

Oscillation Detection and Causality Analysis of Control Systems

by

Elham Naghoosi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in  
Process Control

Department of Chemical and Materials Engineering  
University of Alberta

© Elham Naghoosi, 2016

# Abstract

Management of abnormal events in chemical processes requires detection and diagnosis of abnormal performance of individual elements of the system. Detection of abnormal performance is usually done by means of setting a control limit on measured variables. Abnormality due to any reason in one element, may propagate through feedback loops and interconnections through the process, downgrading the performance of the whole operation. Receiving several alarms due to one abnormality is quite common in industrial operations making it a challenging task to diagnose the root cause of the problem within a limited time.

Causality graphs demonstrating how the measured variables relate to each other help the operators and engineers to short list the main faulty variables which propagate the abnormality to the rest of the process. Causality graphs can be obtained based on deep process knowledge, process schematics or historical data. Qualitative causality graphs based on process knowledge or schematics, even though the most valuable ones, have some limitations and therefore, causality analysis based on historical data has gained a lot of attention.

A part contribution of the thesis is to advance the causality analysis procedures proposing methodologies to extract more reliable information about cause and effect relations between recorded variables. This part of the study considers both causality analysis assuming linear relations between variables as well as nonlinear ones. Considering linear methodologies, a more appropriate model structure and parameter estimation methodology than the existing ones is proposed based on Bayesian framework. Estimating model parameters under Bayesian framework accompanied with a carefully designed prior probability for the parameters can solve some of the issues of traditional procedures. Also, a new procedure is proposed to decompose the energy transfer between variables in a way to obtain a complete picture on the

different paths along which variables can influence each other in addition to providing an estimation of the energy transferred through each path independently. Regarding causality analysis with the assumption of nonlinear relations between variables, a more advanced methodology based on information transfer concepts such as mutual information and transfer entropy is proposed in order to more reliably detect the true relations between the variables.

The work toward diagnosis of abnormalities leads this study toward developing more reliable algorithms to specifically detect and characterize oscillations in control loops. A methodology is developed to detect and estimate oscillation frequencies which could be otherwise hidden within noise and non-stationary trends in industrial variables. Oscillations may occur in control loops due to aggressive controller tuning, external disturbances or due to nonlinearity of the valve or the process itself. In order to help the root cause diagnosis of propagated oscillations, a novel method is developed to distinguish between these three types of oscillations. Oscillations caused by each one of these three sources have specific characteristics distinguishable from other types of oscillations. Examining the oscillation characteristics in the wavelet domain made it possible to develop a comprehensive methodology which is capable of detection and independent diagnosis of different oscillatory components of the variables. The proposed methodologies are verified through various simulations, laboratory experiments and industrial case studies.

# Preface

Chapters 2,3,5 and 6 of this thesis have been previously published as the following journal articles.

1. E. Naghooosi, B. Huang, Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations, *Industrial & Engineering Chemistry Research*, 53.22 (2014): 9427-9438.
2. E. Naghooosi, B. Huang, Diagnosis of oscillations between controller tuning and harmonic external disturbances, *IEEE on Control Systems Technology*, 23.4 (2015): 1283-1293.
3. E. Naghooosi, B. Huang, Interaction analysis of multivariate control systems under Bayesian framework, accepted for publication in *IEEE on Control Systems Technology*, (2016).
4. E. Naghooosi, B. Huang, E. Domlan, R. Kadali, Information transfer methods in causality analysis of process variables with an industrial application, *Journal of Process Control*, 23.9 (2013): 1296-1305.

I was responsible for the theoretical study, algorithm development, data analysis and manuscript composition of all the articles. Dr. Biao Huang was the supervisory author and was involved with manuscript composition. Dr. Elom Domlan and Ramesh Kadali were also supervisory authors of the fourth paper.

# Acknowledgements

I would like to take this opportunity to express my gratitude toward the people who have made this journey possible for me. First and foremost, my sincere thanks go to my supervisor professor Biao Huang for his exceptional care and patience toward students specially during the non-productive times. I have been given the the opportunity to explore my way through PhD under his guidance. The opportunity to work on several industrial projects alongside the research has been quite valuable. Also, the work in the research laboratory provided a unique experience for me which was possible through working in Industrial Research Chair group under supervision of Prof. Biao Huang.

I am also thankful toward my defense committee members Prof. Tongwen Chen, Dr. Vinay Prasad, Prof. Sirish Shah, Dr. Mahdi Tavakoli and Prof. Andreas Kugi for their valuable suggestions. I have greatly benefited from discussions and meetings with members of industrial partners Syncrude Canada Ltd. and Suncor Energy Inc.. I would like to specially appreciate Aris Espejo for his guidance, encouragements and feedbacks regarding my work. I also appreciate the discussions with Dr. Jim Kresta and Darcy Daugela from Syncrude and Dr. Elom Domlan, Dr. Ramensh Kadali and Dr. Enbo Feng from Suncor Energy.

I have enjoyed the friendship of many valuable colleagues. I specially appreciate Dr. Shima Khatibisepehr, Dr. Aditya Tulsyan, Dr. Ruben Gonzalez, Dr. Da Zheng, Dr. Swanand Khare, Dr. Yuri Shardt, Dr. Nima Sammaknejad, Dr. Fadi Ibrahim, Mohammad Rashedi and Anahita Sadeghian.

Last but not least, I would like to appreciate my husband, Dr. Mehdi Dastfan, for all the time being there for me through this journey. Without his encouragements this thesis would not have been accomplished.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction . . . . .	2
1.1.1	Thesis contributions . . . . .	4
1.1.2	Thesis outline . . . . .	5
<b>2</b>	<b>Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations</b>	<b>8</b>
2.1	Abstract . . . . .	9
2.2	Introduction . . . . .	9
2.3	Oscillation detection based on auto correlation function . . . . .	11
2.4	The proposed method . . . . .	13
2.5	Peak detection algorithm . . . . .	15
2.6	Specifying the threshold for the peak detection algorithm . . . . .	18
2.7	Clustering the identified peaks . . . . .	21
2.7.1	Specifying the threshold for the peak clustering algorithm . . .	23
2.8	Addressing possible problem . . . . .	25
2.9	Dampened oscillatory signals . . . . .	27
2.10	Case study . . . . .	29
2.11	Non-sinusoidal signals . . . . .	33
2.12	Summary . . . . .	34
<b>3</b>	<b>Differentiating between oscillations due to controller tuning and harmonic external disturbances</b>	<b>37</b>
3.1	Abstract . . . . .	38
3.2	Introduction . . . . .	38
3.3	Response of a general ARMA process . . . . .	40
3.3.1	The response of ARMA process in reality . . . . .	42
3.3.2	ACF of ARMA processes . . . . .	43
3.3.3	Continuous time systems . . . . .	45

3.4	ACF of harmonics . . . . .	46
3.5	Diagnosis . . . . .	47
3.5.1	Automatic method for diagnosis of oscillation . . . . .	49
3.5.2	Effect of added colored noise . . . . .	51
3.5.3	Estimation of signal to noise ratio as additional test . . . . .	53
3.6	Pure imaginary poles . . . . .	56
3.7	Case study . . . . .	59
3.8	Summary . . . . .	61
<b>4</b>	<b>Wavelet transform based methodology for detection and diagnosis of multiple oscillations in non-stationary variables</b>	<b>64</b>
4.1	Abstract . . . . .	65
4.2	Introduction . . . . .	65
4.3	Introduction to Wavelet transform . . . . .	67
4.3.1	Interpreting the wavelet coefficients . . . . .	70
4.3.2	Data pre-processing . . . . .	73
4.4	Wavelet power spectrum and its application for oscillation diagnosis .	79
4.4.1	Diagnosis of oscillations between controller tuning and external disturbance based on wavelet power spectrum . . . . .	81
4.5	Wavelet coherency . . . . .	84
4.5.1	Bicoherence . . . . .	86
4.5.2	Wavelet bicoherence . . . . .	88
4.6	Diagnosis of oscillations due to nonlinearity . . . . .	91
4.6.1	A note regarding the selection of the scales for which the bicoherence should be estimated . . . . .	94
4.7	Case study . . . . .	96
4.8	Summary . . . . .	97
<b>5</b>	<b>Interaction analysis of multivariate control systems under Bayesian framework</b>	<b>107</b>
5.1	Abstract . . . . .	108
5.2	Introduction . . . . .	108
5.2.1	Literature review on causality analysis methodologies with the assumption of linearity . . . . .	112
5.3	Model structure . . . . .	115
5.4	Detection of instantaneous feedback between variables . . . . .	118
5.5	Bayesian estimation of SVAR models . . . . .	121

5.5.1	Estimation of SVAR model parameters in fully recursive form	122
5.5.2	Solving the equations simultaneously . . . . .	125
5.5.3	Imposing exact restrictions on the posterior distribution . . . . .	128
5.6	Design of the prior distribution . . . . .	130
5.6.1	Prior on elements of $A_0$ . . . . .	130
5.6.2	Prior on elements of $a_+$ . . . . .	131
5.7	Impulse response estimation . . . . .	132
5.7.1	Confidence interval for the impulse responses . . . . .	133
5.8	Impulse response decomposition . . . . .	134
5.9	Case study . . . . .	139
5.10	Summary . . . . .	141
<b>6</b>	<b>Information transfer methods in causality analysis of process variables</b>	<b>145</b>
6.1	Abstract . . . . .	146
6.2	Introduction . . . . .	146
6.3	Mutual information and transfer entropy . . . . .	147
6.4	Practical issues and solutions . . . . .	153
6.4.1	Computation method . . . . .	153
6.4.2	Confidence limit . . . . .	154
6.4.3	Computational load . . . . .	155
6.4.4	Application of dependency and differential dependency curves	155
6.4.5	Parametrising the transfer entropy . . . . .	158
6.4.6	Case of two variables with common parent . . . . .	161
6.5	Industrial application . . . . .	164
6.5.1	Summary of the algorithm for causality analysis . . . . .	171
6.6	Summary . . . . .	172
<b>7</b>	<b>Concluding remarks</b>	<b>173</b>
7.1	Concluding remarks . . . . .	174
7.2	Future work . . . . .	175
<b>Bibliography</b>		<b>177</b>
.1	Appendix . . . . .	185
.1.1	Examining correlation between residuals of a multivariate regression model estimated based on Bayesian framework . . . . .	185

# List of Figures

2.1	Top: $x_2(t)$ in time. Bottom: ACF of $x_2(t)$ . . . . .	13
2.2	Illustration of the general idea . . . . .	14
2.3	Illustration of the problem caused by noise . . . . .	14
2.4	9 points of ACF of a variable along with the order that they are condiered for clustering . . . . .	16
2.5	Illustration of the effect of clustering threshold on detection of peaks	16
2.6	The two cluster centers corresponding to the ACF in Figure 2.5 . .	17
2.7	Illustration of the effect of the decay in estimated ACF . . . . .	21
2.8	Top: $x_3(t)$ in time. Bottom: ACF of $x_3(t)$ . . . . .	22
2.9	Illustration of peak clustering algorithm . . . . .	23
2.10	$x_4(t)$ in time and its ACF after downsampling . . . . .	27
2.11	Top: $x_5(t)$ in time, Bottom: ACF of $x_5(t)$ . . . . .	29
2.12	Time series plot of the case study variables . . . . .	30
2.13	ACF and scaled spectrum of the case study vaiables . . . . .	31
2.14	Components of a variable containing non-sinusoidal oscillations . . .	34
2.15	Algorithm for oscillation detection . . . . .	36
3.1	$y_1(t)$ in time, its ACF and response to non-zero initial values . . . .	44
3.2	Top panel: step response, Middle panel: response to white noise, Bot- tom panel: ACF of the response to white noise of a contiuous time linear sytem controlled with a PID controller . . . . .	45
3.3	Top panel: response of the simulated continuous time system to oscil- latory disturbance, Bottom panel: ACF of the response . . . . .	46
3.4	$y_2(t)$ and $y_3(t)$ in time along with their ACFs . . . . .	48
3.5	ACF of $y_4(t)$ . . . . .	49
3.6	Distribution of the differences between the peak values in the ACF of $x_t = \sin(\frac{2\pi}{10}t) + \varepsilon_t$ . . . . .	51

3.7	Top: output of the simulated continuous time loop, Middle: ACF of the output, Bottom: histogram of the difference between the peaks in the ACF . . . . .	53
3.8	Top panel: step response of the simulated loop, Middle panel: ACF of the step response, Bottom panel: peaks of the ACF of the response and of the equivalent harmonic . . . . .	56
3.9	Top panel: response to noise of the simulated loop, Middle panel: ACF of the response, Bottom panel: peaks of the ACF of the response and of the equivalent harmonic . . . . .	57
3.10	Step response of the continuous control loop introduced in Section 3.3.3 under noisy environment when the gain of the controller is 2.6 . . . . .	58
3.11	Schematic of the multi-tank system under study . . . . .	59
3.12	Level measurement of the tank under study and its ACF when the loop is oscillatory due to the controller . . . . .	60
3.13	Prediction residuals of the ACF plotted in Figure 3.12 and its significance level . . . . .	61
3.14	Level measurement of the tank under study and its ACF using a stable controller in the loop with added oscillatory disturbance . . . . .	61
3.15	Prediction residuals of the ACF plotted in Figure 3.14 along with the significance level . . . . .	62
3.16	The normalized output of an oscillatory loop in an industrial process and its ACF . . . . .	62
3.17	Prediction residuals of the ACF plotted in Figure 3.16 . . . . .	63
3.18	The output of a loop suffering from valve stiction . . . . .	63
4.1	Morlet wavelet function with central frequency of 0.8125 Hz . . . . .	69
4.2	Top: $x(t)$ . Middle: Wavelet coefficients at a scale approximately corresponding to oscillation period of 20. Bottom: Wavelet coefficients at a scale approximately corresponding to oscillation period of 35 . . . . .	71
4.3	Wavelet coefficients of $x(t)$ for scales from 1 to 50 . . . . .	72
4.4	Top: $x(t) = u(t - 250)$ . Bottom: Wavelet coefficients of $x(t)$ . . . . .	73
4.5	Time domain plot of variable loop1 . . . . .	76
4.6	ACF of the Loop1 before and after pre-processing . . . . .	77
4.7	Wavelet coefficients of Loop1 . . . . .	78
4.8	Average absolute wavelet coefficients of Loop1 . . . . .	79
4.9	Wavelet coefficients of Loop1 at scale 130 . . . . .	79
4.10	ACF of Loop1 after removing the time samples from 2600 to 4000 . . . . .	80

4.11	$y(t) = \frac{1}{1-0.33z^{-1}+0.9z^{-2}}\varepsilon_t + 2\sin(2\pi/25t)$ in time along with its ACF and continuous wavelet transform of its ACF . . . . .	82
4.12	Wavelet coefficients of two scales of the transform of the ACF of $y(t) = \frac{1}{1-0.33z^{-1}+0.9z^{-2}}\varepsilon_t + 2\sin(2\pi/25t)$ . . . . .	83
4.13	Ratio of the calculated wavelet power spectrum to the predicted ones at the peaks corresponding to the harmonic oscillation . . . . .	85
4.14	Ratio of the calculated wavelet power spectrum to the predicted ones at the peaks corresponding to the oscillation generated by the linear system . . . . .	85
4.15	Bicoherency of $x(t) = \frac{z^{-2} + 0.3z^{-3}}{z^2 - 0.8z + 0.25}\varepsilon_t$ in frequency domain . . . . .	87
4.16	Total bicoherence of $x(t)$ in frequency domain . . . . .	90
4.17	Total bicoherence of $x(t)$ versus scales . . . . .	91
4.18	Total bicoherence of $x(t)$ in frequency domain . . . . .	92
4.19	Bicoherence of signal $y(t)$ between scales 32 and 20 in time . . . . .	93
4.20	Total bicoherence of signal $y(t)$ . . . . .	94
4.21	Bicoherence of signal $y(t)$ based on Fourier method . . . . .	95
4.22	Required steps for oscillation diagnosis . . . . .	98
4.23	Bicoherence of signal $x(t)$ for scales up to 80 . . . . .	99
4.24	Bicoherence of signal $x(t) = \sin(2\pi \times 0.04 + \frac{\pi}{3}) + \sin(2\pi \times 0.025 + \frac{\pi}{12})$ for scales up to 80 . . . . .	99
4.25	Wavelet coefficients of signal $x(t) = 2\sin(2\pi \times 0.04 + \frac{\pi}{3}) + 2\sin(2\pi \times 0.025 + \frac{\pi}{12}) + 2\sin(2\pi \times 0.065 - \frac{\pi}{4}) + \varepsilon(t)$ . . . . .	100
4.26	4 industrial variables along with the absolute wavelet coefficients . . . . .	101
4.27	ACFs of the 4 industrial variables . . . . .	102
4.28	Wavelet bicoherence of down-sampled Loop 1 by factor of 2 . . . . .	103
4.29	Wavelet bicoherence of down-sampled Loop 2 by factor of 2 . . . . .	103
4.30	Wavelet bicoherence of down-sampled Loop 3 by factor of 2 . . . . .	104
4.31	Wavelet bicoherence of down-sampled Loop 4 by factor of 2 . . . . .	104
4.32	Top: Controller output of Loop 30 in time. Bottom: Wavelet coefficients . . . . .	105
4.33	Wavelet bicoherence of down-sampled Loop 30 by factor of 2 . . . . .	105
4.34	Average of absolute wavelet coefficients of controller output of Loop 30 . . . . .	106
5.1	Simulated process . . . . .	135
5.2	Removing the indirect connections from $u_1$ to $y_2$ . . . . .	137
5.3	The connection between $u_1$ and $y_1$ through $u_3$ . . . . .	137
5.4	Simulated process . . . . .	138
5.5	Total impulse response from $u_2$ to $y_1$ along with 90% confidence interval . . . . .	139

5.6	Direct impulse response from $u_2$ to $y_1$ along with 90% confidence intervals	140
5.7	Impulse response from $u_2$ to $y_1$ along with 90% confidence intervals while only the first loop is closed . . . . .	140
5.8	Impulse response from $u_2$ to $y_1$ while both the loops are closed . . . . .	141
5.9	Impulse response from $u_2$ to $y_1$ through $y_3$ . . . . .	142
5.10	Schematic of the multi-tank system under study . . . . .	142
5.11	Level measurement of the three tanks under study . . . . .	143
5.12	Impulse responses to the middle tank . . . . .	143
5.13	Impulse responses to the bottom tank . . . . .	144
6.1	Time lagged mutual information between a variable by itself . . . . .	149
6.2	Time lagged dependency between input and output of an industrial process . . . . .	150
6.3	Example of two variables that are linearly corelated to each other . .	150
6.4	Time lagged dependency between the second variable with the history of the first variable in Figure 6.3 . . . . .	151
6.5	Time lagged dependency between the second variable with its own history	151
6.6	Time lagged dependency between first and second variables in Figure 6.3 . . . . .	156
6.7	Input and output data of the summation model . . . . .	157
6.8	Dependency and differential dependency between data in Figure 6.7 .	158
6.9	An example of two variables with a common source . . . . .	162
6.10	A contiuously stirred tank . . . . .	162
6.11	Causality graph for the example in Figure 6.9 . . . . .	163
6.12	Separation vessel schematic . . . . .	164
6.13	Dependency between feed flow and feed density . . . . .	165
6.14	Dependency between feed density and underflow density . . . . .	166
6.15	Time lagged dependency between underflow density and middlings density . . . . .	167
6.16	Time lagged dependency between feed density and underflow pressure	168
6.17	Time lagged dependency between feed flow and underflow pressure .	169
6.18	Time lagged dependency between cone flush flow and underflow density	169
6.19	Causality relationships in the separation process . . . . .	170
6.20	Dependency between three tags . . . . .	171

# List of Tables

2.1	Result of applying the algorithm to $x_2(t)$ . . . . .	25
2.2	Result of applying the algorithm to $x_4(t)$ . . . . .	27
2.3	Result of analyzing the industrial data . . . . .	33

# Chapter 1

## Introduction

## 1.1 Introduction

The operational efficiency of chemical processes highly depends on the performance of their control systems. It is reported that the performance of almost 60% of industrial controllers is not close to optimum [1] due to various reasons such as poor controller tuning, hardware malfunctioning, external disturbances, tight limits on the manipulated variables, nonlinearity etc. Estimated loss of 20 billion dollars in petrochemical industries [2] due to abnormal events, has caused the interest of industries and academics to develop more reliable automatic abnormal event management strategies. Abnormal event management includes detection of process performance abnormalities, diagnosis of problems and decision on the required maintenance to return the process back to normal operation [3].

The main step toward an automatic abnormal event management is development of an automatic detection and diagnosis algorithm. Automatic detection of abnormal control performance (fault) is viable through implementation of a performance indicator as is proposed in literature [1, 4]. However, diagnosing the root cause of a propagated fault in a multivariate process is yet a challenging issue. There are two main challenges regarding diagnosis. First, most of the developed diagnostic algorithms are specialized in finding the source of a specific type of abnormality. For example, there are several methods specialized in the diagnosis of an oscillation caused by a sticky valve [5, 6, 7, 8] which however cannot work if the oscillation is due to controller tuning or else. It is mentioned in literature that relying on one type of plant-wide diagnosis procedure assuming a specific type of fault is not only unreliable, but also risky [9]. The main reason is the similarity between the symptoms of different abnormalities and the focus of most of the diagnostic algorithms in literature is on finding the root cause of a specific kind of abnormality.

The second challenge toward diagnosis of abnormal performance in multivariate systems is due to the interconnection between the many loops of the process [10]. A fault can occur in one of the thousands of components in a typical chemical process and easily affect many loops or even the product quality. The interconnections between the loops make it difficult to diagnose the source of abnormality among many variables which have similar symptoms [11].

A common method for diagnosis of abnormal performance in multivariate processes is to utilize process connectivity charts in the form of signed directed graphs (SDG) [12], fault trees [13], Bayesian nets [14] or causal graphs. SDGs are derived based on the process knowledge in terms of first principle modeling, expert knowledge and also P&IDs in order to provide a causal map for the process variables [15]. Fault

trees are also utilized in risk and reliability assessment and provide information on the connection between low level abnormalities and top layer process hazards [16]. Both SDG and fault tree are based on qualitative analysis or deep knowledge about the underlying process. On the other hand, Bayesian nets or data based causal graphs are based on analyzing the historical operational data. Both types of methodologies have shortcomings and limitations. Verification of qualitative causality graphs with process data analysis is considered as an improvement [17].

One shortcoming of SDGs and fault trees is that they are highly dependent on the knowledge of the personnel, which may fail in the case of a new performance abnormality or operation state. Also, the mathematical equations underlying the process are not always available and do not contain much information about causality. Therefore, development of reliable methodologies to analyze the interaction between the variables based on historical data is a promising alternative. [9] has noted interaction analysis as the second step toward a comprehensive control performance assessment, diagnosis and prioritization.

A reliable causality or interaction analysis procedure helps in narrowing down to the top layer variables which are propagating the fault to the bottom layer variables while many variables have similar symptoms. It can also be utilized to differentiate the variation generated within a loop from variations entering the loop as external disturbance. Therefore, it has the potential of diagnosing the source of variations felt within a loop.

Causality analysis based on historical data has been studied in literature in many different fields. Granger [18] was the first to bring the notion of causality to economics proposing a method to distinguish between causality from simple correlation. Since then, causality analysis is considered in various fields from biological studies to engineering with the common purpose of learning how the variables of a complex system truly influence each other and to detect the underlying mechanism of data generation process. A literature review on causality analysis methodologies will be provided in the corresponding chapter of this thesis. Causality analysis is yet an open issue for further study and a part of the study considers advancing the existing procedures.

An issue regarding abnormal performance detection and diagnosis is when the abnormality is of the oscillatory type. Most of the developed causality analysis procedures are appropriate for time series of random data and may not work when there are dominant oscillations in variables. Oscillations require special treatment and therefore, a significant part of this study considers the issue of detection, categorization and diagnosis of oscillations.

