Engineering Applications of Artificial Intelligence*, 14, 23-33.

[103] Rulkov, N.F., Shuschik, M.M., Tsimring, L.S. and Arbandel, H.D.I, 1995. Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, 51, 980-994.

[104] Runger, G.C. and Testik, M.C., 2004. Multivariate extensions to cumulative sum control charts. *Quality and Reliability Engineering International*, 20, 587-606.

[105] Russell, E.L., Chiang, L.H. and Braatz, R.D., 2000. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 51, 81-93.

[106] Salvino, L.W., Cawley, R., Grebogi, C. and Yorke, J.A., 1995. Predictability in time series. *Physics Letters A*, 209, 327-332.

[107] Sauer, T., Yorke, J.A. and Casdagli, M., 1991. Embedology. *Journal of Statistical Physics*, 65, 579-616.

[108] Sauer, T., 1994. Time series prediction using delay coordinate embedding. In [147].

[109] Schaper, C.D., Larimore, W.E., Seborg, D.E. and Mellichamp, D.A., 1994. Identification of chemical processes using canonical variate analysis. *Computers and Chemical Engineering*, 18, 55-69.

[110] Schiff, S.J., So, P., Chang, T., Burke, R.E. and Sauer, T., 1996. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physical Review E*, 54, 6708-6724.

[111] Schmitz, A., 2000. Measuring statistical dependence and coupling of subsystems. *Physical Review E*, 62, 7508-7511.

[112] Schreiber, T., 1999. Interdisciplinary application of nonlinear time series methods. *Physics Report - Review Section of Physics Letters*, 308, 2-64.

[113] Schreiber, T., 2000. Measuring information transfer. *Physical Review Letters*, 85, 461-464.

[114] Schreiber, T. and Schmitz, A., 2000. Surrogate time series. *Physica D*, 142, 346-382.

[115] Scott, D.W., 1992. *Multivariate Density Estimation: Theory, Practice and Visualization.* New York: Wiley.

[116] Seborg, D.E., Edgar, T.F. and Mellichamp, D.A., 2004. *Process Dynamics and Control.* 2nd Ed., New Jersey: Wiley.

[117] Shannon, C.E., 1948. A mathematical theory of communication. *Bell Systems Technology Journal*, 27, 379-423 (Reprint).

[118] Shunta, J.P., 1995. *Achieving World Class Manufacturing through Process Control.* New Jersey: Prentice Hall.

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

[120] Simoglou, A., Martin, E.B. and Morris, A.J., 2002. Statistical performance monitoring of dynamic multivariate processes using state space modelling. *Computers and Chemical Engineering*, 26, 909-920.

[121] Singhal, A. and Seborg, D.E., 2005. Effect of data compression on pattern matching in historical data. *Industrial Engineering and Chemistry Research*, 44, 3203-3212.

[122] Slack, N., Chamber, S., Harland, C., Harrison, A., and Johnston, R., 1998. *Operations Management.* 2nd ed. London, Washington D.C.: Pitman Publishing.

[123] Sorsa, T. and Koivo, H.N., 1991. Neural networks in process fault diagnosis. *IEEE Transactions on Systems, Man and Cybernetics*, 21, 815-825.

[124] Srivastava, M.S., 2002. *Methods of Multivariate Statistics.* New York: Wiley.

[125] Stam, C.J., Pihn, J.P.M. and Pritchard, W.S., 1998. Reliable detection of nonlinearity in experimental time series with strong periodic components. *Physica D*, 112, 361-380.

[126] Stephanopoulos, G., 1984. *Chemical Process Control: An Introduction to Theory and Practice.* Edgwood Cliffs: Prentice-Hall.

[127] Sugihara, G. and May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series. *Nature*, 344, 734-741. Reprint in [87].

[128] Takens, F., 1981. Detecting strange attractors in turbulance. *Lecture Notes in Mathematics*, No 898, Springer Verlag.

[129] Tan, C.C., Thornhill, N.F. and Belchamber, R.M., 2002. Principal component analysis of spectra with application to acoustic emissions from mechanical equipment. *Transactions of the Institute of Measurement and Control*, 24, 333-353.

[130] Theiler, J.S., Eubank, A., Longtin, B., Galdrikian, B. and Farmer, J.D., 1992. Testing for nonlinearity in time series - the method of surrogate data. *Physica D*, 58, 77-94.

[131] Thomas, M.S., Kumar, P. and Chandna, V.K., 2004. Design, development and commissioning of a supervisory control and data acquision (SCADA) laboratory for research and training. *IEEE Transactions on Power Systems*, 19, 1582-1588.