Oscillations happen in control loops of industrial plants due to various reasons such

as poor controller tuning, external oscillatory disturbances or control valve problems. Oscillation is usually caused in one loop and propagates through various interconnected loops. Regardless of the cause, oscillations disturb the normal plant operation imposing a poor operation condition. Diagnosing the root cause of a propagated oscillation in the operation requires detection of all process variables which are oscillating with same or similar frequency. Detection of oscillations and estimation of their frequencies even though seems to be a trivial task, however is a challenge in process variables contaminated with non-stationary noise. The root cause of the oscillation could be any one of the variables oscillating with similar frequency.

There are several methods proposed in literature for root cause diagnosis of oscillations. However, most of the methodologies can only work for a specific type of oscillation such as the ones caused by valve stiction [5, 6, 7, 8]. Root cause diagnosis of oscillations due to aggressive controller tuning or sinusoidal disturbances require different methodologies compared to oscillations due to nonlinearity. Therefore, before trying to find which loop within the system has caused the oscillation, we need to categorize the oscillation meaning to diagnose if the oscillation is caused by nonlinear valve, controller tuning or an external disturbance. Controller tuning induced oscillations have different characteristics compared to oscillations due to valve nonlinearity or external oscillations. Therefore, a part of this study presents methodologies to categorize the oscillation into three different types. A proper root cause diagnosis method can then be selected based on the identified category of the oscillation.

### 1.1.1 Thesis contributions

A complete abnormal control performance diagnostic framework includes three main stages of detection, categorization and diagnosis. The contributions to each section is briefly described in the following.

1. Detection of control performance abnormalities is generally done based on application of a performance index to announce occurrence of abnormal event. There are several methodologies proposed in literature for the detection purpose. This thesis specifically considers detection of oscillatory faults. Even though oscillation detection is well studied in the literature, detection of multiple oscillations in noisy, non-stationary variables is yet a challenging problem. A more reliable methodology which is easily implementable in an industrial platform is proposed to automatically detect oscillations and estimate the frequencies in the presence of multiple oscillations and noise.
2. Categorizing the type of fault is the next step which is required in order to

select the appropriate root cause diagnosis procedure. As was mentioned, many root cause diagnosis algorithms only work for a specific type of abnormality. Therefore, it is important to find out the type of the fault before root cause diagnosis step. If the abnormality is of the oscillatory type, it could have three different categories depending on which element of the loop has caused the oscillation. Oscillations might be caused by controller tuning, valve problems or external oscillatory disturbances. There are methods proposed in the thesis to distinguish these three oscillation categories from each other. Non-oscillatory types of abnormalities are not further divided into different categories in this study.

3. Two chapters of the thesis propose more reliable methodologies for data-based causality analysis specifically suited for engineering process variables. Causality analysis methodologies are divided into linear and non-linear categories. One chapter proposes a methodology to decompose the variations within a loop to variations coming from different sources from different possible paths. The advantage of the proposed method is that it provides a complete picture of the different paths through which variables can influence each other along with an estimation of the energy transferred through each path independently.

Also, a more reliable causality analysis method based on information transfer concepts is proposed in the last chapter in this thesis. It is not always possible to use linear methods for analysis and modeling. Information transfer methods which are based on comparing probability distributions are promising when the processes are nonlinear.

### **1.1.2 Thesis outline**

The thesis is written in paper format according to the guidelines of Faculty of Graduate Study and Research. Contributions of this research that have been or are to be published in peer-reviewed journals include:

1. E. Naghooisi, B. Huang, Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations, *Industrial & Engineering Chemistry Research*, 53.22 (2014): 9427-9438.
2. E. Naghooisi, B. Huang, Diagnosis of oscillations between controller tuning and harmonic external disturbances, *IEEE on Control Systems Technology*, 23.4 (2015): 1283-1293.

3. to be submitted as E. Naghoosi, B. Huang, Wavelet transform based methodology for detection and diagnosis of multiple oscillations in non-stationary variables.
4. E. Naghoosi, B. Huang, Interaction analysis of multivariate control systems under Bayesian framework, accepted for publication in IEEE on Control Systems Technology, (2016).
5. E. Naghoosi, B. Huang, E. Domlan, R. Kadali, Information transfer methods in causality analysis of process variables with an industrial application, Journal of Process Control, 23.9 (2013): 1296-1305.

The second chapter proposes an algorithm for automatic detection of oscillatory variables and estimation of the oscillation periods. The algorithm is based on detecting and clustering the peak values of the auto correlation function of the variables. The advantage of the algorithm is in detecting the oscillatory variables in the presence of multiple oscillations with no frequency-selection filtering requirement as in the literature. The algorithm is capable of providing an estimation of the individual oscillation frequencies present in the data and also the decay rate of the original signal in the case of damped oscillations. Another advantage is that it requires no or little human interference in the detection process and is easily implementable using any programming platform with a low processing power usage. This fact makes the algorithm more suitable for practical applications. The proposed method is verified through a case study.

The third chapter studies the properties of the responses of linear systems to stochastic disturbances and also the properties of the auto correlation functions of the responses. It is shown that the oscillation generated by a loop due to controller tuning has different ACF properties compared to a harmonic oscillation, which can be utilized to distinguish these two types of oscillations. Two hypothesis tests are developed for automatic diagnosis of oscillations in feedback loops. The result is valuable in the sense that distinguishing oscillations caused by poor controller tuning from external oscillatory disturbances helps in deciding the appropriate trouble shooting procedure.

The fourth chapter complements the previous two chapters by presenting a comprehensive oscillation detection and diagnosis procedure based on wavelet transform. The methodology is capable of both detection and independent diagnosis of multiple oscillations in variables. The independent diagnosis of multiple oscillations is viable due to the inherent capability of wavelet transform in decomposing the variables to its components of various frequencies. Two hypothesis tests are proposed in order to

automatically diagnose if the source of the oscillation is controller tuning, valve nonlinearity or external oscillatory disturbance. The tests are mainly based on properties of wavelet power spectrum and wavelet bicoherence. Wavelet power spectrum is utilized to diagnose oscillations between controller tuning and sinusoidal disturbances. Wavelet bicoherence, similar to the classical bicoherence, can detect and quantify presence of nonlinearity in variables and therefore is an appropriate tool to diagnose if the source of the oscillation is nonlinearity in the process. Advantages of the proposed method is illustrated through analysis of data sampled from an industrial process.

The fifth chapter considers causality analysis based on linear methods and enhances it by performing the analysis under Bayesian framework. A novel method is proposed in order to decompose the estimated transfer function between variables into independent transfer functions, each corresponding to a specific path along which the input can influence the output. The advantage of the proposed method is that it can detect the different path along which variables can influence each other and it provides an estimation of the strength of the different connections between the variables. It also provides a way to check the performance of the controller in disturbance rejection. The analysis is performed by estimating Structural Vector Autoregressive models under Bayesian framework. Bayesian approach provides certain advantages in terms of dealing with high dimensional variables and over parameterization problem. An appropriate design of the prior probability for the model parameters also better ensures convergence to a physically interpretable model. A procedure to design the prior distribution for the model parameters is presented in this chapter.

The sixth chapter studies mutual information and transfer entropy for detection of cause and effect relationships between industrial process variables. Mutual information quantifies the amount of dependency between process variables, while transfer entropy detects the direction of information flow between the variables. The chapter overviews the existing definition and limitations of these two quantities and proposes an algorithm based on combining and extending these two quantities for more reliable identification of causal relationship between process variables. It was shown that time lagged mutual information and differential mutual information can potentially identify the variables with probable causality relationships. After identification of the variables with important dependencies, transfer entropy is used to determine the direction of information flow. Time lagged dependency curves help in tuning the parameters of the transfer entropy in this stage. The proposed algorithm provides information about the type of the relationships between process variables and an estimation of the time delay between them.

## **Chapter 2**

**Automatic detection and frequency  
estimation of oscillatory variables  
in the presence of multiple  
oscillations**

## 2.1 Abstract

Various performance assessment indexes are proposed in literature to announce abnormal performance of single or multiple control loops within the process systems [1, 4, 19, 20]. Oscillatory type of abnormalities require special treatment for reliable detection and diagnosis. When abnormal control performance is detected by the monitoring performance index, another methodology is required to detect oscillation and estimate its frequency. Therefore, several methodologies are developed in literature specifically for the purpose of detection and characterization of oscillatory faults.

Automatic detection of oscillatory variables in the presence of multiple oscillations is still a challenging problem despite there are several methods for detection and estimation of single-frequency oscillation. A new method is proposed in this chapter that utilizes the auto correlation function to detect the oscillatory variables and estimate the oscillation periods in the presence of multiple oscillations. The advantage of the developed method is that it requires little human interference in the detection process given the data is pre-processed appropriately. It is also capable of estimation of the decay rate for decaying oscillations and is advantageous over the methods that are based on analyzing the power spectrum for oscillation detection in case of non-sinusoidal oscillations. The proposed method is verified through a case study.

## 2.2 Introduction

Oscillations happen in control loops of industrial plants due to various reasons such as poor controller tuning, external oscillatory disturbances or control valve problems. Oscillation is usually caused in one loop and propagates through the interconnected loops. Regardless of the cause, oscillations disturb the normal plant operation imposing a poor operation condition. The nature of oscillation could also cause an additional harm by damaging the plant hardware. Therefore, detection and diagnosis of oscillation is of great interest in industry.

Finding the root cause of an oscillation requires detection of all the process variables that are oscillating with the same frequency. The root cause could be any one of the oscillatory variables. Detection of oscillatory variables and estimation of the oscillation period is the first step toward troubleshooting oscillatory faults. There are several methods existing in the literature for detection of oscillation from the routine operational data and some of them are mentioned here.

Jelali et al [21] provides a review on the methodologies for detection and diagnosis of oscillations in process variables. Hagglund [22] presented an on-line oscillation

detection method based on the idea that the integral of the absolute controller error (IAE) between its successive zero crossings is higher for an oscillatory signal compared to a random signal. The threshold for the IAE is determined using some prior information about the control loop. An oscillation is detected if the IAE continuously exceeds the threshold.

Thornhill et al. [23] proposed detecting the zero crossings of the ACF (Auto-correlation function) of variables for oscillation detection and period estimation. In the case of multiple oscillations in the variable, the oscillation in the ACF is no longer regular and frequency filtering is required.

Jiang et al. [24] proposed utilizing the spectral envelope method for both oscillation detection and root cause diagnosis. The method is based on spectral analysis and the likely root cause variables are identified as the ones having the most contribution to the spectral envelope at the oscillation frequency.

Srinivasan et al. [25] presented a modified empirical mode decomposition (EMD) process to isolate dominant oscillations in a time series. The basic idea of EMD method is to decompose the data into IMFs (intrinsic mode function) and a non oscillating residual. Li et al. [26] proposed isolating different frequency components of the signal based on the discrete cosine transform of the data and detecting different oscillations by applying zero crossing method to the isolated components. Xia et al. [27] proposed application of independent component analysis to detect and diagnose oscillations. Tangirala et al. [28] proposed the application of non-negative matrix factorization to detect oscillations in process variables.

There are also papers focusing on detection of harmonics in colored noise. Allen et al. [29] proposed hypothesis testing in order to detect sinusoidal components in the data based on the Monte Carlo singular spectrum analysis methodology. Li et al. [30] proposed a method to detect harmonics in colored noise by estimating the mixed spectrum in two steps. At the first step, the sinusoidal part of the signals are detected and estimated by utilizing Fourier transform and at the second step an AR model is fitted to the residuals to estimate the colored noise.

The common disadvantage of most of the mentioned oscillation detection algorithms is for the case of more than one oscillation in the signal. Oscillations with different frequencies destroy the regular pattern of each other and therefore, detection of oscillatory variables with different oscillations is challenging. Thornhill et al. [23] proposed applying filters to divide the data into separate oscillatory signals covering different frequency ranges and estimating the oscillation frequency of each segment individually. However, as is mentioned in [23] this approach requires designing appropriate filters for each variable and may introduce artificial oscillation

frequencies. Therefore, it cannot be fully automated and requires manual interference. Another disadvantage of the existing oscillation detection methods is that they do not provide any information on the probable oscillation source.

The methodology developed in this work is based on utilizing the ACF of the variables to automatically identify the oscillatory variables in the presence of multiple oscillations via clustering methods. ACF of a periodical signal is also periodic with the same frequency as the original signal. The method applies clustering algorithms to the ACF to detect the peaks of the ACF. The time difference between similar peak values provides an estimation of the oscillation period. The advantage of this method over zero crossing detection of the ACF is that it can also identify the oscillatory signals with multiple oscillations without requiring additional data pre-processing such as filtering. Also, the value of the detected peaks can be utilized to extract more information regarding the variables such as decay rate in the case that the signal is a damped sinusoidal signal. The algorithm is easy to implement and does not require heavy computations which makes it suitable for industrial applications.

The remainder of the chapter is organized as follows. The first section is an introduction to the advantages as well as challenges in utilizing the ACF for detection of oscillatory variables and frequency estimation. Section 2.4 describes the proposed method while a subsection presents the peak detection algorithm followed by the peak clustering algorithm. The third section describes the method for estimation of the decay rate of the signal. This chapter also includes a case study in fourth section. The last section concludes the chapter.

## 2.3 Oscillation detection based on auto correlation function

Auto correlation function is a measure of the linear dependence between observations of a variable itself at different time lags. ACF is defined as:

$$\rho_\tau = \frac{1}{\sigma^2} E[(x_t - \mu)(x_{t+\tau} - \mu)] \quad (2.1)$$

where  $\tau$  is the time lag,  $\sigma$  is the standard deviation of the variable and  $\mu$  is its mean value. For a discrete variable, ACF is estimated from the data as

$$\hat{\rho}_\tau = \frac{1}{N\hat{\sigma}^2} \sum_{t=1}^{N-\tau} (x_t - \mu)(x_{t+\tau} - \mu) \quad (2.2)$$

where  $N$  is the number of samples.

ACF has two special characteristics which make it suitable for detection of oscillation in the data. The first characteristic is that the ACF of an oscillatory variable is also oscillatory with the same period. The second one is that the ACF acts like a filter against noise. ACF of white noise equals to zero except at 0 time lag. The ACF of a variable  $X_t$  as in Equation 2.3, which is a summation of several sinusoidal signals with added white noise, can be obtained as shown in Equation 2.4.

$$X_t = \sum_{i=1}^k A_i \cos(\omega_i t + \phi_i) + \varepsilon_t \quad (2.3)$$

$$\rho_x(\tau) = \frac{1}{\sigma_x^2} \sum_{i=1}^k 0.5 A_i^2 \cos(\omega_i \tau) \text{ for } \tau > 0 \quad (2.4)$$

At zero time lag, another term is added to the summation equal to the noise variance. Thus, ACF contains the same oscillation frequencies while it is free from white noise. The filtering nature of the ACF makes it more suitable for oscillation detection compared to the original time series.

Estimation of the oscillation period from the ACF can be done in two ways: First approach is to detect the time lags where ACF is zero and estimate the period as the time difference between two successive zero crossings multiplied by 2. Second approach is to find the peak values of the ACF of the signal and estimate the period as the time difference between the adjacent peak values.

The main disadvantage of the first approach [23] is its limitation to variables with single oscillatory component. Figure 2.1 plots a variable which is the summation of two sinusoidal signals along with its ACF. The signal is  $x_2(t) = 2\sin(0.2t) + 2.2\sin(0.299t) + n(t)$  where  $n(t)$  is colored noise generated as  $n(t) = \frac{1-0.2z^{-1}}{1-0.1z^{-1}+0.8z^{-2}} \nu_t$  and the variance of  $\nu_t$  is 0.2. The two sinusoidal signals have a period of 31.5 and 21 samples respectively. As can be seen in Figure 2.1, the zero crossings of the ACF do not indicate presence of a regular oscillation in the variable. Thornhill et al. [23] propose applying filters to divide the data into two separate oscillatory signals and estimate the period of each one of them. This approach requires designing appropriate filters for each variable and may introduce artificial oscillation frequency. Due to the need to design filters, the method cannot be automated to detect the oscillatory variables among many process variables in a typical industrial unit.

The other disadvantage of zero crossing detection method in general is the detection of spurious zero crossings due to the noise which requires a pre-processing of the data for noise removal. These two problems are also common in the other oscillation detection methods as investigated in [21].

The above discussion illustrates that more reliable methods are required that can automatically detect oscillatory variables in the presence of multiple oscillations.

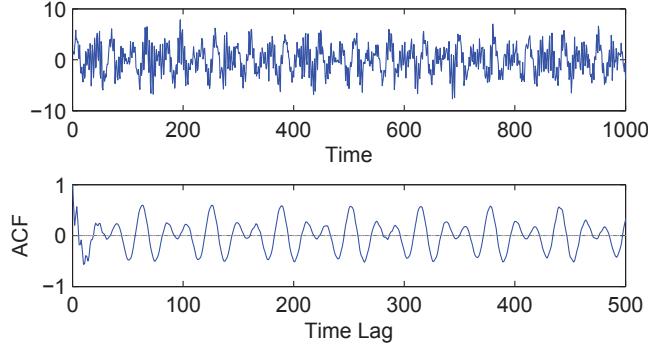


Figure 2.1: Top:  $x_2(t)$  in time. Bottom: ACF of  $x_2(t)$

## 2.4 The proposed method

The algorithm to be developed is intended to run off-line and with little human interference. Each variable is analyzed for oscillation detection and frequency estimation. The variables with similar oscillation frequencies will be grouped together at the end for further analysis to diagnose the root cause.

The input to the algorithm is the routine operation data. There are some minimum requirements of the data for a reliable analysis. The experience shows that a reliable oscillation period estimation from the ACF requires presence of at least 5 periods in the ACF. The maximum time lag of the ACF also needs to be smaller than the length of the data to ensure a reliable estimation of the ACF. The bias of ACF estimation, as in Equation 2.2, increases with the time lag as  $E[\hat{\rho}_\tau] = (1 - \frac{|\tau|}{N})\rho_\tau$  where  $\rho_\tau$  is the true ACF value and  $\tau$  is the time lag. Therefore, the length of the data should be several times of the maximum time lag of the ACF to be calculated to avoid a large bias. If  $N$  is chosen to be 3 times of maximum  $\tau$ , it implies existence of at least 15 oscillation periods in the data. However, smaller length of data can still be used for analysis by considering that the estimation of the period may be less reliable.

The main difference of the proposed algorithm from the existing oscillation detection methods is in detecting oscillatory variables in the presence of multiple oscillations. A variable which is a summation of two oscillatory signals with different oscillation periods is still oscillatory with a period equal to the least multiple of both the individual periods. The same applies to variables including several oscillatory components.

The ACF of a signal with multiple oscillations has several peaks with different values depending on the individual frequencies and their respective power. The peaks corresponding to the main oscillation period (the period which is an integer multiple of all the smaller periods) have the largest values compared to the other peaks in

the ACF. As can be seen in Figure 2.1, although there are different local peaks in the ACF, the highest peaks correspond to the main oscillation of the signal which has 63 samples in the oscillation period. The average of the time differences between the adjacent peaks of the highest values equals to 63 with a standard deviation of 0.8. Therefore, a method based on detection of the peaks and clustering them based on their values is able to automatically detect the oscillatory variables even in the presence of multiple oscillations.

The overall picture of the method is illustrated in Figure 2.2 on the ACF of  $x_2(t)$ . The first step is to find the time lags corresponding to the peaks in the ACF which are marked by circles in Figure 2.2. After that, the peaks with higher values which correspond to the main oscillation should be clustered together while the smaller value peaks form another cluster as shown by the peaks covered by each pair of horizontal lines.

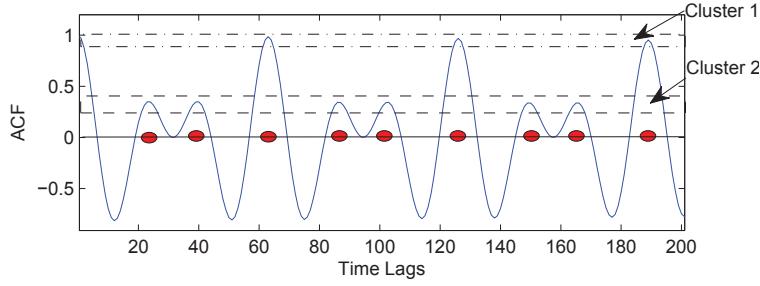


Figure 2.2: Illustration of the general idea

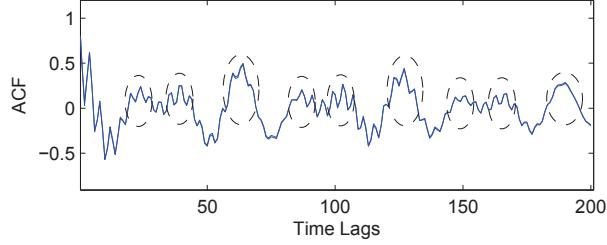


Figure 2.3: Illustration of the problem caused by noise

In reality data contains different kind of noises which makes it more difficult to detect the peaks and cluster them afterwards. Figure 2.3 plots the ACF of the same data as in Figure 2.2 while some colored noise is added to the signal. The colored noise causes some spurious peaks in the ACF which do not represent an oscillation in the data. The algorithm should be able to differ between the spurious peaks and the ones due to the oscillation in the data. To be able to detect the peaks, the data is

divided into different clusters as shown by ellipses in Figure 2.3 and then an algorithm is developed to determine the true peaks based on the clustered data. The method for detecting the peaks and finally estimating the oscillation frequency is presented in the following sections.

## 2.5 Peak detection algorithm

The next step after selection of the data set and estimation of the ACFs is to detect the oscillatory ACFs. A non-oscillatory ACF does not contain peaks with regular time distance between them while the ACF of an oscillatory variable has several peaks on top of a non-oscillatory trend. The approach for peak detection is to cluster the ACF values near the time lags of the peaks.

Some characteristics of this clustering process make it difficult to use conventional clustering algorithms. The aim of the algorithm is that each peak in the ACF be classified into a cluster regardless of the peak value. This clustering is based on the time lags and since there is no prior information about the existence of any peak in the ACF, it is not possible to have an estimation of the number of clusters beforehand. Therefore, an algorithm is required that can capture the shape of the ACF and cluster it according to the time lags of the peaks. Following is the description of the proposed algorithm.

The ACF is sorted in a descending order based on their values. The first cluster is formed by the maximum ACF value along with its corresponding time lag. The next value to be clustered is the second largest value and so on. The clustering rule is based on the time difference between the value to be clustered and the values already in the cluster. The reason for choosing this clustering criterion is the need to cluster the data according to the time. If the time difference of an ACF value from the values in the cluster is less than a threshold, then it cannot represent a new peak and therefore is a member of the existing cluster. A value of ACF that does not fit into any existing cluster forms a new cluster.

For an illustration on this algorithm, consider the ACF of an oscillatory variable with 6 samples per period. Figure 2.4 plots 9 points of the ACF of the variable along with the order that the points are considered for clustering. The first cluster is formed containing the point that is marked by 1. After that, the point marked by 2 has the largest value and since its time difference from point 1 is more than 2 samples (the threshold), it forms a second cluster. 3 is the next point to be clustered which falls into the same cluster as point 1 since its time difference from point 1 is 1 which is less than the threshold. All the samples are clustered in the same way in order of their

value. Here, if the clustering threshold has been set to be larger than 5 samples (time difference between points 1 and 2), all the points would have fallen into one cluster and the two peaks could not be distinguishable.

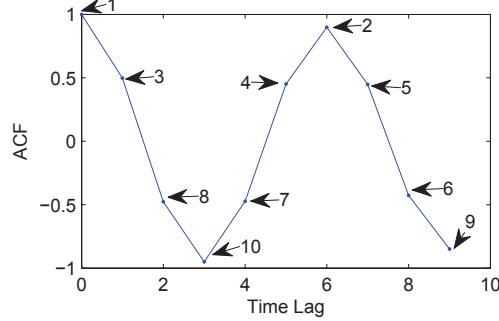


Figure 2.4: 9 points of ACF of a variable along with the order that they are considered for clustering

The threshold should be at least equal to 2 to account for the effect of noise. A too large threshold may result in clustering two distinct peaks of the ACF into one cluster. This emphasizes the need for a small enough sampling time to have at least 4 samples per period. A too small threshold in the case of many samples in one oscillation period leads to detecting spurious peaks due to the effect of noise. The effect of the clustering threshold in detection of peaks is illustrated in Figure 2.5 on two adjacent peaks of Figure 2.3.

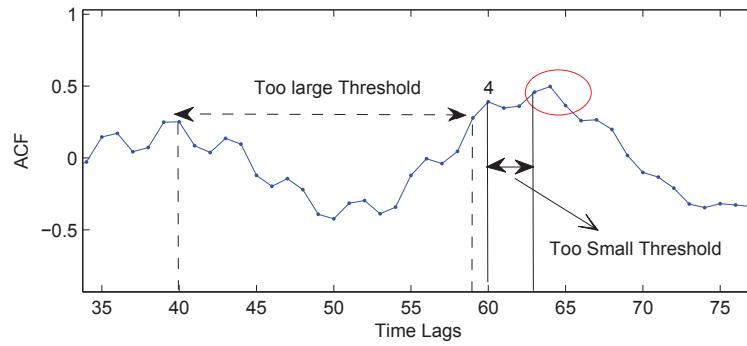


Figure 2.5: Illustration of the effect of clustering threshold on detection of peaks

The circle in Figure 2.5 contains a few largest points which form the first cluster. Point 4 is the next largest point to be clustered and if the clustering threshold is smaller than 3, then point 4 forms another cluster by itself which is spurious. Also if the threshold is larger than the one shown by dashed arrow, then the two distinct

peaks would fall into the same cluster and therefore the oscillation will remain undetected. The method for determining the appropriate threshold will be described in Section 2.6.

The whole ACF of a non-oscillatory data set gets clustered into only one cluster by applying this procedure, since there are no peaks in the ACF to form different clusters. The number of clusters for an oscillatory variable equals to the number of distinct peaks in its ACF. After all the ACFs are clustered, the cluster centers (represented by value and time) are determined as the mean of the two largest ACF values in each cluster and the mean of their corresponding time lags respectively. The average of the two largest values is considered for further filtering the effect of noise. However, averaging over more values is not recommended since it causes problem if there are few samples in an oscillation period. The cluster centers are now regarded as the peaks of the ACF. Figure 2.6 shows how the ACF points in Figure 2.5 are divided into two distinct clusters. The respective cluster centers are marked by stars which are regarded as ACF peaks.

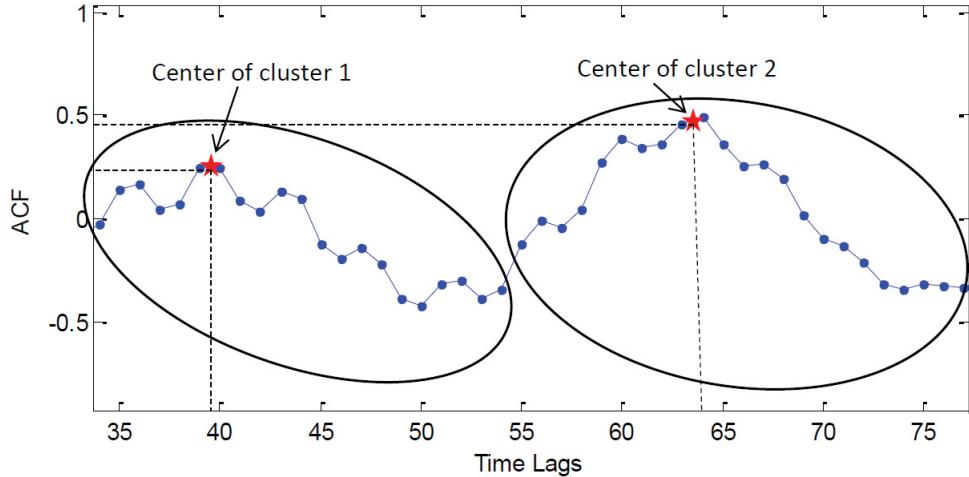


Figure 2.6: The two cluster centers corresponding to the ACF in Figure 2.5

The time differences between the adjacent peak values are approximately equal to the oscillation period if only one oscillation is present in the data. The period is estimated as the average of the time differences between the peaks and the standard deviation can be used as a regularity indicator in the same way as presented in [23]. The regularity index is defined as  $r = \mu/3\sigma$ .  $\mu$  is the average of the time differences between adjacent peaks and  $\sigma$  is the standard deviation of time differences.

When the data contains only one oscillation frequency, the estimated period has

a high regularity index. In addition to a high regularity index, the identified peak values should also be close to each other or have a small variation. Therefore, there are two indicators for existence of multiple oscillations in the signal. The first one is that the estimated period has a high standard deviation. This implies that the time differences between adjacent peaks are not regular. The second one is that the standard deviation of the peak values is high compared to their mean value. A similarity index can be defined to check the similarity between the identified peak values. The similarity index is defined as  $s = \sigma/\mu$ . standard deviation larger than 1/5 of the mean value implies that it is very likely that there are multiple oscillations in the ACF which are generating peaks with different values.

When the estimated period has high regularity but the peak values have a high variation, it is concluded that the data is periodic. However, the estimated period might not be the main oscillation period. Thus, additional analysis has to be done for a more reliable estimation of the oscillation period.

## 2.6 Specifying the threshold for the peak detection algorithm

The performance of the peak detection algorithm directly depends on its threshold. As was mentioned, a large clustering threshold might miss detecting the actual peaks in the ACF while a small threshold may detect false peaks in the ACF. Thus, specifying the threshold is a trade off between missed and false peak detection. One way to determine the threshold is based on some preliminary analysis on the data. Knowledge of the probable oscillation period in the data, helps in determining an appropriate threshold.

Power spectrum of the data provides some information on the possible oscillations in the data. Existence of peaks in the power spectrum does not imply that there are some oscillatory components in the data as there might exist peaks in the estimated spectrum of a purely random noise. Direct use of the peaks in the spectrum for oscillation detection has other disadvantages too. Even in the case that there is one dominant oscillation in the data, the power is distributed to the neighbouring frequencies in the power spectrum depending on the methods used for estimation of the power spectrum. Therefore, it cannot be directly utilized to get a reliable estimation of the oscillation period. The other disadvantage is that the spectrum provides a general presentation of the data in the frequency domain which does not imply if an oscillation persists in the data or fades away. If the oscillation is not persistent, there is no need to detect it from the diagnosis point of view. However,

power spectrum can provide an estimation of the probable oscillation frequencies. The peak in the spectrum corresponding to the highest frequency is used here to determine the clustering threshold.

As was mentioned, oscillations in a variable are always present in its ACF free from the white noise. Here the spectral density of the ACF is utilized instead of the spectrum of the data itself to detect the probable oscillations as is shown in Equation 2.5.

$$I(\omega) = \frac{1}{2\pi T} |\sum_{\tau=0}^{T-1} \hat{\rho}_{\tau} e^{-i\omega\tau}|^2 \quad (2.5)$$

where  $T$  is the length of the ACF. Oscillations with less power compared to noise are more likely to show up as peaks in this estimation compared to the power spectrum of the original signal.

A simple outlier detection method can be used to find the peaks in the spectrum. The peak with the highest frequency determines the possible minimum oscillation period that needs to be detected in the data. Although the peak might be due to the noise, it could also be due to an oscillation in the data. Therefore, the clustering threshold should be smaller than the period corresponding to the highest frequency peak so that it differs between neighboring peaks of the highest frequency oscillation.

There is another important issue that needs to be considered in specifying the clustering thresholds. ACF values as estimated by Equation 2.2 gradually decay with increase of the time lag even if the signal is a pure sinusoidal signal. This decay can be estimated using Equation 2.2 for a pure sinusoidal signal. Assume that the time lag corresponding to a peak value  $p_1$  is  $\tau_1$  and the time lag corresponding to the neighboring peak  $p_2$  is  $\tau_2$ . The oscillation period is equal to  $\tau_2 - \tau_1$  and

$$p_1 = \frac{1}{N\hat{\sigma}^2} \sum_{t=1}^{N-\tau_1} (x_t - \mu)(x_{t+\tau_1} - \mu) \quad (2.6)$$

$$p_2 = \frac{1}{N\hat{\sigma}^2} \sum_{t=1}^{N-\tau_2} (x_t - \mu)(x_{t+\tau_2} - \mu) \quad (2.7)$$

Since  $\tau_1$  and  $\tau_2$  are multiples of the oscillation period and the signal is a pure sinusoidal wave,  $x_t = x_{t+\tau_1} = x_{t+\tau_2}$ . Thus,  $p_1 = \frac{1}{N\hat{\sigma}^2} \sum_{t=1}^{N-\tau_1} (x_t - \mu)^2$  and  $p_2 = \frac{1}{N\hat{\sigma}^2} \sum_{t=1}^{N-\tau_2} (x_t - \mu)^2$  which implies  $p_1 = \frac{N-\tau_1}{N}$  and  $p_2 = \frac{N-\tau_2}{N}$ . The ratio between two peaks in the ACF of a sinusoidal wave can be obtained as

$$p_2 = \frac{N - \tau_2}{N - \tau_1} p_1 \quad (2.8)$$

Therefore, the estimated ACF of a pure sinusoidal signal can be written as

$$\hat{\rho}_\tau = \frac{N - \tau}{N} \cos(\omega\tau) \quad (2.9)$$

If we consider white noise ( $\varepsilon_t$ ) added to the oscillation ( $x_t$ ), then  $p_1$  will be written as in Equation 2.10 which is not equal to  $\frac{N - \tau_1}{N}$  (it is assumed that the mean of the noise and signal is equal to 0). However, The ratio between  $p_2$  and  $p_1$  still equals to  $\frac{N - \tau_2}{N - \tau_1}$  since  $p_2$  can also be written as  $\frac{\frac{1}{N} \sum_{t=1}^{N-\tau_2} x_t^2}{\frac{1}{N} \sum_{t=1}^N x_t^2 + \sigma_\varepsilon^2}$ .

$$p_1 = \frac{\frac{1}{N} \sum_{t=1}^{N-\tau_1} (x_t + \varepsilon_t)(x_{t+\tau_1} + \varepsilon_{t+\tau_1})}{\frac{1}{N} \sum_{t=1}^N (x_t + \varepsilon_t)(x_t + \varepsilon_t)} = \frac{\frac{1}{N} \sum_{t=1}^{N-\tau_1} x_t^2}{\frac{1}{N} \sum_{t=1}^N x_t^2 + \sigma_\varepsilon^2} \quad (2.10)$$

Thus, the ACF of a harmonic with added white noise can be written as

$$\hat{\rho}_\tau = \frac{N - \tau}{N - \tau_r} p_r \cos(\omega\tau) \quad (2.11)$$

where  $p_r$  is a peak in the ACF except for the peak at zero time lag with its corresponding time lag  $\tau_r$ .

The effect of this decay on determining the clustering threshold is the fact that even two adjacent peaks corresponding to one oscillation frequency have different values. According to the clustering rule, the higher ACF values will be clustered first. Considering two adjacent peaks, there might be a few samples between the two peaks that have a higher ACF value compared to the second peak value. Thus, the time difference between the second peak and the already clustered peaks will be smaller than the estimated oscillation period. An example of this is plotted in Figure 2.7. Figure 2.7 plots the ACF of a sinusoidal signal with 36 samples per period in solid line. A horizontal line is plotted at the same height of the last peak in the ACF. All the ACF samples above the horizontal line are clustered before the last peak since they have larger values. Therefore, the minimum time difference between the last peak and already clustered points is less than one oscillation period.

To obtain an estimation of this time difference, consider an extreme example where the length of the ACF ( $l$ ) is exactly 2 times the oscillation period. This implies that the oscillation frequency of the variable is  $\frac{2\pi}{l/2}$ . There are 3 peaks at time lags 0,  $l/2$  and  $l$ , respectively and assume that the length of the data  $N = 2l$ . This is the case of the ACF plotted in Figure 2.7. This case represents the minimum oscillation frequency that can be detected by this method. The value of the last peak ( $p_3$ ) can be obtained as  $p_3 = \frac{N-l}{N-l/2} P_2 = \frac{2}{3} P_2$ . Now, we need to find the time lag  $x$  between  $P_2$  and  $P_3$  so that  $\rho_{l/2+x} = \frac{2}{3} P_2$ . It is possible to ignore the decay in the ACF to find the value of  $x$  and find the time lag  $x$  from  $P_2 \sin(\frac{2\pi}{l/2}(l/2+x)) = \frac{2}{3} P_2$ . It is similar to approximating

the ACF with  $P_2 \sin(\frac{2\pi}{l/2})$  as plotted in dotted line in Figure 2.7. This approximation is valid for the points close to  $P_2$  itself. The value of  $x$  is approximately obtained as  $0.06l$ . Therefore, ACF values up to the point  $0.5l + 0.06l$  are clustered before  $P_3$ . Thus, the clustering threshold should be smaller than  $l - 0.56l = 0.44l$ . This limit is  $0.06l$  smaller than  $0.5l$  which is the oscillation period that can be estimated from the spectrum of the signal.

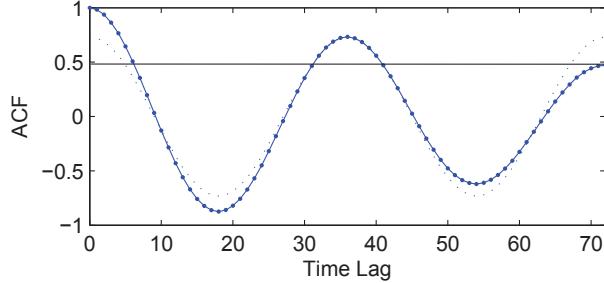


Figure 2.7: Illustration of the effect of the decay in estimated ACF

Considering that the length of the data should be larger than 2 times the length of the ACF, the decay in the ACF will be smaller than the estimated value. Therefore, it is safe to take the clustering threshold equal to  $0.88p_L$  where  $p_L$  is the minimum oscillation period in the data detected from the spectrum. However, considering the uncertainty in the estimated oscillation period and the effect of noise, it is better to take a smaller threshold. In this work, the clustering threshold is considered as  $0.5p_L$  as long as it is greater than 2. The diagram in Figure 2.15 summarizes the algorithm for detection of the peaks in the ACF.

## 2.7 Clustering the identified peaks

The next step of the algorithm is to cluster the identified peaks based on the similarity of their values. When there are multiple oscillations in the signal, the peaks corresponding to the main oscillation have the highest values. However, there are other peak values which correspond to oscillations of different frequencies. By clustering the peaks based on their values, it is possible to detect the similar peaks and estimate their corresponding oscillation periods. Therefore, another clustering algorithm is required when existence of multiple oscillations is confirmed by the peak detection algorithm.

One of the challenges in this clustering is due to the decay in the ACF. The natural decay in the estimated ACF makes it impossible to use existing clustering methods which are based on minimizing the distance of the point from the cluster

centers. In the ACF, each new peak that truly belongs to a cluster has a smaller value compared to the peaks already in the cluster because of the larger time lag and decay of ACF. The decay could be significant when the length of data is short. Therefore, being within a threshold from the latest clustered point is used as the clustering rule instead of closeness to the cluster center.

It is important that the new peak value to be clustered should be compared with the clustered peaks in order of closeness in time. For an illustration of this point, consider a data set as shown in the top panel of Figure 2.8. The bottom panel of Figure 2.8 shows the ACF of  $x_3(t) = 3\sin(0.2t) + \sin(0.1t)$  which is estimated from a data set containing 1500 samples. The larger peak values correspond to the main oscillation with a period of 63 samples. There is one smaller peak in between any two larger peaks due to the sinusoidal signal with a period equal to half of the main oscillation period ( $3\sin(0.2t)$ ). As can be seen in Figure 2.8, the ACF has a strongly decaying behavior because of the very short length of the data used for calculation of the ACF.

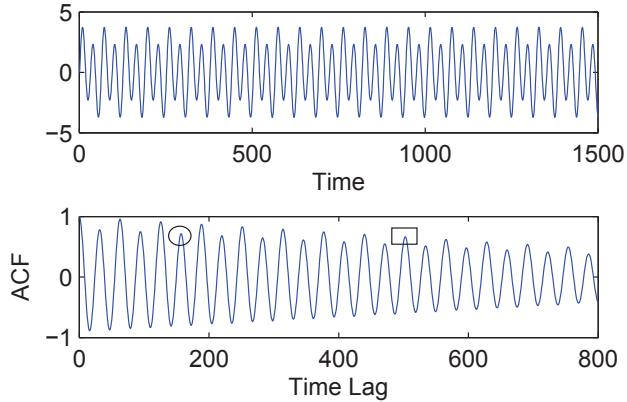


Figure 2.8: Top:  $x_3(t)$  in time. Bottom: ACF of  $x_3(t)$

Two peaks are marked in the ACF plot with a circle and a square. The circle marks a peak value corresponding to the oscillation with higher frequency while the square marks a peak value corresponding to the main oscillation frequency. The clustering algorithm should be able to correctly differ between these two peaks so that they will be assigned to two different clusters. However, as can be seen in Figure 2.8, the peak marked by the square has even a smaller value than the peak marked by the circle due to the decay in the ACF.

Since both the smaller and larger peaks are decaying simultaneously, comparing the new peak value with the clustered peaks at the nearest time can correctly cluster the new peaks. Following is a description of the clustering algorithm: The peak values

are first sorted in order of their time lags. The first cluster is formed containing the peak with the smallest time lag. The next peak value in time is compared to the existing cluster. If the difference between this peak value and the value in the existing clusters is less than a threshold, then the peak falls into the same cluster. Otherwise, the peak forms another new cluster by itself. This procedure continues until all peaks are clustered. This algorithm is illustrated in Figure 2.9 based on 4 peaks of the ACF from Figure 2.1 where the peaks are marked by circles. Again a method for specifying the threshold is required as discussed in the next section.

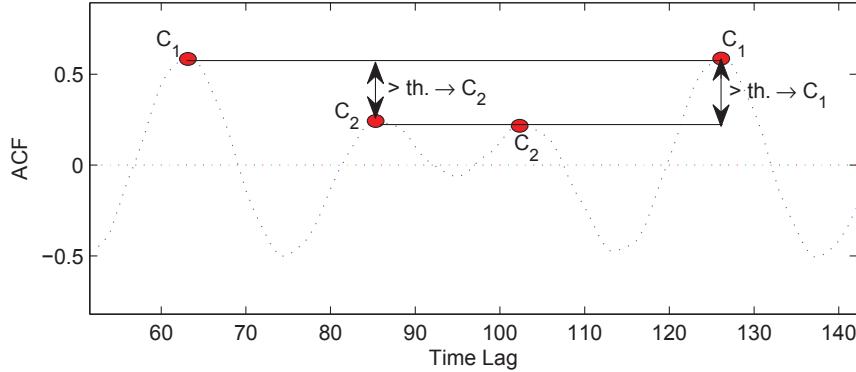


Figure 2.9: Illustration of peak clustering algorithm

### 2.7.1 Specifying the threshold for the peak clustering algorithm

Specifying the clustering threshold directly influences the result of the algorithm. A too large threshold does not differ between the peaks that truly correspond to different oscillation frequencies while a too small threshold can miss the peaks which should be in one cluster.

A threshold value larger than the decay in the ACF can correctly cluster the peaks corresponding to one oscillation into one cluster. Thus, the value of the clustering threshold should be larger than the expected decay in the ACF values. The second consideration on the threshold is that it should be smaller than the minimum difference between the peak values that actually belong to different clusters.

Estimation of the decay for a single sinusoidal wave can help in determining a threshold. Suppose that one cluster is formed containing the first peak  $p_1$  and now clustering the second peak  $p_2$  is to be considered. If  $p_2$  corresponds to the same oscillation as  $p_1$ , then its value should be close to  $\frac{n-\tau_{p_2}}{n-\tau_{p_1}} p_1$  as was described in Subsection

2.6. If  $p_2$  is not close to this value, then it does not belong to the same oscillation represented by the first peak and it should form another cluster. Therefore, if the difference between any two peaks is in the range of 0 to  $(1 - \frac{n-\tau_{p_2}}{n-\tau_{p_1}})p_1$ , the two peaks should be in one cluster. However, the noise in the signal changes the expected value of the second peak. To account for the effect of noise, the absolute value of the difference between the two peaks is compared to  $(1 - c\frac{n-\tau_{p_2}}{n-\tau_{p_1}})p_1$  which implies that the value of the second peak could be as small as  $c(\frac{n-\tau_{p_2}}{n-\tau_{p_1}})p_1$  where  $c$  is positive constant less than or equal to 1. The value of  $c$  depends on the signal to noise ratio and is chosen as 0.7 in this work based on experience. Therefore, the clustering threshold is adaptive and depends on the time difference between the two peaks and the value of the first peak.

To cluster any new peak value, its difference from the clustered peak value that is closest in time is calculated along with the suitable threshold. If the difference is larger than the specified limit, the procedure is repeated considering the next closest clustered peak in time. The peak value forms a new cluster if it does not fall into any of the existing clusters.

It is possible that a cluster only contains one peak value in the case that the data is too noisy. After all the peaks are clustered, the clusters that only contain one peak should merge to the other clusters. Selection of the appropriate cluster cannot be based on the peak value but based on the time lags. The individual peak cluster is merged to the other clusters one by one. The oscillation period for each cluster is recalculated along with the standard deviation before and after addition of the peak. A peak remains in the cluster that has the most reduction in standard deviation after addition of the peak.

The oscillation period is estimated for each cluster separately. Also the mean and standard deviation of the peak values in the cluster is calculated. The cluster that includes the highest peak values, is the one corresponding to the main oscillation period. In the case that any one of the clusters indicates an oscillation with high regularity value, the signal is identified as an oscillatory signal.

Table 2.1 shows the result of applying the second clustering algorithm to the data in Figure 2.1. The first row in Table 2.1 is estimated from the identified peaks before clustering them based on their values. The average of the time differences between the identified peaks is 21.05 which is the smallest individual oscillation period in the data. The average of the peak values is 0.34 with a standard deviation of 0.16 which results in a similarity index equal to 0.47. This similarity index value implies existence of peaks with different values which could be the result of multiple oscillations in the signal.

Table 2.1: Result of applying the algorithm to  $x_2(t)$

period	STD of period	Mean of peaks	STD of peaks
21.05	2.74	0.34	0.16
63	0.77	0.56	0.02
31.58	14.5	0.23	0.06

The peak values are divided into two clusters after applying the second clustering algorithm. One cluster contains the peaks with values close to 0.56 and standard deviation of 0.02. The oscillation period estimated for this cluster is 63 samples with a low standard deviation. This period is the main oscillation period of the signal. The other cluster contains the peaks with values close to 0.23. The oscillation period estimated for this cluster is 31.58 which is very close to the second individual oscillation period in the signal.

## 2.8 Addressing possible problem

When the values of peaks corresponding to different oscillations are very close to each other, the algorithm might not correctly cluster the peak values. One way to approach this issue is to design band pass filters for the data to separate the different oscillations. This approach has the problem of possibly introducing artificial oscillations in the data as well as requiring manual interaction. The proposed approach to rectify this issue in this work is to remove the fastest oscillation of the ACF by down-sampling the data. In this approach, the regular oscillation with minimum period is treated as noise and is removed by down-sampling the original data. The ACF is estimated again and the same clustering algorithm is performed on the ACF of the down-sampled data. This procedure is repeated until all the oscillations are removed and all the ACF values of the last down-sampled data fall into one cluster. The easiest method for down-sampling is utilizing the moving average filter which replaces each  $p_L$  number of samples ( $p_L$  is the least oscillation period) by their mean value as  $y_k = \sum_{i=0}^{p_L-1} u_{k-i}$ .