[132] Thornhill, N.F. and Hägglund, T., 1997. Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5, 1343-1354.

[133] Thornhill, N.F., Shah, S.L. and Huang, B., 2001. Detection of distributed oscillations and root cause diagnosis. *Preprints CHEMFAS4*, Korea, 7-8 June, 167-172.

[134] Thornhill, N.F., Shah, S.L., Huang, B. and Vishnubhotla, A., 2002. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 10, 833-856.

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

[136] Thornhill, N.F., Huang, B. and Zhang, H., 2003. Detection of multiple oscillations in control loops. *Journal of Process Control*, 13, 91-100.

[137] Thornhill, N.F., 2004. Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control Systems Technology*, 13, 434-443.

[138] Thornhill, N.F., Choudhur, M.A.A.S. and Shah, S.L., 2004. The impact of compression on data-driven process analysis. *Journal of Process Control*, 14, 389-398.

[139] Tse, P.W., Peng, Y.H. and Yam, R., 2001. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis - their effectiveness and flexibilities. *Journal of Vibration and Acoustics - Transactions of the ASME*, 123, 303-310.

[140] Ulbikas, J., Čenys, A. and Sulmova, O.P., 1998. Chaos parameters for EEG analysis. *Nonlinear Analysis: Modeling and Control*, 3, Vilinius, IMI, 1-8.

[141] Venkatasubramanian, V., Rengaswamy, R., Yin K. and Kavuri, S.N., 2003. A review of process fault detection and diagnosis. Part I: Quantitative model-based methods. *Computers and Chemical Engineering*, 56, 293-311.

[142] Venkatasubramanian, V., Rengaswamy, R., Yin K. and Kavuri, S.N., 2003. A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies. *Computers and Chemical Engineering*, 56, 313-326.

[143] Venkatasubramanian, V., Rengaswamy, R., Yin K. and Kavuri, S.N., 2003. A review of process fault detection and diagnosis. Part III: Process history based methods. *Computers and Chemical Engineering*, 56, 327-346.

[144] Wachs, A. and Lewin, D.R., 1999. Improved PCA methods for disturbance and failure identification. *AIChE Journal*, 45, 1688-1700.

[145] Wang, H. and Lin, W., 2000. Applying observer based FDI techniques to detect faults in dynamic and bounded stochastic distribution. *International Journal of Control*, 73, 1424-1436.

[146] Watson, M.J., Liakopoulos, A., Brzakovic, D. and Georgakis, C., 1998. A practical assessment of process data compression techniques. *Industrial Engineering and Chemistry Research*, 37, 267-274.

[147] Weigend, A.S. and Gersenfeld, N.A., 1994. Time Series Prediction: Forecasting the Future and Understanding the Past. *Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis held in Santa Fe, New Mexico, May 14-17 1992.* Reading, MA: Addison Wesley.

[148] Westerhuis, J.A., Kourti, T. and MacGregor, J.F., 1998. Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, 12, 301-321.

[149] Wiener, N., 1948. *Cybernetics.* New York: Wiley.

[150] Wiesenfeldt, M., Parlitz, U. and Lauterborn, W., 2001. Mixed state analysis of multivariable time series. *International Journal of Bifurcation and Chaos*, 11, 2217-2226.

[151] Wise, B.M. and Gallagher, N.B., 1996. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6, 329-248.

[152] Xia, C. and Howell, J., 2002. Loop status monitoring and fault localisation. *Journal of Process Control*, 13, 679-691.

[153] Xia, C. and Howell, J., 2003. Isolating multiple sources of plant-wide oscillations via spectral independent component analysis. *Control Engineering Practice*, 13, 1027-1035.

[154] Yoon, S. and MacGregor, J.F., 2004. Principal component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal*, 50, 2891-2903.

[155] Zang, X. and Howell, J., 2004. Correlation dimension and Ljapunov exponent based isolation of plant-wide oscillations. *Proceedings of DYCOPS7*, 5-7 July, Boston, MA.

[156] Zhang, J.Q. and Yan, Y., 2001. A wavelet-based approach to abrupt fault detection and diagnosis of sensors. *IEEE Transactions on Instrumentation and Measurement*, 50, 1389-1396.

[157] Zupan, J. and Gasteiger, J., 1991. Neural networks - a new method for solving chemical problems or just a passing phase. *Analytica Chimica Acta*, 248, 1-30.

[158] Åström, K.J., 1967. Computer control of a paper machine - an application of linear stochastic control theory. *IBM Journal for Research and Development*, 11, 389-405.

[159] Åström, K.J. and Hägglund, T., 1995. *PID Controllers: Theory, Design, and Tuning.* Research Triangle Park, NC: ISA.