An important point to be considered when down-sampling the data is the change in the Nyquist frequency. Nyquist frequency equals to half of the sampling frequency. Down-sampling the data by a factor  $p_L$  decreases the Nyquist frequency to  $\frac{1}{p_L} \times f_{N_1}$  where  $f_{N_1}$  is the Nyquist frequency of the original data. Reduction of the Nyquist frequency causes aliasing problem. Aliasing occurs when the frequency of an oscillation is greater than the Nyquist frequency. In that case, the oscillation frequency causes a peak at the aliased frequency which can be obtained from Equation 2.12.

$$f_a = |nf_s - f| \quad (2.12)$$

where  $f_a$  is the aliased frequency in  $H$ ,  $f_s$  is the sampling frequency in  $H$ ,  $f$  is the oscillation frequency and  $n = [\frac{f}{f_s}]$ . Since  $f$  is not known, it is not possible to calculate the aliased frequency.

By down-sampling the data, the oscillations, which were originally lower than the new Nyquist frequency, may fall above it and cause the aliasing problem. That is the reason that a low-pass filter with a cut-off frequency equal to the new Nyquist frequency is always applied to the data before down-sampling to prevent the aliasing.

The problem with applying the low pass filter is that the oscillations present in the data with a frequency above the new Nyquist frequency will be removed and eventually remain undetected. To resolve this issue, it is possible to take the moving average of the data with a rate equal to  $\text{round}(p_L/2)$  (the nearest integer to  $p_L/2$ ). With this filtering, the new Nyquist frequency is  $\frac{1}{p_L}$ . Therefore, there is no need to remove the oscillations with a period between  $p_L$  and  $2p_L$ . The effect of the oscillation with the period of  $p_L$  in the down-sampled data is also removed by the threshold in the peak detection algorithm. By this procedure, the oscillations with periods less than or equal to  $p_L$  are removed as the noise in the data and the oscillations with periods between  $p_L$  and  $2p_L$  are not lost.

To illustrate the effectiveness of this approach consider the example in Figure 2.10. The variable is  $x_4(t) = 2\sin(0.6283t) + 2\sin(0.169t) + 2\sin(1.047t) + n(t)$ . The oscillation periods of the individual sinusoidal signals are 10, 37 and 6 samples respectively. The main oscillation period is  $10 \times 37 \times 3 = 1110$  samples. The ACF of the original variable is shown in Figure 2.10 along with the ACFs of the down-sampled data. As can be seen in Figure 2.10, the peaks of the ACF corresponding to different oscillations have similar values. Also, the different oscillation periods are so close to each other that it is impossible to separate them by applying band-pass filters.

Table 2.2 shows the result of applying the oscillation detection algorithm to  $x_4(t)$ . The first row correspond to the first run of the algorithm on the original data. The algorithm identifies the oscillation with 6 samples per period. In the second run, the data is down-sampled by 3 samples and the algorithm is applied to the down-sampled data. The result is in the second row of Table 2.2. The regular oscillation identified in the second run has 10 sample per period which is another individual oscillation in the variable. The third run of the algorithm after down-sampling the already down-sampled data by 2 samples, results in identification of an oscillation period equal to 37.

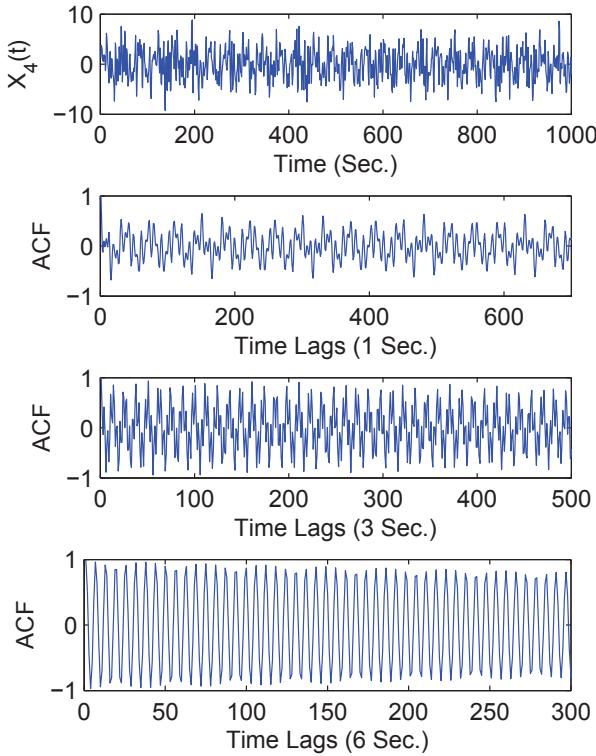


Figure 2.10:  $x_4(t)$  in time and its ACF after downsampling

Table 2.2: Result of applying the algorithm to  $x_4(t)$

Run no.	period	STD of period	Mean of peaks	STD of peaks
1	6.08	1.14	0.19	0.18
2	10	1.41	0.35	0.29
3	37.02	2.26	0.78	0.09

## 2.9 Damped oscillatory signals

In reality, the oscillations are not always a pure sinusoidal signal. Oscillatory variables may have a dampening factor multiplied to them such as  $\exp^{-0.005t} \sin(2\pi f)$ . As was mentioned before, another advantage of the proposed method is that the detected peak values can be used to estimate the decay rate of the signal. Note that this decay exists in the data itself and is different from the decay in the ACF due to the estimation procedure.

The auto correlation of the damped sinusoidal signal has the form of Equation 2.13 where  $x_t$  represents the sinusoidal part of the signal and the decay rate is noted by  $\beta$ . Since only the peaks are analyzed here,  $\tau$  is a multiple of the oscillation period

which implies  $x_{t+\tau} = x_t$ . Therefore,  $\rho_\tau$  can be written as in Equation 2.14 which equals to  $\beta^\tau$ . Equation 2.14 shows that the rate of the decay of the peaks in the ACF is the same as of the original signal.

$$\rho_\tau = \frac{E[\beta^t x_t \beta^{t+\tau} x_{t+\tau}]}{E[(\beta^t x_t)^2]} \quad (2.13)$$

$$\rho_\tau = \frac{\beta^\tau E[\beta^t x_t \beta^t x_t]}{E[(\beta^t x_t)^2]} = \beta^\tau \quad (2.14)$$

The problem of the decay due to estimation of the ACF should again be considered here. The same approach as in Section 2.6 is followed here for estimation of the ratio between two adjacent peak values  $p_1$  and  $p_2$ .  $p_1$  can be estimated as in Equation 2.15.

$$p_1 = \frac{\sum_{t=1}^{N-\tau_1} \beta^{\tau_1+t} x_{t+\tau_1} \beta^t x_t}{\sum_{t=1}^N \beta^{2t} x_t^2} = \frac{\beta^{\tau_1} \sum_{t=1}^{N-\tau_1} \beta^{2t} x_t^2}{\sum_{t=1}^N \beta^{2t} x_t^2} = \frac{N - \tau_1}{N} \beta^{\tau_1} \quad (2.15)$$

$p_2$  is also estimated in the same way as  $p_1$  and equals to  $\frac{N - \tau_2}{N} \beta^{\tau_2}$ . Thus,  $p_2$  can be written as a function of  $p_1$  as in Equation 2.16.

$$p_2 = \frac{N - \tau_2}{N - \tau_1} \beta^{\tau_2 - \tau_1} p_1 \quad (2.16)$$

Again if there is white noise added to the signal, the exact values of the peaks are not as shown in Equation 2.15, but the ratio between them is as in Equation 2.16. To correctly estimate the decay rate of the original signal from the ACF, we need to first compensate the decay in ACF due to estimation. The compensation can be performed by multiplying each peak to  $\frac{N - \tau_1}{N - \tau_p}$  where  $\tau_1$  is the time lag of the second peak in ACF and  $\tau_p$  is the time lag of the successive peaks in the ACF. Here the second peak is taken as the reference for other peaks since its time lag is not too large and therefore the effect of the decay due to its estimation is not considerable. After that, the peak values and their corresponding time lags can be used to estimate the parameter  $\beta$  by fitting the data into the model of Equation 2.17.

$$\hat{\rho}_\tau = a \beta^\tau \quad (2.17)$$

where  $\beta$  is always between 0 and 1. A smaller value of  $\beta$ , implies that the signal fades away faster while a value equal to 1 means that the signal is pure oscillatory with no dampening factor. The implication of  $\beta$  is in contrary to its value. To obtain a definition for decay rate which has a smaller value when the decay in the original data is smaller, it is possible to use a decay index as  $1 - \beta^\tau$  instead of  $\beta$  itself ( $\tau$  is the oscillation period). This decay index has also a meaning and can be obtained from the difference between two adjacent peaks as shown in Equation 2.18.

$$p_2 = \beta^{\tau_2 - \tau_1} p_1 \longrightarrow p_1 - p_2 = (1 - \beta^{\tau_2 - \tau_1}) = 1 - \beta^\tau \quad (2.18)$$

Therefore, the difference between two adjacent peaks in the ACF (after compensating the estimation decay) is  $1 - \beta^\tau$ .

The method is applied on a simulated example where the signal is  $x_5(t) = 5e^{-0.005t} \sin(2\pi/15) + w(t)$ .  $w(t)$  is white noise with variance equal to 1. The time series plot of the data along with its ACF is shown in Figure 2.11. The algorithm detects an oscillation with 14.97 samples per period in the data with a standard deviation equal to 1.4. The identified peaks are used for estimation of the decay rate in the data. The decay rate is obtained as 0.278 before compensating the decay due to the estimation. The decay rate of the original data is  $e^{-0.005} = 0.9950$  which is equivalent to a decay index value of 0.07. The obtained decay rate is much smaller than the true value because of the greater decay that the estimated ACF has compared to the original data. After compensation of the estimation decay in the ACF, the decay rate is obtained as 0.9963 with 95% confidence interval as [0.9928, 0.995]. The true value is indeed in the confidence interval of the estimation. The decay index is also obtained as 0.059.

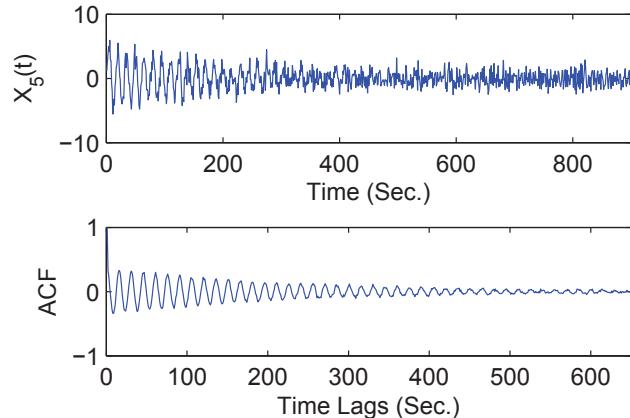


Figure 2.11: Top:  $x_5(t)$  in time, Bottom: ACF of  $x_5(t)$

## 2.10 Case study

The algorithm is also applied to an industrial data set. The data is originated from Eastman Chemical Company and consists of 30 variables which are sampled every 20 seconds. Figure 2.12 is the time series plot of the variables and Figure 2.13 plots the auto correlation functions and the spectrum of the individual variables. The

spectrum of most of the variables shows a peak value around the frequency of 0.0188 which approximately equals to 330 samples. There are also very small peaks with approximately 66 samples per period in the spectrum of tags 20, 17, 16, 15 and 1. Tags 18, 16 and 15 also contain very small peaks close to frequency of 0.356 which is almost 17.6 samples per period. Table 2.3 lists the result of applying the algorithm to the data without any pre-processing or filtering.

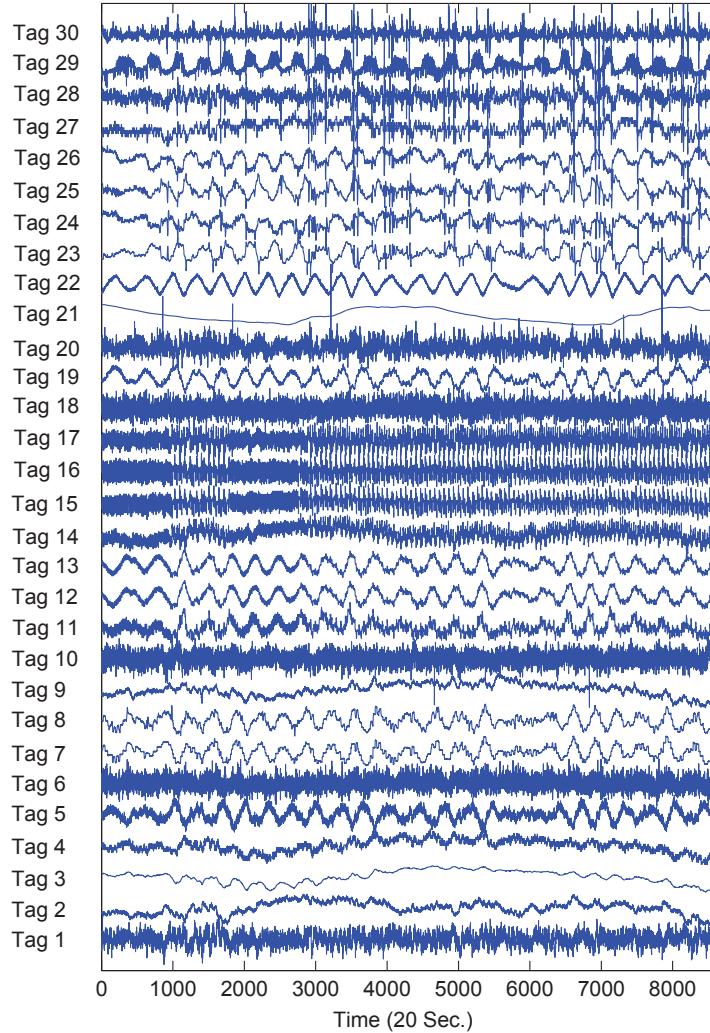


Figure 2.12: Time series plot of the case study variables

As Table 2.3 shows, the algorithm successfully identifies the variables containing peaks in their spectrum and estimates the corresponding oscillation periods even for the ones containing several oscillations. Oscillations in tags 2-4, 9-10, 14-15 and 21 are hidden by a dominant low frequency trend in the data which prevents detection of oscillations. It is necessary to filter the data if it has a dominant trend in order to

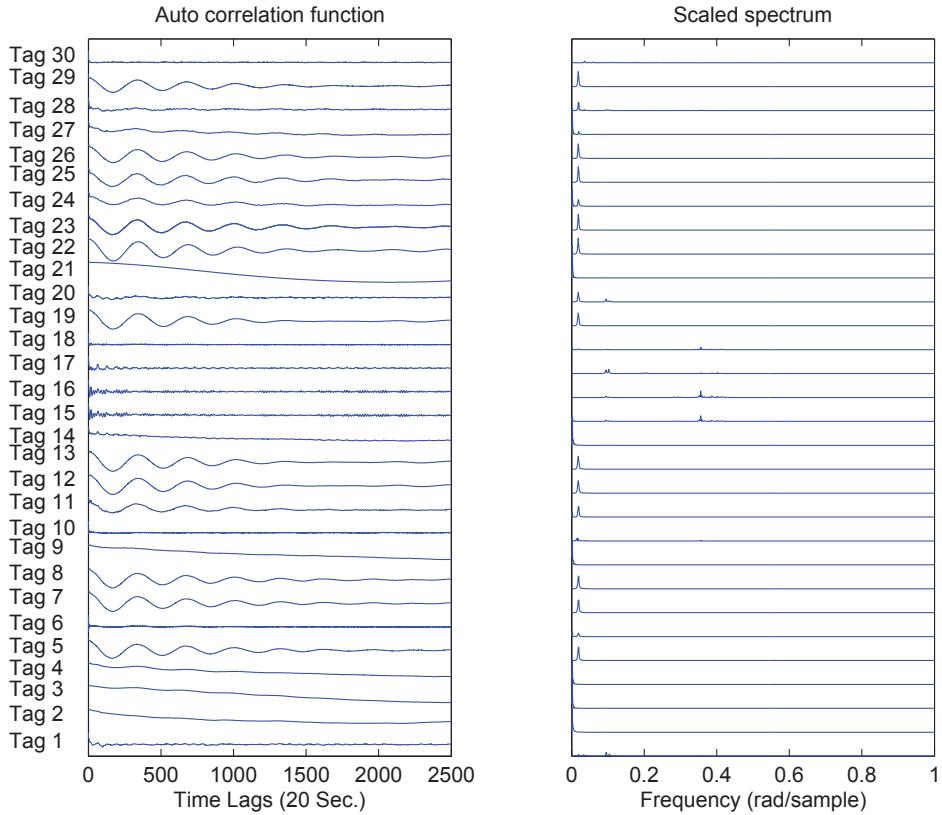


Figure 2.13: ACF and scaled spectrum of the case study variables

be able to detect oscillations.

The cut-off frequency for the filter can be determined from the length of the data. Note that we are using the band-pass filter only to remove the non-stationary trend in the data rather than to separate oscillations. While this filter may be designed automatically, it does not distort the data and this point will be elaborated shortly. The length of the data is about 8600 samples. As was mentioned, the length of the ACF to be calculated should be at least 1/3 of the length of the data which is almost 2800. This implies that only oscillations with a period up to  $2800/5=570$  can be reliably estimated from the ACF. The cut-off frequency should be above this limit. In this work a filter is applied to the data to remove the oscillations with a period above 1000 samples. Also the oscillations with periods less than 2 samples are removed. The result of analyzing the filtered data is also listed in Table 2.3. As can be seen in Table 2.3, the oscillations that were hidden by the trend in the data are also detected after applying the filter.

An important point in selecting the filter is to avoid over filtering the data. For

example, applying a narrow pass-band filter to the data and keeping oscillations in range of 2 to 35 samples per period, causes detection of oscillations in this range in almost all the variables. These oscillations do not contain any considerable power in the original variables which is the reason that there is no peak corresponding to these oscillations in the power spectrum even after removing the trend in the data. The reason that these oscillations are detected after applying the narrow pass-band filter is due to the magnification they receive by removing other frequencies from the spectrum. Filtering may also inserts new oscillations in the data depending on the cut-off frequency and the type of the filter. Therefore, applying filters to the data other than one general filter for removing the trend in the data is not recommended. Even if the filter does not generate artificial oscillations, it is not desired to identify any oscillations in the data that do not actually have any power in the original data, but they are detected simply due to the magnification they receive from the filter.

Table 2.3: Result of analyzing the industrial data

Tag	Large period		Medium period		Small period	
	org.	Filt.	org.	Filt.	org.	Filt.
no. 1	$320 \pm 42$	$324 \pm 57$	$66.7 \pm 16$	$66.7 \pm 16$	-	-
2	-	$622 \pm 37$	-	-	-	-
3	-	$335 \pm 15$	-	-	-	-
4	-	$301 \pm 53$	-	-	-	-
5	$314 \pm 35$	$314 \pm 35$	-	-	-	-
6	$313.8 \pm 26$	$304 \pm 37$	-	-	-	-
7	$334 \pm 0$	$306 \pm 33$	-	-	-	-
8	$306 \pm 42$	$306 \pm 42$	-	-	-	-
9	-	$279 \pm 73$	-	-	-	-
10	-	$418 \pm 92$	-	-	$17 \pm 5$	$17.7 \pm 5.6$
11	$309 \pm 37$	$309 \pm 37$	-	-	-	-
12	$307 \pm 58$	$358 \pm 32$	-	-	-	-
13	$358 \pm 60$	$358 \pm 60$	-	-	-	-
14	-	$648 \pm 101$	-	$63 \pm 9.7$	-	-
15	-	$354 \pm 114$	$65 \pm 7.6$	$65 \pm 7.4$	$7.6 \pm 2.5$	$17.6 \pm 2.5$
16	$303 \pm 40.8$	$303 \pm 40.8$	$65 \pm 9.8$	$65 \pm 9.8$	$17.6 \pm 2.6$	$17.6 \pm 2.6$
17	-	$408 \pm 136$	$65.9 \pm 11.2$	$65.9 \pm 11.2$	$32.5 \pm 4.9$	$32.5 \pm 4.9$
18	$357 \pm 83$	$341 \pm 38$	-	-	$16.4 \pm 4.8$	$16.6 \pm 4.7$
19	$355.6 \pm 55$	$355.6 \pm 25$	-	-	-	-
20	$351 \pm 46$	$351 \pm 46$	$67 \pm 12.5$	$67 \pm 12$	-	-
21	-	$468.6 \pm 59.6$	-	-	-	-
22	$353 \pm 18$	$353.8 \pm 18.2$	-	-	-	-
23	$337 \pm 0.7$	$337 \pm 0.7$	-	-	-	-
24	$349 \pm 35$	$348.6 \pm 34.6$	-	-	-	-
25	$339 \pm 1.41$	$339 \pm 1.41$	-	-	-	-
26	$359 \pm 32$	$358.8 \pm 31.9$	-	-	-	-
27	$332 \pm 27$	$328 \pm 23$	-	-	-	-
28	$330 \pm 37$	$330 \pm 37$	-	-	-	-
29	$357 \pm 41$	$357 \pm 37$	-	-	-	-
30	-	-	-	$120 \pm 38$	-	-

## 2.11 Non-sinusoidal signals

Nonlinear oscillations which are mainly produced by valve stiction do not have an exact sinusoidal form. It is reported that the controller output in a feedback loop suffering from valve stiction has a triangular shape while the process output has a square wave type of oscillation [31]. Non-sinusoidal oscillations have several harmonics in addition to the main oscillation. This fact may cause problem while applying

methods based on the analysis of power spectrum for oscillation detection. However, since ACF does not decompose the original signal, it will still be oscillatory with the same period as the original signal. For an illustration, Figure 2.14 plots a signal which is a summation of 4 different types of signals along with its components. As can be seen in Figure 2.14, the power spectrum has several peaks corresponding to the harmonics of triangular and square waves. Determining the main oscillation frequencies present in the original signal from the power spectrum is difficult while the developed algorithm correctly detects the oscillation periods of triangular and square waves as 15 and 40 samples respectively.

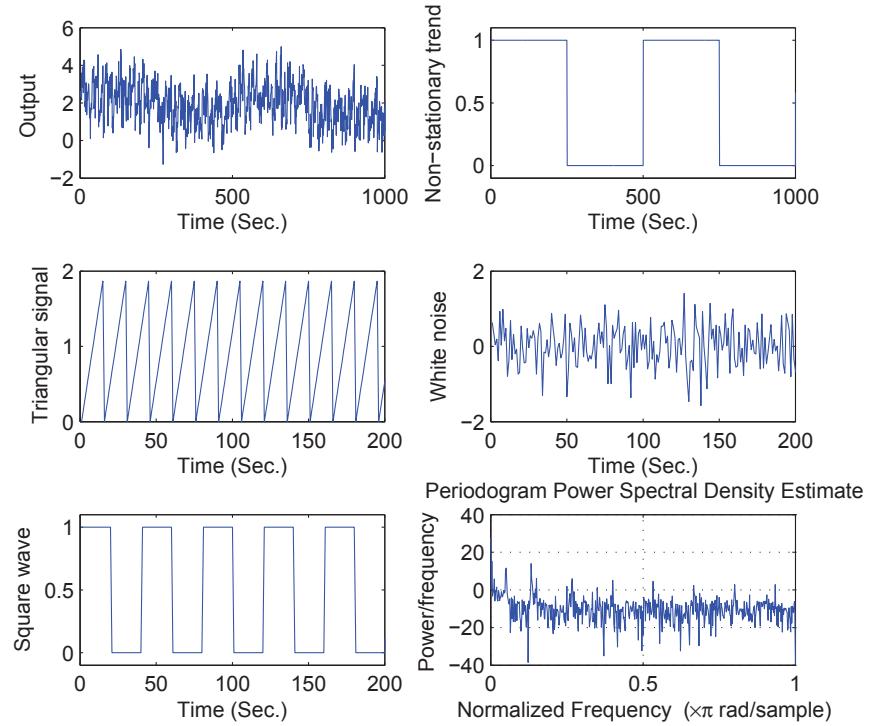


Figure 2.14: Components of a variable containing non-sinusoidal oscillations

## 2.12 Summary

The chapter proposed an automatic algorithm for detection of the oscillatory variables and estimation of the oscillation periods. The algorithm is based on detecting and clustering the peak values of the auto correlation function of the variables. The advantage of the algorithm is in detecting the oscillatory variables in the presence of multiple oscillations with no frequency-selection filtering requirement in order to

separate oscillations. The algorithm is capable of providing an estimation of the individual oscillation frequencies present in the data and also the decay rate of the original signal in the case of damped oscillations. Another advantage of the algorithm is that it is easily implementable using any programming platform with a low processing power usage. This fact makes the algorithm more suitable for practical applications

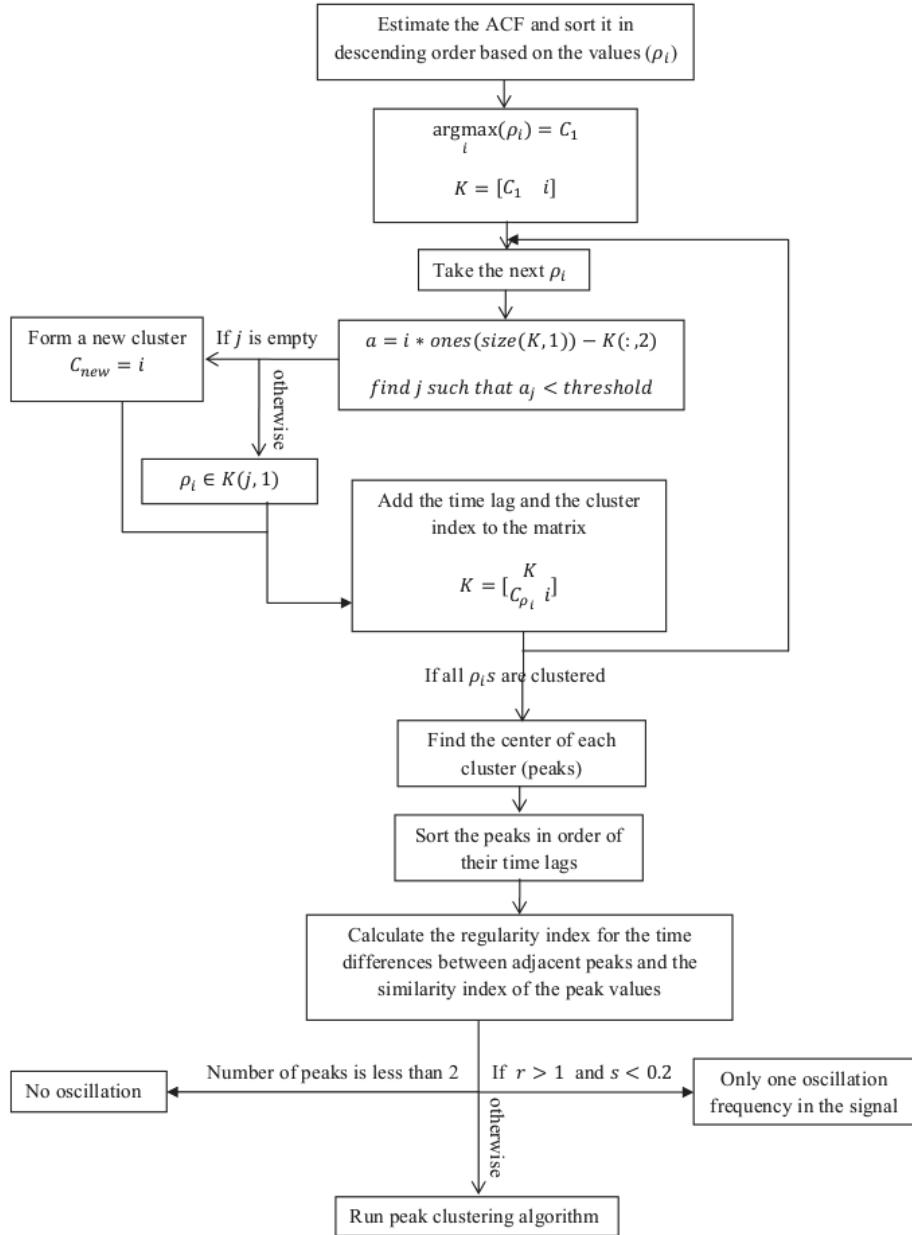


Figure 2.15: Algorithm for oscillation detection

# **Chapter 3**

**Differentiating between oscillations  
due to controller tuning and  
harmonic external disturbances**

### 3.1 Abstract

After oscillation detection, the next stage is to categorize the type of the oscillation. in this chapter, oscillatory faults are divided into three different categories based on the element of the loop which causes the oscillation: oscillations due to controller tuning, valve problems and external oscillatory disturbances. Selection of the appropriate root cause diagnosis and troubleshooting methodologies requires the knowledge of which element in the loop has caused the oscillation. The motivation is that most of the developed diagnostic algorithms are specialized in finding the source of a specific type of abnormality. For example, there are several methods specifically targeting root cause diagnosis of an oscillation caused by a sticky valve [6, 5, 31] which cannot work if the oscillation is actually caused by a poorly tuned controller. This section considers developing algorithms in order to identify the category of the oscillation before trying to find the loop which is the root cause of the fault.

The properties of linear systems described by discrete-time stochastic dynamic models and the conditions which yield an oscillatory response are studied here. Feed-back loops which are oscillatory due to the controller tuning fall into this category of systems. Therefore, analyzing linear systems with oscillatory responses, reveals the properties of self-oscillatory feedback loops. It is shown that in the presence of random disturbances, the oscillation due to the controller has a varying amplitude and phase even though the oscillation frequency is constant. The auto correlation function of this type of oscillatory signals has a decay ratio and does not have a deterministic amplitude of oscillation. These properties can be utilized to distinguish between controller induced and externally introduced oscillations. Two hypothesis tests are developed to facilitate a solution for the above stated oscillation diagnosis problem.

### 3.2 Introduction

Oscillations in a feedback loop can be introduced by controller tuning, nonlinearity such as valve induced problems or external harmonic disturbances. Oscillations due to the valve problems have the distinct property of being nonlinear which make them distinguishable from other types of oscillations [31]. Nonlinear oscillations have several harmonics in addition to the main frequency oscillation which have phase coupling. Methods like non-linearity ranking utilizing surrogate data [32] or bicoherence [6] are suitable for detection and diagnosis of nonlinear oscillations which are most often caused by a sticky valve. Jelali et al. [21] provides a literature review on

developed methods for oscillation detection and diagnosis in control loops with an emphasis on valve stiction detection.

The other two types of oscillations, external harmonics and oscillations due to the controller tuning, are very similar to each other and there is a lack of methods for distinguishing these two types of oscillations. Oscillations due to the controller tuning happen when the controller is tuned in such a way that makes the closed loop to have at least a pair of complex poles. The imaginary part of the complex pair of poles causes oscillations in the closed-loop system. When a loop has consistent linear oscillation, the question is if it is due to the controller tuning or a harmonic disturbance.

The problem of distinguishing oscillations between controller tuning and harmonic disturbances has also been studied in some other papers as explained below. Estimating the ultimate frequency of the oscillation due to the controller by modelling the loop is a possible diagnosis method. Karra et. al. [33] proposed a solution based on a system identification approach in order to distinguish between external oscillatory disturbances and oscillations due to controller tuning. It is also proposed that a marginally stable controller has a poor performance at all frequencies based on minimum variance benchmark which can be used as a criterion for distinguishing oscillations [34]. Babji et. al. [35] proposed a method based on the assumption that an oscillation due to the controller has a higher amplitude compared with an external oscillatory disturbance. However, the assumptions used in these methods do not hold in general and applying the previously proposed methodologies is not always viable.

It is worth mentioning that there are many papers that consider detection of harmonics in colored noise [36, 29, 30], which bear some similarity with the subject to be studied here. However, these methods try to compensate the effect of colored noise in order to detect and estimate the frequency of harmonics in the data, which is an oscillation detection problem. This chapter is to identify oscillatory colored noise from harmonics.

The chapter studies the condition for a loop to be oscillatory due to the controller. It is shown that a consistent oscillation in the time domain due to the controller does not require the loop to have poles with pure imaginary values. Any complex pair of poles will cause persistent oscillation in the loop even if the real part is not zero in the presence of stochastic disturbances. Although the fact that systems with complex conjugate poles driven by white noise oscillate at their natural frequency is very well known in literature with the name of resonance [37], it has not been used in oscillation root cause diagnosis of control systems.

The difficulty in distinguishing externally introduced harmonics and oscillations

due to controller tuning is due to the fact that they are similar to each other in the time domain. However, it will be shown that an oscillation due to controller tuning has a different auto correlation function from a harmonic disturbance of the same frequency. Two hypothesis tests are developed to distinguish these two types of oscillations from each other. The method is also useful in detection of harmonics in colored noise which is developed on the basis of our previous work on oscillation detection [38].

The rest of the chapter is organized as follows. Section 3.3 studies the response of a general ARMA process to the input in time domain. In particular, Section 3.3.2 studies the form of the ACF of the response of a general ARMA process. Section 3.4 studies the form of the ACF of harmonic disturbances. Section 3.5 discusses the possible methods to distinguish oscillations due to the controller tuning from harmonic disturbances. The case of pure imaginary poles in the system is investigated in Section 3.6. A case study is also included in Section 3.7. Conclusions are drawn in Section 5.10.

### 3.3 Response of a general ARMA process

This section reviews the known properties of responses of systems with complex conjugate poles to stochastic disturbances as the foundation for the diagnosis methodology to be discussed later. Discrete time linear stochastic systems are usually expressed in the form of ARMA models. An ARMA model consists of a moving average and an autoregressive part. The forms of the responses of these two parts in time domain are investigated separately and then combined to form the response of the whole process. The form of the response of a  $p_{th}$  order AR (autoregressive) model can be obtained as a generalization of the response of a  $2^{nd}$  order AR model. Thus, first the response of a general  $2^{nd}$  order AR model as defined in Equation 3.1, is investigated.

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = \varepsilon_t \quad (3.1)$$

Since the input to ARMA processes is white noise which does not have a closed form expression, the difference equation is solved in between two time samples [39]. The response of the process in between two input samples is obtained by solving Equation 3.1 assuming that  $\varepsilon_t$  is constant within the sampling interval. It is known that the solution to a difference equation consists of two homogeneous and particular solutions. The homogeneous solution of a difference equation only depends on the system properties and its initial values while the particular solution is obtained by

considering the form of the input to the process [39]. First, we consider the homogeneous solution and then the form of the particular solution is investigated.

The homogeneous solution to Equation 3.1 has three different forms based on the values of the two roots of the equation (poles of the system) [40].

- The two roots are real and distinct. The solution to the equation is obtained as:

$$y_t = b_1 \alpha_1^t + b_2 \alpha_2^t \quad (3.2)$$

where  $b_1$  and  $b_2$  are constants determined from the initial conditions and  $\alpha_1$  and  $\alpha_2$  are the two roots.

- The two roots are real but equal, then,

$$y_t = (b_1 + b_2 t) \alpha^t \quad (3.3)$$

where  $\alpha$  is the repeated root and  $b_1$  and  $b_2$  are determined from the initial values.

- The two roots are complex conjugate which yeilds a response as in Equation 3.4.

$$y_t = b^* \alpha^t + b \alpha^{*^t} \quad (3.4)$$

where  $*$  notes conjugate transpose. If the roots are expressed as  $r e^{i\theta}$  where  $r = \sqrt{a_2}$  and  $\cos(\theta) = -a_1/(2r)$  and  $\sin(\theta) = \frac{|a_1^2 - 4a_2|^{0.5}}{2r}$ , then the response can be written as:

$$y_t = r^t (d_1 \cos t\theta + d_2 \sin(t\theta)) = g_1 r^t \cos(t\theta + g_2) \quad (3.5)$$

where  $d_1$ ,  $d_2$ ,  $g_1$  and  $g_2$  are determined from the initial values. Therefore, the response is oscillatory but it has a decay ratio equal to  $\sqrt{a_2}$ . When  $a_2 = 1$ , the response is a pure sinusoidal wave. It should be noted that this decay ratio is different from damping factor of the second order transfer function formulation.

Considering a  $p_{th}$  order process, for each pair of complex conjugate roots repeated  $n$  times, the following term is included in the response [40].

$$r^t [\beta_1 \cos(t\theta + \phi_1) + \beta_2 t \cos(t\theta + \phi_2) + \dots + \beta_n t^{n-1} \cos(t\theta + \phi_n)] \quad (3.6)$$

The one-sample ahead particular solution of a general AR model in response to a constant input is also a constant value. Therefore, the combination of the homogeneous and particular solution is of the form of the homogeneous solution added with a constant.

The response of an ARMA process defined as in Equation 3.7 also has the same form as the response of its AR part.

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q} \quad (3.7)$$

Moving average part of the process only changes the particular solution which is still a constant value within sampling interval. The change in the constant value added to the homogeneous solution changes the mean value and the coefficients of the homogeneous solution by changing the initial values at each sample time. Therefore, analysis of AR processes is sufficient in order to investigate the properties of oscillatory linear processes since only the autoregressive part of the process can generate the oscillation.

### 3.3.1 The response of ARMA process in reality

The input to a linear feedback loop described by ARMA models in reality, is not only one constant value. The input is the noise entering the loop. By receiving a new noise value, the difference equation is solved again to produce the corresponding response. The form of the system response remains the same at all sample times depending only on the system properties. The difference in the response is caused by different initial and input values at each sampling instant.

In the case of a second order AR process with imaginary poles,  $g_i s$  in Equation 3.5 change at each sampling instant while  $\theta$  and  $r$  are constant. This change in parameters yields an oscillation with varying amplitude and phase but with constant frequency and decay ratio. The response of such a system can be shown as in Equation 3.8 where the change in amplitude and phase is shown by  $\beta_t$  and  $\phi_t$  as functions of time:

$$y_t = \beta_t r^t \cos(t\theta + \phi_t) + \varepsilon_t \quad (3.8)$$

To summarize the above discussions, the responses to the stochastic disturbances by ARMA processes with complex poles are always oscillatory with constant frequency and varying amplitude. It is not required that the real parts of the poles to be zero in order to obtain persistent oscillation in the response. If the input is constant, the response converges to a constant value depending on the input value. However, consecutive varying inputs to the process keep the response constantly oscillating if the time difference between consecutive inputs is not long enough for the process to settle down. In practice, there is always some kind of noise entering the loop at each sampling instant (piece wise constant or even continuous change) which does not let the process to settle on a constant value and therefore the response is constantly oscillating.

### 3.3.2 ACF of ARMA processes

For a  $p_{th}$  order autoregressive model defined as in Equation 3.9, the ACF has the form of the Equation 3.10 according to Yule-Walker theorem [40].

$$Y_t = \frac{1}{1 - a_1 z^{-1} - \dots - a_p z^{-p}} \varepsilon_t \quad (3.9)$$

$$\rho_j = a_1 \rho_{j-1} + a_2 \rho_{j-2} + \dots + a_p \rho_{j-p} \quad (3.10)$$

where  $\rho$  denotes the ACF and  $j \geq p$ . Equation 3.10 is the same difference equation as that the process itself. When the poles of the system ( $\alpha_i$  is Equation 3.11) are distinct, the solution of Equation 3.10 can be obtained as in Equation 3.11:

$$\rho_j = g_1 \alpha_1^j + g_2 \alpha_2^j + \dots + g_p \alpha_p^j \quad (3.11)$$

where  $g_i$ s are determined from the initial values.

Considering a moving average process as defined in Equation 3.12, it can be shown that the ACF has the form of Equation 3.13 [40].

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.12)$$

$$\rho_j = \frac{\sigma_\varepsilon^2}{\sigma_Y^2} (\theta_j + \theta_{j+1} \theta_1 + \theta_{j+2} \theta_2 + \dots + \theta_q \theta_{q-j}) \text{ for } j = 0, 2, \dots, q \quad (3.13)$$

The ACF of a MA process is non-zero only for the time lags equal or less than the order of the MA process. Thus, the ACF of an ARMA process is the same as the ACF of its autoregressive part after the first few lags depending on the order of the MA part. Therefore, it is sufficient to analyze the AR part of the process in order to study the form of the ACF.

When there are complex roots in the AR part of the process, the ACF will be oscillatory. To learn the behavior of the ACF in the case of complex roots, we can consider a 2<sup>nd</sup> order AR process as in Equation 3.1. The ACF of a general  $AR_2$  process is of the form of Equation 3.11 and the two coefficients are determined from  $\rho_0 = 1$  and  $\rho_1 = -a_1/(1 + a_2)$ . Thus, the ACF of an  $AR_2$  process can be written as in Equation 3.14.

$$\rho_j = \frac{(1 - \alpha_2^2) \alpha_1^{j+1} - (1 - \alpha_1^2) \alpha_2^{j+1}}{(\alpha_1 - \alpha_2)(1 + \alpha_1 \alpha_2)} \quad (3.14)$$

When the two roots are complex conjugate (happens if  $a_2 > a_1^2/4$ ) the roots can be written as  $\alpha_1 = \sqrt{a_2} \exp(i\theta)$  and  $\alpha_2 = \sqrt{a_2} \exp(-i\theta)$  where  $\cos(\theta) = -a_1/2\sqrt{a_2}$ . The ACF has the form of Equation 3.15 by substituting the roots into Equation 3.14.

$$\rho_j = \frac{a_2^{j/2} (\sin(j+1)\theta - a_2 \sin(j-1)\theta)}{(1 + a_2) \sin \theta} \quad (3.15)$$

The ACF is oscillatory and it has a decay ratio of  $a_2^{j/2}$ . For the process to be stable, it is required that  $a_2$  is less than 1. If  $a_2$  is exactly equal to 1, then the ACF is pure oscillatory and does not converge to zero. Note that the decay ratio of the ACF and its frequency are exactly the same as that of the response of the system. The decaying oscillation in the ACF of oscillations generated by linear systems is the special property of these type of oscillations which will be utilized later for the oscillation diagnosis purpose.

The theoretical form of the ACF of the response of a linear system to white noise is similar to its response to non-zero initial values since they are both homogeneous solutions of the system's equation. However, the estimated ACF from the data does not converge to zero even for small decay ratios. The tail of the estimated ACF will always be oscillatory with small amplitudes depending on the decay ratio. For example, middle panel of Figure 3.1 shows the estimated ACF of  $y_1(t) = \frac{1}{1-0.3z^{-1}+0.9z^{-2}}\varepsilon_t$  where  $\varepsilon_t$  has a variance of 1. As can be seen in Figure 3.1, the ACF has a decaying behavior up to time lag of 100 and has smaller amplitude oscillations for larger time lags while the response of the system to a non-zero initial value converges to zero much faster.

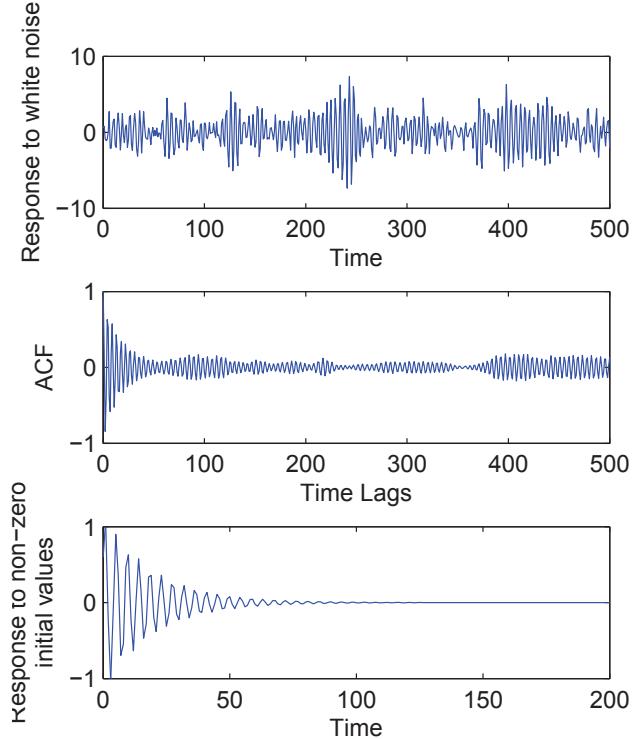


Figure 3.1:  $y_1(t)$  in time, its ACF and response to non-zero initial values

### 3.3.3 Continuous time systems

It should be mentioned that continuous systems also behave in the same way as described for the discrete time systems. A pair of complex roots ( $a \pm bj$ ) in the differential equation describing a linear system, adds the term  $g_{1t}e^{at}\cos(bt + g_{2t})$  to the system's response to varying input values. For example, the response of a system described by  $\frac{Y(s)}{U(s)} = \frac{e^{-2s}}{25s^2+6s+1}$  which is controlled by a PID controller of  $1 + \frac{0.1}{s} + \frac{50}{1+\frac{100}{s}}$  is plotted in Figure 3.2. The top panel of Figure 3.2 plots the step response of the system and the middle panel plots the response of the system to white noise with variance of 1. The bottom panel of Figure 3.2 plots the ACF of the response to the white noise. It should be noted that the data are sampled at every second in order to estimate the ACF and plot the results. The oscillation period of the ACF can be estimated as 22.5 samples.

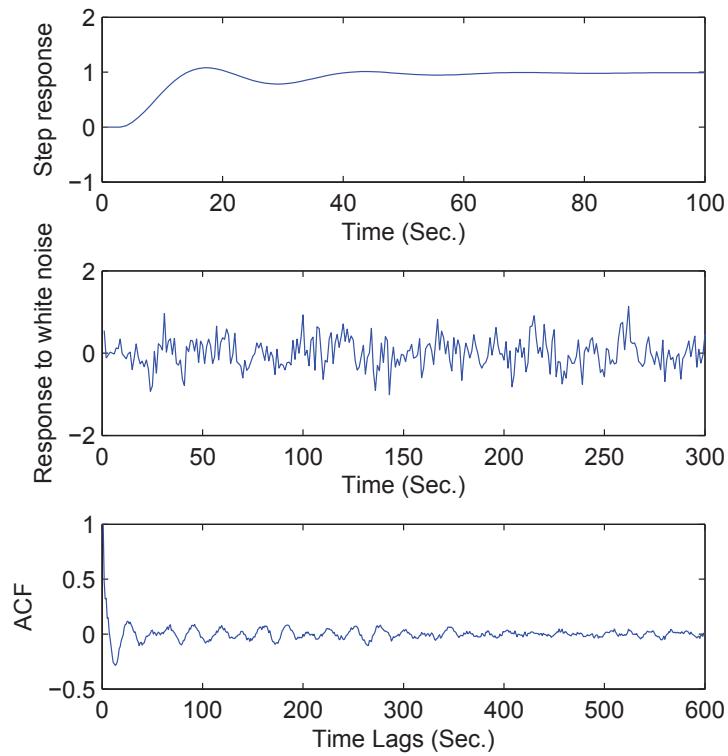


Figure 3.2: Top panel: step response, Middle panel: response to white noise, Bottom panel: ACF of the response to white noise of a continuous time linear system controlled with a PID controller

To compare the response of the simulated linear system to a sinusoidal disturbance with the response of the system to white noise, Figure 3.3 plots its response to a

sinusoidal signal and its ACF. It is observed that the response of the system to a sinusoidal disturbance is a sinusoidal signal with constant amplitude and phase as is expected. The ACF of the response also is different from the ACF of the response to white noise.

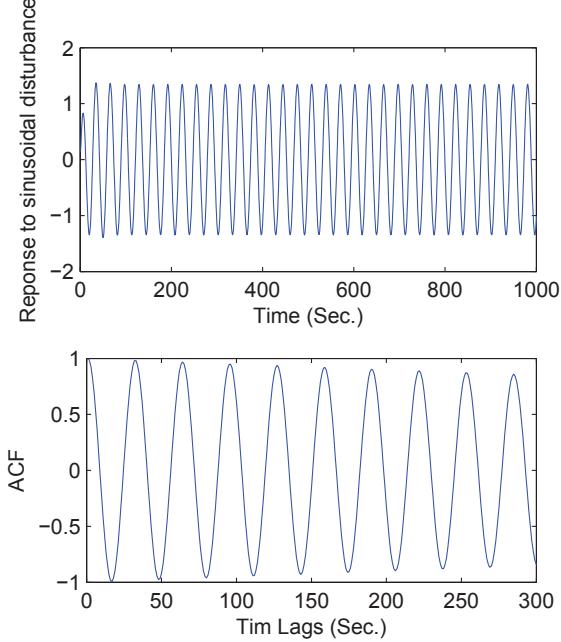


Figure 3.3: Top panel: response of the simulated continuous time system to oscillatory disturbance, Bottom panel: ACF of the response

### 3.4 ACF of harmonics

A harmonic process means a pure sinusoidal signal with added white noise. The ACF of a harmonic process as in Equation 3.16 can be obtained as shown in Equation 3.17.

$$X_t = \sum_{i=1}^k A_i \cos(\omega_i t + \phi_i) + \varepsilon_t \quad (3.16)$$

$$\rho_x(\tau) = \frac{1}{\sigma_x^2} \sum_{i=1}^k 0.5 A_i^2 \cos(\omega_i \tau) \quad (3.17)$$

where  $\sigma_x$  is the standard deviation of  $X_t$  and  $\tau$  is time lag. The ACF of a harmonic process is oscillating with exactly the same frequency and a constant amplitude. This fact implies that the ACF of a harmonic process is deterministic and can be exactly determined by the knowledge of its frequency and amplitude. However, in practice, the ACF calculated from data of finite length is also a function of the length of the data and the time lags as discussed in [38].

The ACF is estimated by Equation 3.18 from the data.

$$\hat{\rho}_\tau = \frac{1}{N\sigma^2} \sum_{t=1}^{N-\tau} (x_t - \mu)(x_{t+\tau} - \mu) \quad (3.18)$$

where  $N$  is the number of samples,  $\mu$  is the estimated mean of the data and  $\sigma$  is the standard deviation. This estimation is known to have the minimum estimation error compared to alternative methods [41]. It artificially introduces a decay in the estimated ACF due to the finite number of data samples used for estimation, even when the signal is pure sinusoidal. This decay is linear with respect to time.

Each peak value of the ACF of a pure sinusoidal signal has a value equal to  $\frac{N-\tau}{N}$  where  $\tau$  is the corresponding time lag. The reason is that  $\rho_\tau = \frac{1}{N\sigma^2} \sum_{t=1}^{N-\tau} (x_t - \mu)(x_{t+\tau} - \mu) = \frac{1}{N\sigma^2} \sum_{t=1}^{N-\tau} (x_t - \mu)^2 = \frac{N-\tau}{N}$ . Therefore the ACF of a single sinusoidal signal can be written as  $\frac{N-\tau}{N} \cos(2\pi f\tau)$  where  $f$  is the oscillation frequency.

When there is white noise added to the signal, the peak values in ACF are no longer equal to  $\frac{N-\tau}{N}$  because of the effect of white noise at zero time lag. However, the effect of noise is not present at the second peak and so on, which facilitates the estimation of ACF. For example, the ACF of a signal  $\cos(2\pi ft) + \varepsilon_t$  (where  $\varepsilon_t$  is white noise) can be written as shown in Equation 3.19 [38].

$$p_r \frac{N-\tau}{N-\tau_r} \cos(2\pi f\tau) \quad (3.19)$$

where  $p_r$  is the value of a peak in the ACF (except for the peak at lag 0) and  $\tau_r$  is its corresponding time lag.

The above discussion illustrates that the ACF of a harmonic process plus white noise has a deterministic behavior. That is why in this work the ACF of variables is utilized for further analysis instead of the original data. While the original variables have random behavior due to the effect of white noise, the shape of their ACF can be easily determined with the knowledge of their respective frequency and one of the peak values.

The peak values in the ACF can be obtained following our previous work in [38] where estimation of the oscillation period was performed by detecting the peaks in the ACF. Detection of the peaks was performed based on clustering the ACF around the time lags of the peaks. The oscillation period can be estimated as the average of the time difference between adjacent peaks in the ACF.

## 3.5 Diagnosis

This section discusses possible methods for distinguishing controller induced oscillations from harmonics. As was mentioned in Section 3.3.1, one difference between

these two oscillation types is the varying amplitude and phase of controller induced oscillations. This property can be visually verified in the time domain. However, the difficulty here is the effect of noise on both types of oscillations. Harmonic processes with added white noise also have a varying amplitude in the time domain due to the effect of noise. Thus, the time series plot of a harmonic process with added noise will be similar to the plot of an oscillation due to the controller tuning in terms of the varying amplitude. That is why these two different types of oscillations are not distinguishable in the time domain.

The other difference between these two types of oscillations is the difference in the ACF as described in Sections 3.3.2 and 3.4. Figure 3.4 shows two oscillatory signals of same frequency.  $y_2(t)$  is the output of a loop consisting of a linear process model as  $\frac{1}{z^2+0.6z+0.05}$  controlled with a PI controller as  $1.48 + \frac{1}{z-1}$  in the presence of stochastic disturbance.  $y_3(t)$  is a single harmonic plus white noise as  $y_3(t) = 0.5\sin(2\pi/3.5) + \varepsilon_t$ . As can be seen in Figure 3.4, in the time domain plot they are similar to each other. However, the ACF of the sinusoidal signal is constantly oscillating for all the time lags, while the ACF of the other signal fades away after the first few lags.

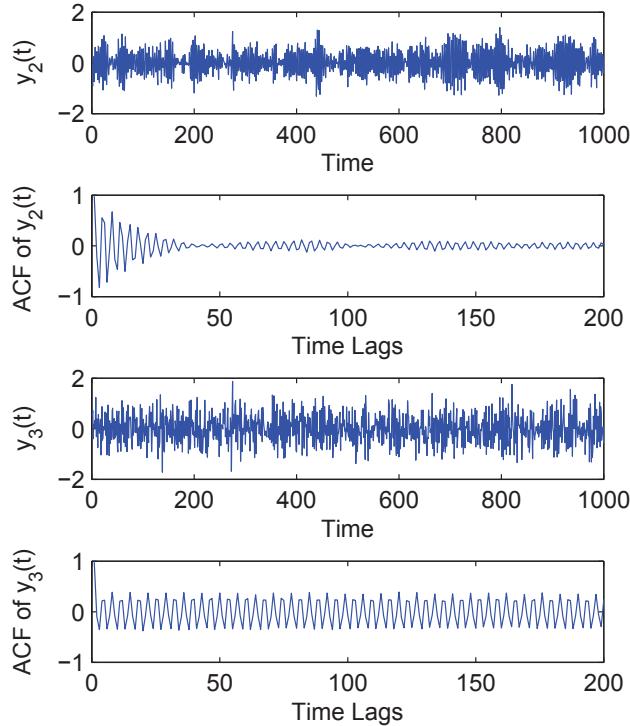


Figure 3.4:  $y_2(t)$  and  $y_3(t)$  in time along with their ACFs

This difference in the ACF makes it possible to diagnose the type of the oscillation by examining the tail of the ACF. If the ACF after the first few lags is still oscillatory with almost constant amplitude, then the oscillation is due to a harmonic process and not the controller tuning.

One issue remains here for the controller induced oscillations when the decay ratio is close to 1. As was mentioned, the ACF of the response of linear systems calculated from data does not converge to zero even for large time lags but continues oscillating with small amplitudes. The smaller the decay ratio, the smaller the amplitude will be. The oscillation amplitude for large decay ratios will be larger. An example of this case is plotted in Figure 3.5 where the decay ratio equals to 0.995 ( $y_4(t) = \frac{1}{1-0.3z^{-1}+0.99z^{-2}}\varepsilon_t$ ). This is the disadvantage of examining the tail of the ACF to diagnose oscillations.

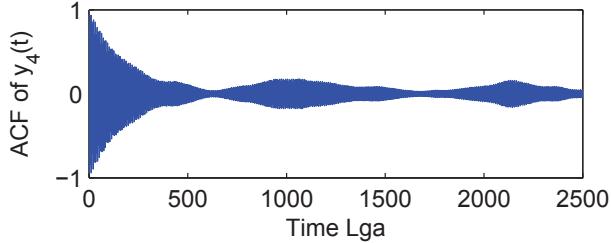


Figure 3.5: ACF of  $y_4(t)$

There is another difference in the ACF between the harmonic processes and oscillations due to the controller tuning. The oscillation due to controller tuning has a varying amplitude in the estimated ACF. This fact can be observed in Figures 3.5 and 3.1. Therefore, a method based on diagnosis of the oscillation with varying amplitude in the ACF is capable of distinguishing these two types of oscillations.

An automatic and reliable method is required to check for variations in the amplitude. One possible method is based on direct examination of the peak values in the ACF corresponding to the oscillation. If the peak values are constant, then it can be concluded that the oscillation is due to a harmonic process. Otherwise, the oscillation is caused by controller tuning.

### 3.5.1 Automatic method for diagnosis of oscillation

The previous section showed that the ACF of a harmonic signal is deterministic while the ACF of a controller induced oscillation is stochastic. A method is required for automatic distinction between these two oscillation types.

A hypothesis test needs to be developed in order to diagnose the type of the

oscillation by examining the ACF. Development of a hypothesis test requires a test statistic and the knowledge of the probability distribution of the test statistic under the null hypothesis. The null hypothesis is that the oscillation is a harmonic process and not due to controller tuning.

Here, instead of examining ACF values at all time lags, we only consider the peak values. The test will be conducted by comparing the differences between peak values in the ACF from their confidence interval. Assuming there are  $k$  peaks in the ACF, there will be  $\frac{k*(k-1)}{2}$  distinct pairs of peaks in the ACF (it should be noted that the peak at 0 time lag should not be considered in the analysis). Assuming a harmonic process, the difference between any two peak values in the ACF of harmonics is theoretically zero. However, there are random differences between estimated peak values (disregarding the effect of the linear decay inserted due to the estimation for now).

In order to test the differences between the peak values, we need to first detect the peaks and second, compensate the difference between the peaks due to the decay caused by the estimation method as was described in Section 3.4. To this end, the value of the peak at the larger time lag ( $p_2$ ) is compared to the adjusted value ( $p_1^a$ ) of the peak at the smaller time lag ( $p_1$ ).  $p_1^a$  is obtained as in Equation 3.20.

$$p_1^a = p_1 \frac{N - \tau_{p_2}}{N - \tau_{p_1}} \quad (3.20)$$

$p_2 - p_1^a$  will no longer be affected by estimation induced decay.

We can now determine the confidence interval of the difference between two ACF peaks. A confidence interval can be derived based on the variance of the estimated ACF. It is known that the ACF of the response of a general linear process follows normal distribution considering each time lag individually [42]. The variance of the distribution depends on the true values of the ACF. However, it is known that the ACF of white noise has expected value of 0 and variance equal to  $\frac{1}{N}$  where  $N$  is the number of data samples [41].

Considering a harmonic process with added white noise ( $y_t = x_t + \varepsilon_t$  where  $x_t$  is the harmonic part),  $x_t$  has a deterministic ACF while  $\varepsilon_t$  has ACF values distributed as  $\mathcal{N}(0, 1/N)$ . ACF of  $y_t$  which is a weighted summation of the two ACFs, is approximately distributed as  $\mathcal{N}(\mu, 1/N)$  where  $\mu$  is the ACF value of  $x_t$ . The true variance of ACF of  $y_t$  is smaller than  $\frac{1}{N}$  because the ACF of  $\varepsilon_t$  is multiplied by  $\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_x^2}$  in order to obtain the ACF of  $y_t$ . This makes the variance of ACF of  $y_t$  to be equal to  $1/N \times (\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_x^2})^2$  which is smaller than  $\frac{1}{N}$ . In the presence of small signal to noise ratio, i.e.  $\sigma_x^2 \ll \sigma_\varepsilon^2$ , we can assume ACF of  $y_t$  is approximately distributed as  $\mathcal{N}(\mu, 1/N)$ .

If this assumption does not hold, methods for estimating the signal to noise ratio will be provided in the next section.

The difference between two nominal peaks in the ACF, after adjusting the value of the peak at smaller time lag ( $p_2 - p_1^a$ ), is the subtraction of two normally distributed random variables (assuming the oscillation is a harmonic process). Thus the distribution of the differences between peak values in the ACF of a harmonic process is as in Equation 3.21.

$$p_2 - p_1^a \sim \mathcal{N}\left(0, \frac{1}{N} \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_x^2}\right)^2 \left(1 + \left(\frac{N - \tau_{p_2}}{N - \tau_{p_1}}\right)^2\right)\right) \quad (3.21)$$

where  $p_2$  peak is selected far from the initial peak  $p_1$  so the two variables can be assumed independent.

If an estimation of the factor  $\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_x^2}$  is available as will be discussed shortly, the true value of the confidence interval can be obtained. Otherwise, an approximate confidence interval equal to  $\frac{3\sqrt{2}}{\sqrt{N}}$  can be used in order to check if the differences between peak values in the ACF are significant. For example, Figure 3.6 plots the histogram of the differences between the peak values in the ACF of  $x_t = \sin(\frac{2\pi}{10}t) + \varepsilon_t$  where variance of  $\varepsilon_t$  equals to 1. The confidence interval obtained based on Equation 3.21 equals to 0.063 and as can be seen in Figure 3.6 the difference between no two peaks in the ACF exceeds the significance level. The approximate confidence interval  $(\frac{3\sqrt{2}}{\sqrt{N}})$  equals to 0.094.

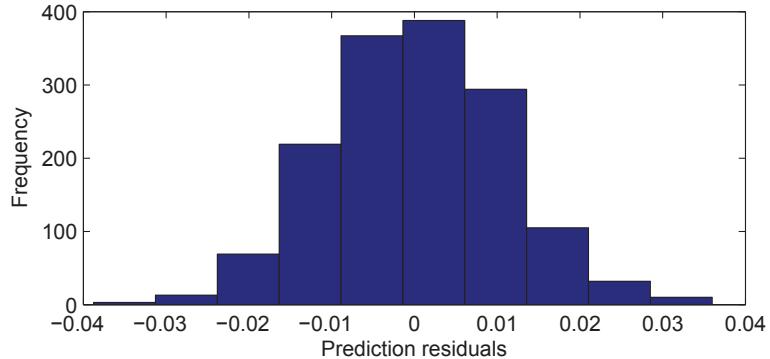


Figure 3.6: Distribution of the differences between the peak values in the ACF of  $x_t = \sin(\frac{2\pi}{10}t) + \varepsilon_t$

### 3.5.2 Effect of added colored noise

An investigation of the effect of colored noise on both types of oscillations is required. ACF has the advantage of filtering white noise from the data. If the noise is white,

there will be no effect since the noise does not appear in the ACF of any signal except at 0 time lag. However, we need to consider the effect of added colored noise on the ACF of both types of oscillations.

Colored noise is the output of processes such as an ARMA process with white noise as its input and therefore has all the properties mentioned in Section 3.3. If the added colored noise is also oscillatory (i.e. the noise process has complex roots), we need to filter the signal in order to decompose it into its individual oscillatory components. Readers are referred to [23] for more information on the filter design and the method for dealing with the presence of multiple oscillations in the data. Here, it is assumed that the noise is not oscillatory.

A signal with added colored noise can be written as in Equation 3.22 where  $x_t$  represents the oscillatory part of the signal (could be a harmonic or generated from an ARMA process) and  $w_t$  represents the added colored noise. Since  $x_t$  is independent from  $w_t$ , the autocovariance function of  $y_t$  ( $R_y(\tau)$ ) is the summation of the autocovariances of the two independent parts as shown in Equation 3.23. Therefore, the ACF of  $y_t$  is also a weighted summation of two independent functions.

$$y_t = x_t + w_t \quad (3.22)$$

$$R_y(\tau) = R_x(\tau) + R_w(\tau) \quad (3.23)$$

$w_t$  has a fading ACF which can be represented as  $\alpha^\tau$  where  $\alpha$  is the largest pole of the noise transfer function. Therefore, its effect is present in the smaller time lags till the value of  $\alpha^\tau$  is negligible.

The fading ACF of colored noise disturbs the ACF of the oscillatory part of the variable ( $x_t$ ). Therefore it is necessary to analyze the ACF after the transient behavior in the first few lags is faded and the ACF becomes monotone. The monotonous behavior in the case of a harmonic process means an almost constant oscillation amplitude which implies peaks with similar values. On the other hand, monotonous behavior in the case of oscillation due to controller tuning means that although the oscillation has varying amplitude, it does not show a constantly decaying behavior. The part of the ACF that satisfies this condition is called the tail of the ACF. The proposed hypothesis test should be performed on the tail of the ACF so that the result is not affected by the added colored noise in the data.

It should be noted that the non-stationary trend in the data can also disturb the ACF and make oscillation detection or diagnosis difficult. It is required to remove the non-stationary trends and abrupt changes from data by filtering [23] or other methods for a reliable oscillation detection or diagnosis based on ACF.

The developed diagnosis methodology also requires additional pre-processing of the data if there are multiple oscillations present in order to correctly diagnose each oscillatory component. If there are multiple oscillations, one should apply the approach such as [23] to separate the oscillations into individual ones first and then apply the proposed method to each oscillation for the diagnosis. Alternatively, one can also apply wavelet transform approach to separate oscillations. Proper wavelet transform of the ACF can deal with the presence of multiple oscillations, non-stationary trends in the data and noise [43].

### 3.5.3 Estimation of signal to noise ratio as additional test

By excluding the first few peaks in the ACF from the analysis in order to remove the possible effect of added colored noise in the data, the null hypothesis may not be rejected even when the oscillation is truly due to controller tuning. For example, consider a process model of  $\frac{e^{-2s}}{25s^2+6s+1}$  which is controlled by a PID of  $1 + \frac{0.1}{s} + \frac{50}{1+\frac{100}{s}}$  in a loop. There is white noise with variance of 0.2 added to the output as measurement noise. Figure 3.7 plots the output of the loop sampled every second and its ACF in the middle panel. The bottom panel shows the histogram of the differences between peak values in the ACF.

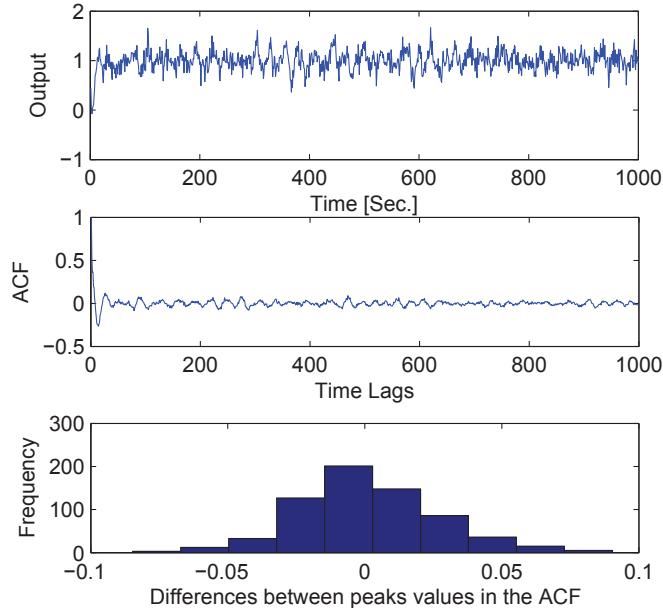


Figure 3.7: Top: output of the simulated continuous time loop, Middle: ACF of the output, Bottom: histogram of the difference between the peaks in the ACF

As can be seen in Figure 3.7, since the ACF decays very fast, the peaks in the ACF have very small values. The confidence interval for the differences between the peak values in the ACF equals to 0.056 based on Equation 3.21 while the approximate confidence interval is 0.094. By excluding the first few high value peaks in order to avoid the possible effect of added colored noise in the data, the differences between the remained peaks exceed the true confidence interval but are less than the approximate one. Considering that an estimation of the true confidence interval may not be always available and the approximate confidence interval might be used in practice, the null hypothesis may not be rejected which implies that the oscillation is a harmonic and not generated by a linear system. The reason for this problem is the fact that the peaks have very small values at the tail of the ACF and therefore their differences are also very small.

Correct diagnosis of the oscillation is possible by more analysis based on estimating the signal to noise ratio (SNR) from the ACF and comparing it to the SNR estimated from the power spectrum of the variable. Signal to noise ratio is defined as in Equation 3.24.

$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2} \quad (3.24)$$

$\sigma_{signal}^2$  is the variance of the oscillatory part of the variable ( $y_t = x_t + \varepsilon_t$ ) which will be noted as  $\sigma_x^2$ .  $\sigma_{noise}^2$  will be noted as  $\sigma_\varepsilon^2$  afterwards.

It is possible to estimate  $SNR$  based on the peak values in the ACF. The value of a peak in the ACF of a sinusoidal signal with added white noise can be obtained as in Equation 3.25 [38]:

$$\rho_{\tau_p} = \frac{\frac{N - \tau_p}{N} \sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \quad (3.25)$$

where  $\tau_p$  is the time lag of the peak and  $N$  is the sample size of the data. Thus, SNR can be estimated from the peak values in the ACF as in Equation 3.26.

$$\frac{\sigma_x^2}{\sigma_\varepsilon^2} = \frac{\frac{N}{N - \tau_p} \rho_{\tau_p}}{1 - \frac{N}{N - \tau_p} \rho_{\tau_p}} \quad (3.26)$$

For a more reliable estimation of SNR from the peak values in the ACF, an average of the peak values in the ACF should be used instead of only one peak value. For that purpose, the length of the ACF is divided by 3 and the peak values in the middle portion of the ACF are used. Each peak value is multiplied by  $\frac{N}{N - \tau_p}$  where  $\tau_p$  is its corresponding time lag. The average of the adjusted peak values ( $P$ ) is then used to estimate SNR as  $\frac{P}{1 - P}$ .

Another way to estimate SNR is based on utilizing the power spectrum of the variable ( $y_t$ ). Parsval's theorem states the relation between energy of the signal in the time domain and its power spectrum as in Equation 3.27.

$$\sum_{n=0}^{N-1} y_n^2 = \frac{1}{N} \sum_{n=0}^{N-1} |Y_k|^2 \quad (3.27)$$

where  $Y_k$  is the discrete Fourier transform of the variable  $y_n$ . Assuming that the data is detrended, its standard deviation equals to  $\frac{1}{N^2} \sum_{n=0}^{N-1} |Y_k|^2$  which is the summation of the power spectrum of the data divided by the sample size.

Estimating SNR from the power spectrum is possible through estimation of the oscillation frequency by an oscillation detection methodology. The oscillation detection method proposed in [38] provides an estimation of the oscillation period ( $T$ ) along with a standard deviation for the estimated period ( $\sigma_T$ ). Summation of the power spectrum of the variable in the frequency range of  $[1/(T + 2\sigma_T) : 1/(T - 2\sigma_T)]$  subtracted by the mean of the power spectrum at other frequencies, gives an estimation of the power of the oscillation in the data. Dividing this value by the summation of power spectrum at other frequencies gives an estimation of the signal to noise ratio.

If the oscillation is generated by the controller tuning, the estimated SNR from the ACF will be considerably less than the SNR estimated from the power spectrum. The reason is that the oscillation amplitudes in the tail of the ACF of this type of oscillatory signals are very small and theoretically should be 0. Therefore, the SNR estimated from the peak values in the tail of the ACF of this type of oscillations will be considerably less than the true SNR value. However, if the oscillation is a harmonic process, the estimated SNR based on the ACF will be close to the SNR estimated based on the power spectrum.

For example, the SNR estimated based on the ACF of the response of the simulated control loop plotted in Figure 3.7 equals to 0.0576 which is considerably smaller than the SNR estimated from the power spectrum of the variable which is 0.775. The SNR estimated from the power spectrum of a sinusoidal signal with added colored noise as  $\sin(\frac{2\pi}{24}t) + \frac{1}{z^{-2} - 0.9z^{-1} + 0.18}\varepsilon_t$  (varince of  $\varepsilon_t$  is 1), equals to 0.29 while the SNR estimated from its ACF equals to 0.25. The true value of SNR for this variable is 0.2. We can observe that the SNR of a sinusoidal signal with added noise estimated from ACF is close to its value estimated from power spectrum while it is not the same for the oscillation due to a controller tuning.

### 3.6 Pure imaginary poles

So far the assumption was that the poles of the transfer function of the loop are complex but not pure imaginary. The reason for this assumption was that pure imaginary poles rarely happen in practice and behave similar to unstable poles. This subsection studies the ACF properties of the response of a system with pure imaginary poles.

Pure imaginary poles cause a consistent oscillation in the step response and a consistent oscillation with varying amplitude when responding to noise. An example of this case is plotted in Figure 3.8 which shows the step response of a loop consisting of a process model of  $\frac{3e^{-2s}}{34s^2+17s+2}$  controlled by a PID controller of  $3.2734 + \frac{0.2}{s} + \frac{80}{1+\frac{100}{s}}$ .

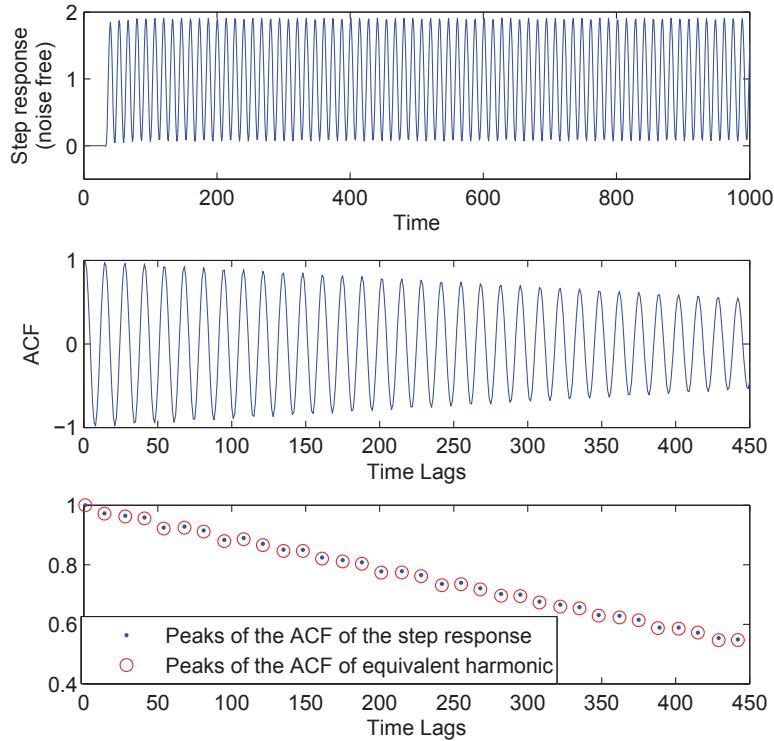


Figure 3.8: Top panel: step response of the simulated loop, Middle panel: ACF of the step response, Bottom panel: peaks of the ACF of the response and of the equivalent harmonic

The response to a noise free step input is persistently oscillating with constant amplitude. The ACF of this response is expected to be the same as the ACF of a pure sinusoidal signal which has the same frequency and sample size. This can

be examined by generating a sinusoidal signal with the same frequency as the step response case with the same sample size. Bottom panel of Figure 3.8 shows the peaks of the ACF of the step response and the peaks of the ACF of the equivalent sinusoidal signal. The peak values are the same for all the time lags as is expected.

Figure 3.9 plots the response of the same loop to the step input when white noise with variance equal to 0.1 is added to the input. The peaks of the ACF of this response along with the peaks of the ACF of a harmonic with the same frequency and sample size are plotted in the bottom panel. As can be seen in Figure 3.9 the decay in the ACF corresponding to the response of the control loop is larger compared to the decay in the ACF of the equivalent harmonic. The reason for this could be the fact that the amplitude of the oscillation changes with time in the time domain due to the varying noise values. Even though the ACF is expected to behave as the ACF of a harmonic process, the estimated ACF has a larger decay because of the less predictable behavior of the oscillation when there is noise in the environment.

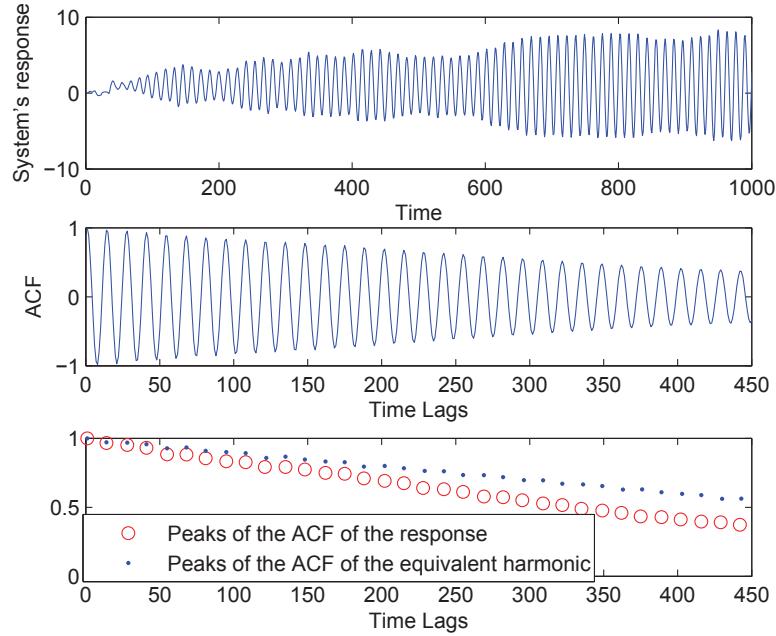


Figure 3.9: Top panel: response to noise of the simulated loop, Middle panel: ACF of the response, Bottom panel: peaks of the ACF of the response and of the equivalent harmonic

Considering the fact that there is always noise in the environment, the larger decay rate in the ACF of the response of a marginally stable loop to noisy input compared to the ACF of its equivalent harmonic process can be utilized as a criterion to distinguish between these two types of oscillations. When the result of the introduced hypothesis

tests agrees with the oscillation being caused by harmonics, the decay rate of the ACF can be checked again to ensure that the oscillation is not due to pure imaginary poles. This test can be performed by simulating a sinusoidal signal with the same frequency as the system's response and the same sample size with added white noise. The variance of the noise should be equal to the variance of the actual noise in the data which can be estimated as proposed in Section 3.5.3. The actual ACF can be compared with the ACF of simulated signal.

One note regarding this issue is when a step test is done on the system to determine the ultimate frequency of the loop in order to design controllers. In the noise free environment, the step response is constantly oscillatory only if the poles are pure imaginary. For example, the simulated loop in Section 3.3.3 with the mentioned PID controller has a constant oscillation in its step response when the proportional gain is increased to 3.178. However, when there is white noise with variance of 0.02 added to the unit step input, the response of the loop has a persistent oscillation even for smaller gain values. For example, the step response of the loop with added white noise considering a proportional gain of 2.6 for the controller is plotted in Figure 3.10.

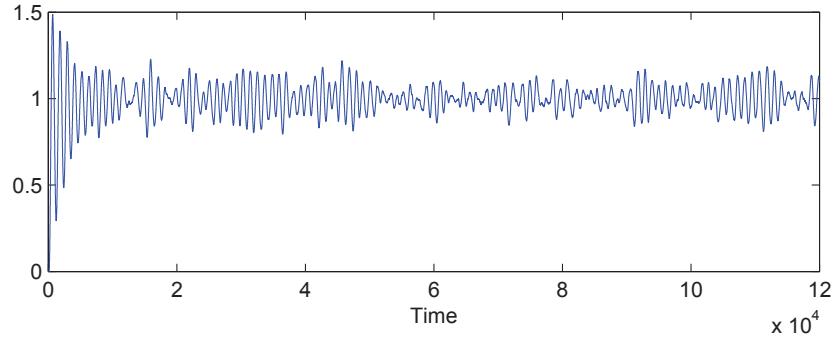


Figure 3.10: Step response of the continuous control loop introduced in Section 3.3.3 under noisy environment when the gain of the controller is 2.6

The conclusion of this subsection is that increasing the gain of the controller until the step response of the loop has persistent oscillation in an industrial environment does not necessarily mean that the loop is already marginally unstable at that condition. Depending on the noise in the environment, the loop can show persistent oscillation in the response when the poles are far from being pure imaginary. Therefore, it is important to check the ACF of the loop's response when trying to find the ultimate frequency and gain of the loop for the controller design purposes.

### 3.7 Case study

A multiple tank system as shown in Figure 3.11 is used in order to verify the results of this study. The equipment is composed of three tanks and the level of the middle tank is controlled by a PID controller in a loop. After reaching to a steady state operation, the gain of the controller is increased until the output is oscillatory. The level of the tank under control (after detrending) and its ACF are plotted in Figure 3.12.

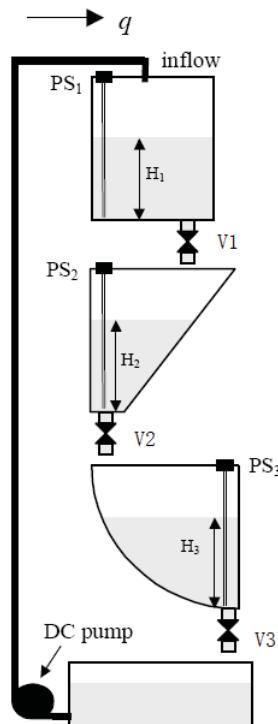


Figure 3.11: Schematic of the multi-tank system under study

As can be seen in Figure 3.12, the ACF shows the same characteristics as discussed in the previous sections. Figure 3.13 plots the prediction residuals of the ACF using Equation 3.19 and taking the value of  $23_{rd}$  peak in the ACF as the reference. As can be seen in Figure 3.13, prediction residuals exceed the significance level (0.09) at several time lags which results in a rejection of the null hypothesis.

To observe the difference of the response of this system to oscillatory external disturbances, an external oscillation of  $0.01\sin(2\pi/160t)$  is inserted into the loop after re-tuning the controller to return to a non-oscillatory steady state. The level measurement and its ACF are plotted in Figure 3.14. The ACF of the level measurement in this case shows a pure sinusoidal signal with a linear decay as expected.

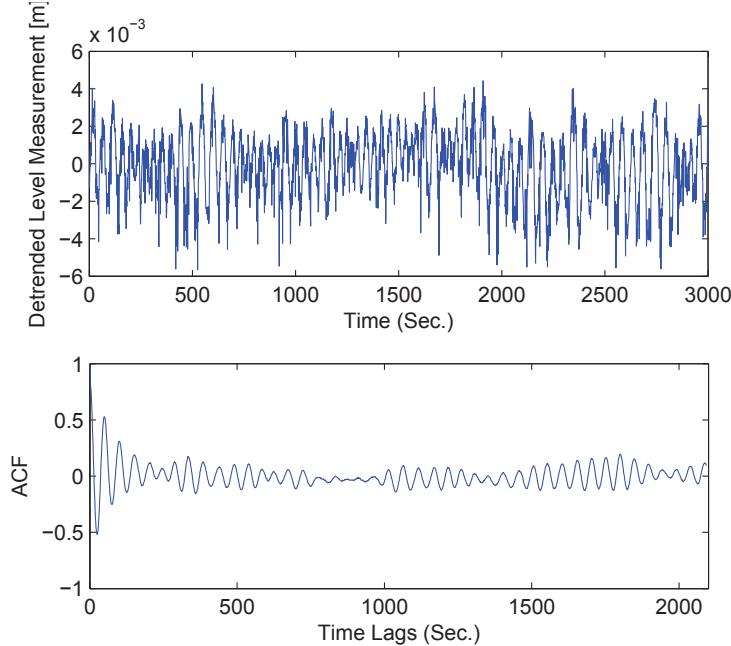


Figure 3.12: Level measurement of the tank under study and its ACF when the loop is oscillatory due to the controller

Figure 3.15 plots the residuals of predicting ACF of the response to sinusoidal disturbance based on the value of its third peak. In contrary to Figure 3.13, the residuals are far less than the significance level (0.09) and the null hypothesis is not rejected. Again this verifies our methods.

External harmonic disturbances in industrial processes could happen due to an oscillation originated by a valve stiction which becomes more linear when propagates through the plant. For example Figure 3.16 top panel shows the normalized output of an oscillatory loop in an industrial process which will be called loop 1. Figure 3.16 bottom panel plots the ACF of the output of loop 1. The histogram of the differences between the peak values of this ACF is plotted in Figure 3.17 which has an approximate normal shape. The confidence limit estimated for the differences between the peak values is 0.07. It is observed that there is no sample in the histogram exceeding the confidence limit. Performing the second test based on estimation of SNR results in the SNR value of 0.14 estimated from the ACF which is close to SNR value of 0.16 estimated from the power spectrum of the signal. Based on these tests, the oscillation in loop 1 is considered to be due to an external disturbance.

Another loop in the same process is also oscillatory with similar frequency as plotted in Figure 3.18 which will be called loop 2. Since loop 2 has a nonlinear oscillation, it can be concluded that the root cause of the oscillation is a sticky valve

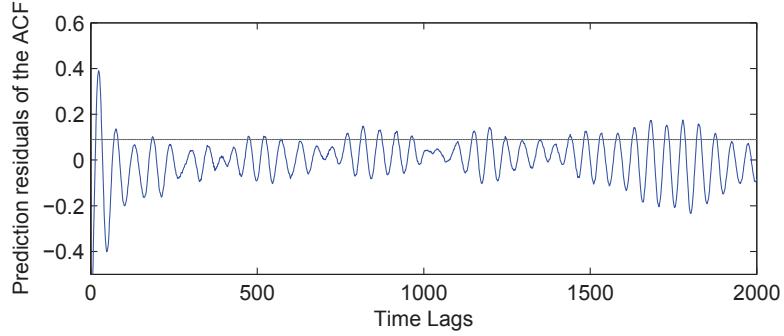


Figure 3.13: Prediction residuals of the ACF plotted in Figure 3.12 and its significance level

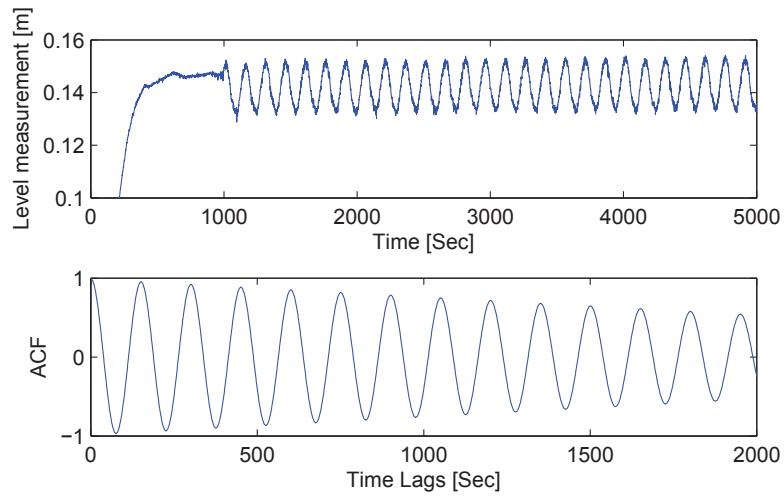


Figure 3.14: Level measurement of the tank under study and its ACF using a stable controller in the loop with added oscillatory disturbance

in loop 2 and loop 1 is disturbed through loop 2.

### 3.8 Summary

This chapter studied the properties of the responses of linear systems to stochastic disturbances and also the properties of the auto correlation functions of the responses. It was shown that the oscillation generated by a loop due to controller tuning has different ACF properties compared to a harmonic oscillation, which can be utilized to distinguish these two types of oscillations. Two hypothesis tests were developed for automatic diagnosis of oscillations in feedback loops. The result is valuable in the sense that distinguishing between oscillations caused by poor controller tuning from external harmonics helps in deciding the appropriate trouble shooting procedure.

The disadvantage of this method is in the case of existence of more than one oscil-

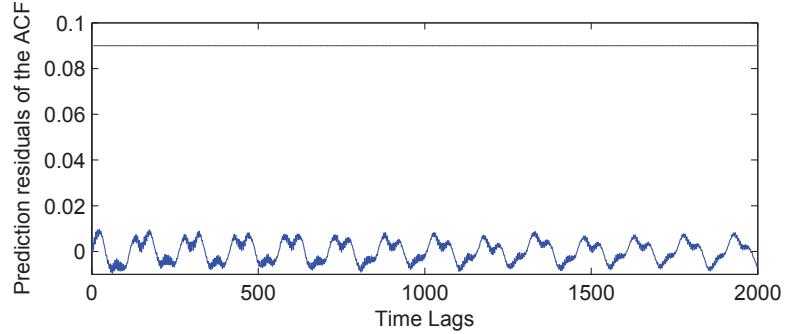


Figure 3.15: Prediction residuals of the ACF plotted in Figure 3.14 along with the significance level

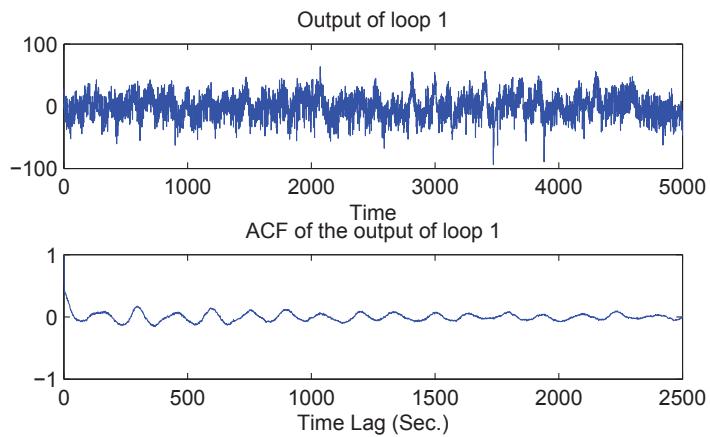


Figure 3.16: The normalized output of an oscillatory loop in an industrial process and its ACF

lation in the signal. The varying amplitude of the oscillation caused by linear systems due to controller tuning makes it difficult to reliably differentiate between the peaks in the ACF corresponding to different oscillations. Therefore, this method can be reliably utilized if there is only one oscillation present in the data. Another methodology based on wavelet transform which is capable of handling multiple oscillations and non-stationary trends will be proposed in the next chapter.

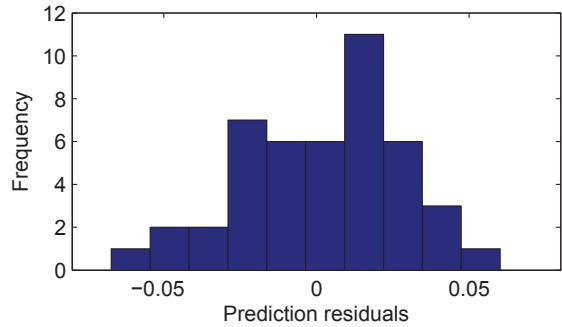


Figure 3.17: Prediction residuals of the ACF plotted in Figure 3.16

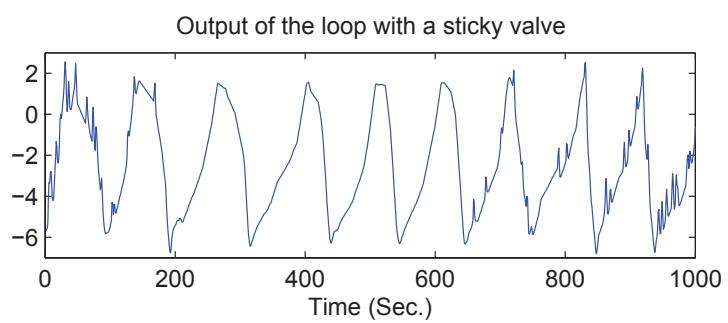


Figure 3.18: The output of a loop suffering from valve stiction

## **Chapter 4**

**Wavelet transform based  
methodology for detection and  
diagnosis of multiple oscillations in  
non-stationary variables**

## 4.1 Abstract

The assumption of data being stationary is the requirement of many statistical analysis procedures. However, variables recorded in an industrial environment do not usually satisfy this assumption. Non-stationary trends, temporal behavior and abrupt changes are very common in industrial data which make it difficult to detect and diagnose abnormalities. In order to deal with these issues, this chapter proposes using wavelet transform as a powerful tool for analysis of non-stationary data, specifically for the purpose of oscillation detection and diagnosis.

Detection of multiple oscillations in variables has been studied in literature while diagnosis of individual oscillatory components is yet not resolved. This chapter proposes a methodology based on wavelet transform which is capable of both detection and diagnosis of multiple oscillations in variables. Wavelet transform provides a time dependent, multi-resolution decomposition of signals which facilitates handling non-stationary trends, analyzing intermittent oscillations and dealing with presence of multiple oscillations.

Some hypothesis tests are proposed in order to diagnose if the source of oscillation is poor controller tuning, valve nonlinearity or external oscillatory disturbance. The tests are mainly based on applications of wavelet power spectrum and wavelet bicoherence. Wavelet power spectrum is utilized to diagnose oscillations between controller tuning and sinusoidal disturbances. Wavelet bicoherence, similar to the classical bicoherence, can detect and quantify presence of nonlinearity in variables and therefore is an appropriate tool to diagnose if the source of the oscillation is nonlinearity in the process. The proposed methodology is capable of individual diagnosis of different oscillatory components of the variable. Advantages of the proposed method is illustrated through analysis of data sampled from an industrial process.

## 4.2 Introduction

Wavelet transform facilitates individual diagnosis of multiple oscillations due to its intrinsic nature in decomposing the variables. Wavelet transform is extensively applied in signal and image processing especially for the purpose of noise filtering and data compression. It is also applied for fault detection and diagnosis in engineering applications. Peng et al. [44] reviewed the application of wavelet in machine condition monitoring. Yang et al. [45] proposed a fault detection algorithm based on the use of wavelet transform for noise filtering and pre-processing of data followed by kernel entropy component analysis. Alabi et al. [46] also integrated wavelet transform with

generic dissimilarity measure to achieve a better performance monitoring tool. Zhu et al. [47] presents a review on the literature regarding monitoring tool conditions based on wavelet transform. Bakshi [48] integrated wavelet transform with PCA in order to monitor the slow and fast features of the data individually along with better noise removal. Lin et al. [49] proposed a wavelet denoising method in order to extract the weak features in variables for the purpose of fault diagnosis.

Wavelet transform has also been utilized for the purpose of oscillation detection. Matsu et al. [50] proposed visual inspection of wavelet transform of the variables as a method to detect oscillations in noisy environment. Guo et al. [51] used wavelet packet transform for detection of multiple non-stationary oscillations in process variables. Plett [52] proposed a hypothesis test based on using cross wavelet transform and coherence in order to detect transient oscillations in variables. Diagnosis of multiple oscillations is less studied in literature. Karra et al. [33] proposed an algorithm for detection as well as diagnosis of multiple oscillations in variables based on combination of several methods and system identification procedures. This work presents a comprehensive algorithm which is capable of both detection and independent diagnosis of multiple oscillations in the presence of noise and nonstationary trends in variables based on wavelet transform. The idea of using wavelet transform for diagnosis of oscillations between controller tuning and external disturbances was initially introduced in [53] which is extended in this work.

Diagnosis of oscillations is performed based on the different properties of oscillations caused due to valve nonlinearity, controller tuning or sinusoidal disturbances. There are several methods for diagnosis of oscillations due to valve stiction. Choudhury et al. [31] reviewed the methods in literature for valve stiction detection and diagnosis. One of the methods which is only based on analyzing the operational data is to examine bicoherence of the variable. Bicoherence is used in order to detect if there is nonlinearity in the variable and therefore, it can diagnose the oscillations due to valve nonlinearity. Diagnosis between oscillations due to controller tuning and oscillations due to other sources is also studied in literature. Naghooisi et al. [54] proposed a methodology to diagnose oscillation caused by controller tuning based on examining the auto correlation function (ACF) of the variables. The method proposed in [54] works with the assumption of a single oscillation frequency in the variable. The current work uses similar concepts in the wavelet domain in order to achieve a comprehensive methodology which not only is capable of detection of multiple oscillations in noisy and nonstationary data sets, but is also capable of diagnosis of the individual oscillatory components.

The chapter is organized as follows. First, a preliminary introduction to wavelet

transform is given. Wavelet power spectrum and its properties are reviewed in Section 4.4 followed by the methodology for diagnosis of oscillations between controller tuning and external disturbances based on using wavelet power spectrum. Section 4.5 reviews the concepts of coherence, bicoherence and wavelet bicoherence. The method to use wavelet bicoherence for the purpose of detecting nonlinear oscillations is presented in 4.6. Section 4.7 presents a case study and Section 4.8 summarizes the chapter.

### 4.3 Introduction to Wavelet transform

Since wavelet transform is considered as an advancement over Fourier transform, the properties of Fourier transform are briefly reviewed first. Fourier transform of a signal  $x(t)$  with the following definition is a widely applied tool to analyze stationary signals for different purposes.

$$f(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (4.1)$$

The underlying assumption of Fourier analysis is that the signal has constant spectral properties over the whole time period. In definition of Fourier transform the integration (or the average in discrete Fourier transform) is taken over the whole sample time. Therefore it provides an average representation of the signal independent of time. For non-stationary signals, whose spectral properties change with time, the information regarding the temporal behavior is lost in Fourier analysis.

An improvement over Fourier transform to make it more suitable for non-stationary signals is called windowed or short-time Fourier transform (STFT). In STFT a window function concentrated on a desired point in time  $b$  is multiplied to the signal and then the Fourier transform is taken as shown in Equation 4.2. The window is designed such that it decays with time. Therefore, its multiplication to the signal damps the contribution of the samples that are far away from the desired point in time leading to a localized representation of the signal.

$$STFT(b, \omega) = \int_{-\infty}^{\infty} g(t-b)x(t)e^{-j\omega t} dt \quad (4.2)$$

The window function  $g(t)$  can be of different types such as exponential  $g(t) = \exp(-\frac{t^2}{2\sigma^2})$ . Selection of  $\sigma$  which is the width of the function changes the time and frequency resolution. A small value for  $\sigma$  shortens the width of the window leading to good time resolution for the high frequency components. However, for the low frequency components the length of the window might be too short leading to poor frequency resolution for these components. On the other hand, a wide window provides a good resolution for low frequency components but poor time resolution

for high frequency components. Therefore, the trade-off between time and frequency resolution is the problem of STFT.

Wavelet transform (WT) provides a solution for the problem of time-frequency resolution. In WT, the width of the window which is called scale is allowed to change based on the frequency currently being analyzed. Higher frequency components are analyzed by windows with shorter width while wider windows are used to analyze low frequency features of the signal. Windows with a variety of widths centered at one sample time are used in WT in order to provide good frequency resolution over the whole frequency range of interest. Scale in wavelet domain is a varying parameter to be adjusted to the frequency of interest opposite to STFT where the window width is constant. Wavelet transform was first introduced by [55] and then extensively developed and applied in different fields such as Mathematics, Engineering, Econometric etc..

Continuous wavelet transform (CWT) of a signal  $x(t)$  at time  $b$  and scale  $a$  ( $w_x(b, a)$ ) is defined as

$$w_x(b, a) = \int_{-\infty}^{+\infty} x(t) \Psi_{a,b}^*(t) dt \quad (4.3)$$

The wavelet function  $\Psi_{a,b}$  is a scaled and translated version of mother wavelet function  $\Psi$  as

$$\Psi_{a,b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (4.4)$$

Mother wavelet function  $\Psi$  is usually chosen as a decaying oscillatory function. Changes in the scale  $a$  and time  $b$  provide three dimensional representation of the signal in time-frequency domain. The values of wavelet coefficients are a function position in the time domain, scale and the choice of mother wavelet function in addition to the signal itself. The relation between scale and frequency is approximately given by

$$f_a = \frac{f_c}{a\Delta} \quad (4.5)$$

where  $f_c$  is the central frequency of the wavelet function,  $\Delta$  is the sampling period and  $a$  is the scale.

As an example of mother wavelet function, Morlet wavelet function is plotted in Figure 4.1. Morlet wavelet function is constructed as multiplication of a sinusoidal of frequency  $f$  to a Gaussian weighting function as in the following equation.

$$\Psi(t) = \pi^{-1/4} \exp(i\omega_c t) \exp(-t^2/2) \quad (4.6)$$

CWT coefficients are actually the inner product of the signal with shifted and scaled versions of the mother wavelet function. The wavelet coefficients quantify the

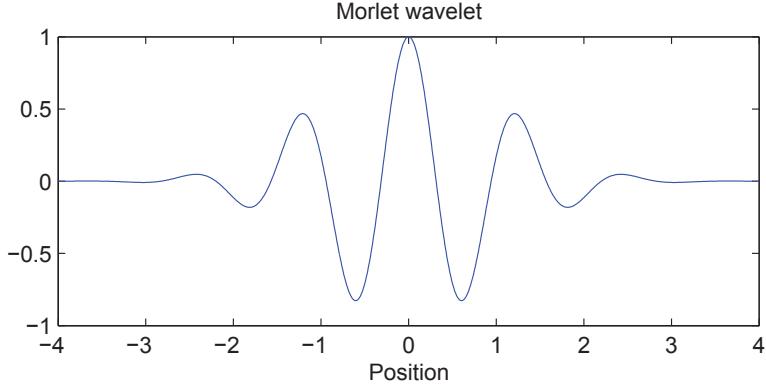


Figure 4.1: Morlet wavelet function with central frequency of 0.8125 Hz

similarity between the signal and the wavelet function in the same way that Fourier transform measures the similarity between the signal and sinusoidal waves of different frequencies. If the similarity is high, the absolute value of the coefficients will be high which usually happens if the scale of the wavelet function is close to the period of the signal's oscillation. Lower scales measure the energy of the signal at higher frequencies while larger scales measure the signal's energy at lower frequencies.

Smaller scales compress the wavelet function and can provide a good time resolution for the high frequency components. Larger scales stretch the wavelet function more and therefore better capture the coarser features of the signal and represent the energy of the signal at low frequencies. Therefore, changing the scale makes it possible to study the evolution of fast changing features of the data in time as well as its long term behavior.

The choice of the mother wavelet function can be done subjectively depending on which feature of the data is of interest to be analyzed. The mother wavelet function can be any function which satisfies the wavelet admissibility condition as:

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\widehat{\Psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (4.7)$$

where

$$\widehat{\Psi}(\omega) = \int \Psi(t) \exp(-j\omega t) dt \quad (4.8)$$

The standard mother wavelet functions are usually oscillatory functions which decay with time. Therefore, wavelet transform at a time sample presents the contribution of different frequencies to the signal similar to Fourier transform. However, the decay in the wavelet function preserves the presentation to a short time interval. If the purpose of wavelet transform is to study the oscillatory components of the signals, it is better to choose smooth oscillatory wavelet functions. Since this work

is mainly concerned with using wavelet transform for oscillation diagnosis, Morlet mother function is selected.

Continuous wavelet transform (CWT) has difficulties in estimation due to the continuous change over both time and frequency domains and thus it is usually estimated at finite number of time samples and scales. Discretized CWT of a signal is obtained as shown in Equation 4.9.

$$w_x(b, a) = \sum_{n=0}^{N-1} x_n \left( \frac{\Delta}{a} \right)^{1/2} \Psi^* \left[ \frac{(n - b)\delta t}{a} \right] \quad (4.9)$$

where  $a$  is the scaling factor and  $b$  is the translation factor. Changing  $a$  between 0 to  $N - 1$  changes the width of the wavelet function and therefore provides different frequency resolution.  $b$  varies between 1 to  $N$  (number of data samples), moving the wavelet function in time horizon and providing a time dependent representation of the signal.

If the scales and the time positions are chosen as  $a = 2^j$ ,  $b = k2^j$  where  $k, j \in z$ , the wavelet transform can be estimated in a faster way and is called Discrete Wavelet Transform (DWT). In CWT the wavelet functions are not orthogonal and therefore the transform contains redundant information. In DWT the wavelet functions are required to be orthogonal providing a more compact transformation of the signal. The disadvantage of DWT is that the number of scales and time points chosen for the transformation are few for the purpose of this work. Therefore, CWT is preferred in this work as well as the literature in similar subjects since it provides the flexibility of choosing all the scales of interest and studying the inter-scale relations.

### 4.3.1 Interpreting the wavelet coefficients

This section provides some explanation on how the wavelet coefficients can be interpreted. Consider a variable constructed as  $x(t) = \sin(\frac{2\pi}{20}t) + u(t - 250)\sin(\frac{2\pi}{35}t)$  where  $u(t)$  is the unit step function. The variable is plotted in top panel of Figure 4.2 which shows that it consists of a persistent oscillation with period of 20 with an additional oscillation with period of 35 from sample 250 onward.

The wavelet coefficients at the scale of 16, which is the closest scale to oscillation period of 20, calculated using Morlet wavelet function, are plotted in the middle panel. The coefficients show a persistent oscillation during the whole time interval as is expected. Wavelet coefficients at scale of 28, approximately corresponding to oscillation period of 35, are also plotted at the bottom panel. The coefficients are very close to 0 approximately up to the time sample of 250 when the oscillation actually appears in the data. The wavelet coefficients correctly indicate the presence of

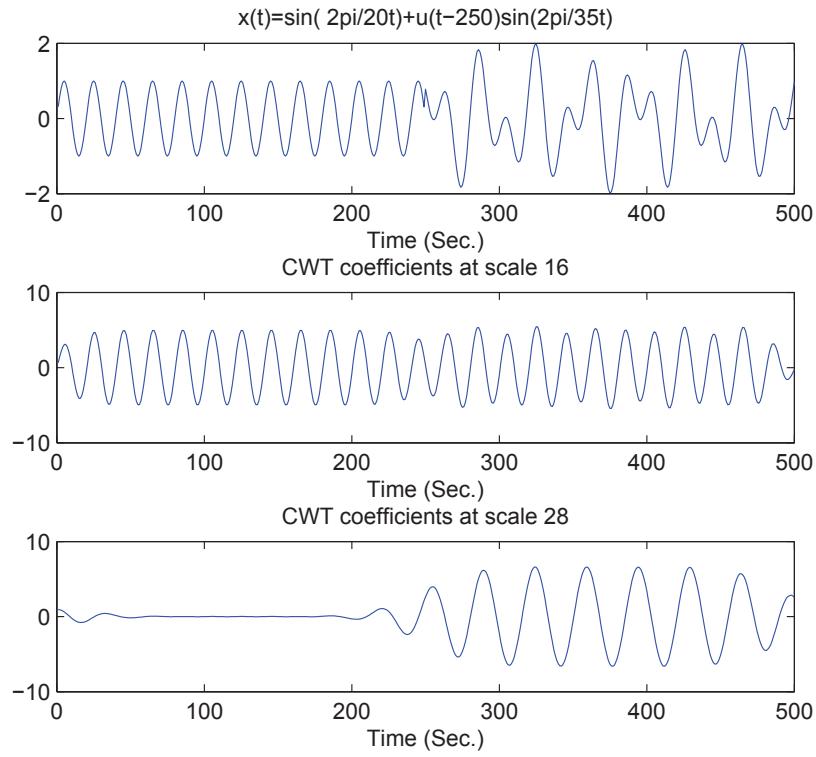


Figure 4.2: Top:  $x(t)$ . Middle: Wavelet coefficients at a scale approximately corresponding to oscillation period of 20. Bottom: Wavelet coefficients at a scale approximately corresponding to oscillation period of 35

oscillation with period of 35 for the second half of the whole period. The fluctuations at the beginning and the end of the time interval are due to a phenomenon explained as cone of influence in the literature.

Usually the wavelet coefficients are estimated for a range of scales and the absolute value of coefficients is color coded in a figure as shown in Figure 4.3 which shows the wavelet transform of  $x(t)$  for scales from 1 to 50. The scales close to 16 and 28 where the signal actually has its highest energy have brighter colors which indicate larger wavelet coefficients. The map of wavelet coefficients shows which scales or frequencies have the highest contribution to the signal and at which time samples.

Now consider the wavelet transform of a step function as plotted at the top panel of Figure 4.4 which is shown at the bottom panel. Figure 4.4 shows large wavelet coefficients at all the scales around the time of the step change as well as at the end of the time interval. This phenomenon is called cone of influence (COI).

COI at each scale  $a$  and position  $b$  is the set of wavelet coefficients which are influenced by the signal's value at position  $b$ . COI depends on the width of the

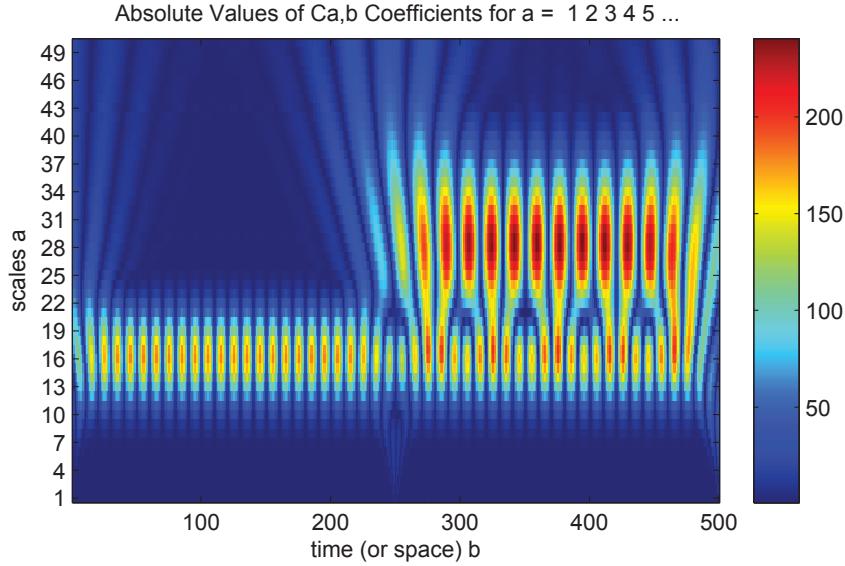


Figure 4.3: Wavelet coefficients of  $x(t)$  for scales from 1 to 50

wavelet function. For example if the support of a wavelet function at a specified scale  $a$  is  $[-4, 4]$ , samples within  $[b - 4, b + 4]$  influence the value of the wavelet coefficient estimated at position  $b$ . In other terms, signal's value at position  $b$  influences wavelet coefficients estimated at positions within  $[b - 4, b + 4]$ . Therefore, when a sudden change happens in the signal similar to a step change, there will be nonzero wavelet coefficients at all scale around the time of the step change. At larger scales, the number of coefficients influenced by the step change increases since the width of the wavelet function is larger. In general, the wavelet coefficients within the interval  $[-Ca + b, Ca + b]$  are influenced by the signal's value at position  $b$  where  $2C$  is the support of the mother wavelet function and  $a$  is the scale.

The cone shape of the wavelet coefficients around the step change is because of the increase in the number of influenced coefficients with increasing the scale. This phenomenon is also observed at the end of the time interval. The reason for this is that the signal is padded with zeros when estimating the coefficients due to the finite length of the signal. Therefore, there is a sharp change at the end of the time interval. This sharp change would have appeared at the beginning of the time interval if the value of the signal was not 0. Therefore, the coefficients estimated at the beginning and end of the time interval are not reliable for the analysis due to the influence of the added zeros. For the rest of the analysis, at each scale  $Ca$  first and final coefficients are removed in order to avoid their influence on the analysis. For the standard Morlet wavelet in MATLAB  $C$  equals to 4.

Even though the COI has the mentioned adverse effect, it is helpful in order to

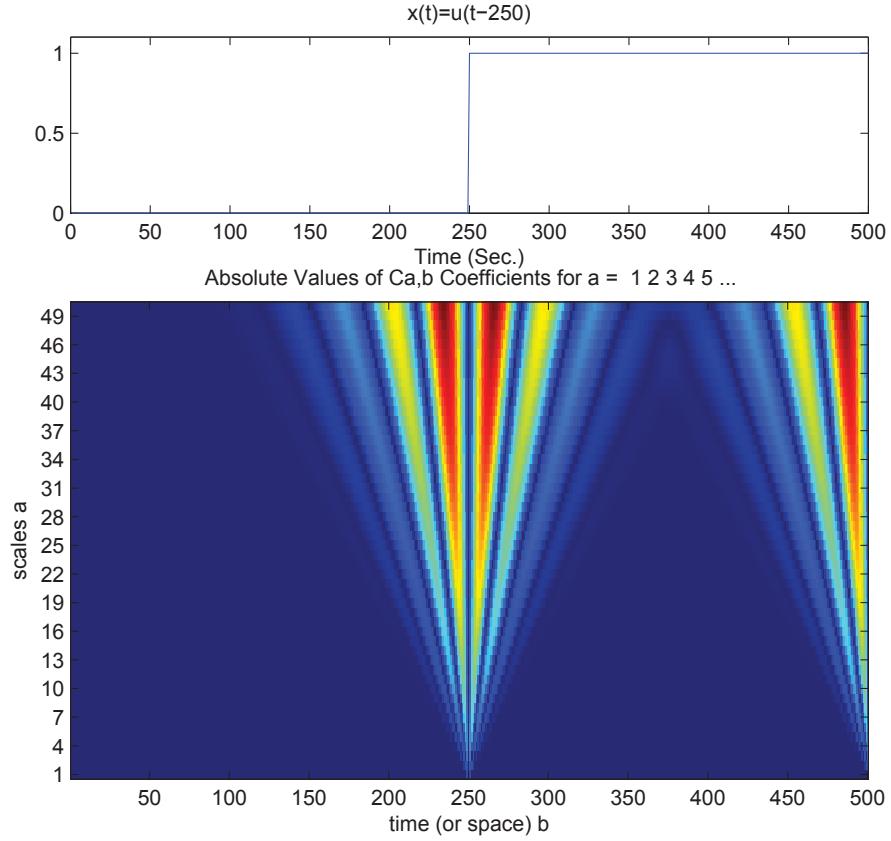


Figure 4.4: Top:  $x(t) = u(t - 250)$ . Bottom: Wavelet coefficients of  $x(t)$

detect sudden changes in the variables. Abrupt changes are common in industrial variables due to setpoint changes, manual control, change in the feed etc.. An application of wavelet transform is in detection of abrupt changes since wavelet coefficients have large values around the time of abrupt changes [56].

Abrupt changes in variables are specifically detrimental to most oscillation detection algorithms since they distort the Fourier transform of the variables in all the frequencies as well as the estimated ACF. However, abrupt changes may not be visible from the signal itself and therefore, it is another advantage of wavelet transform that it makes it easier to detect the abrupt changes. Such changes should be removed from data before analyzing it for fault detection purposes.

### 4.3.2 Data pre-processing

Data pre-processing is a mandatory step before fault detection based on statistical analysis of the data. Industrial variables are usually contaminated with noise, non-

stationary trends as well as abrupt changes. Even though using robust fault detection algorithms may help in correct fault detection and diagnosis problem, removing the non-stationary trends and noise from data is important. Wavelet transform is extensively applied for signal or image noise filtering. Also, extracting the non-stationary trend of the variables is another application of wavelet transform. This section provides a short review on the literature on these subjects.

Noise is usually considered as an additive component to the true signal as

$$\hat{x}_n = x_n + \varepsilon_n \quad (4.10)$$

where  $x_n$  is the true signal and  $\hat{x}_n$  the noisy signal. Estimating  $x_n$  with minimum error from  $\hat{x}_n$  is the target of noise filtering algorithms. Application of wavelet transform for signal or image noise filtering is extensively studied in the literature and yet is a subject of interest in academia. It is mentioned that [57] noise filtering based on wavelet transform has a better performance compared to common frequency domain filtering methods.

Noise usually affects the estimated wavelet coefficients at all scales. One of the most famous filtering procedures is hard/soft thresholding originally proposed by [58]. In hard thresholding the wavelet coefficients with absolute values less than a specified threshold are pushed to zero as:

$$y_f = \begin{cases} y & |y| > T \\ 0 & |y| < T \end{cases} \quad (4.11)$$

In soft thresholding not only the small coefficients are shrunk to 0 but other coefficients are also decreased by the amount of the threshold as:

$$y_f = \begin{cases} sgn(y)|y| - T & |y| > T \\ 0 & |y| < T \end{cases} \quad (4.12)$$

where  $sgn$  is the sign function.

Hard thresholding tends to better keep the signal features while it may cause abrupt changes. Soft thresholding on the other hand is better in removing noise and making the signal smoother. There are many different algorithms proposed for optimal selection of the threshold [59], the most famous one being  $T = \sqrt{2\sigma^2 \log N}$  proposed by [58] where  $N$  is the number of samples and  $\sigma^2$  is the noise variance. [60] proposed a method based on Bayesian framework for optimal selection of the threshold and [61] proposed an interscale orthonormal thresholding method which is optimal in the sense of minimizing the MSE. [62] reviewed some of the noise filtering methods based on the wavelet transform. Even though there are advanced methods proposed in the literature for noise filtering, hard thresholding algorithm is applied

in this work since it better preserves the features of the data and has a satisfactory performance.

Extraction of the non-stationary trends in the variables is another application of wavelet transform. Trend is usually defined as an additive non-stationary element to the signal as:

$$x_n = tr_n + r_n \quad (4.13)$$

where  $n$  is the time sample,  $tr$  denotes the trend and  $r$  is the residual. The residual part of the signal is assumed to be stationary which could contain both deterministic and stochastic elements. Trend is also interpreted as the smooth change in the mean value of the signal [63].

Extracting the trend of variables helps in studying the long term behavior of the variables as well as obtaining the stochastic stationary part of the signals for statistical analysis. Different methodologies are proposed in the literature for trend extraction. [63] has presented a review of some of the mostly applied trend extraction algorithms including wavelet transform. [64] has presented a novel trend extraction algorithm based on wavelet transform and compared its performance to linear and nonlinear filtering approaches.

The basic trend extraction method based on wavelet transform can be summarized in the following steps:

- Compute the wavelet transform of the signal for a proper range of scales
- Make the wavelet coefficients belonging to small and intermediate scales equal to 0 to only keep the information of the large scales
- Calculate the inverse wavelet transform of the remaining wavelet coefficients.

Since in this work the purpose is to extract the stationary portion of the data for further analysis by removing the trend, the wavelet coefficients of the large scales are pushed to 0 instead of small scales. Decision on which scales should be forced to 0 can be done based on the prior knowledge on the largest oscillation period to be detected based on the analysis. The other way to choose the level is based on the length of the data. Since oscillation diagnosis requires presence of minimum 5 oscillation periods in the data, scales larger than 1/4 of the length of the data can be pushed to 0 to obtain a stationary signal. Even though the application of more advanced methods could be valuable, experience shows that simply making the wavelet coefficients for very large scales equal to 0 works well as a de-trending procedure.

As an example to illustrate the importance of data pre-processing and the advantage of using wavelet method, consider the output of a flow control loop in industry

as plotted on the top panel of Figure 4.5. The trend of the variable estimated based on wavelet transform is also plotted in red dotted line and the detrended variable in the middle panel. As can be seen from the figure, the trend in the variable is well removed but the sharp change close to time sample 3000 is still present. Sharp changes similar to this usually cannot be removed by detrending and they disturb the result of statistical analysis of the data. There are two simple ways to deal with these sharp changes. One way is to use portions of the data which do not contain these temporary changes. The other is to manually remove these changes from data and then detrend the data. Removing a portion of the data and attaching the remnant together can disturb the analysis but if the data length is not enough it is possible to do so. The third panel of Figure 4.5 plots the detrended data after removing the data from time sample 2650 to 2850.

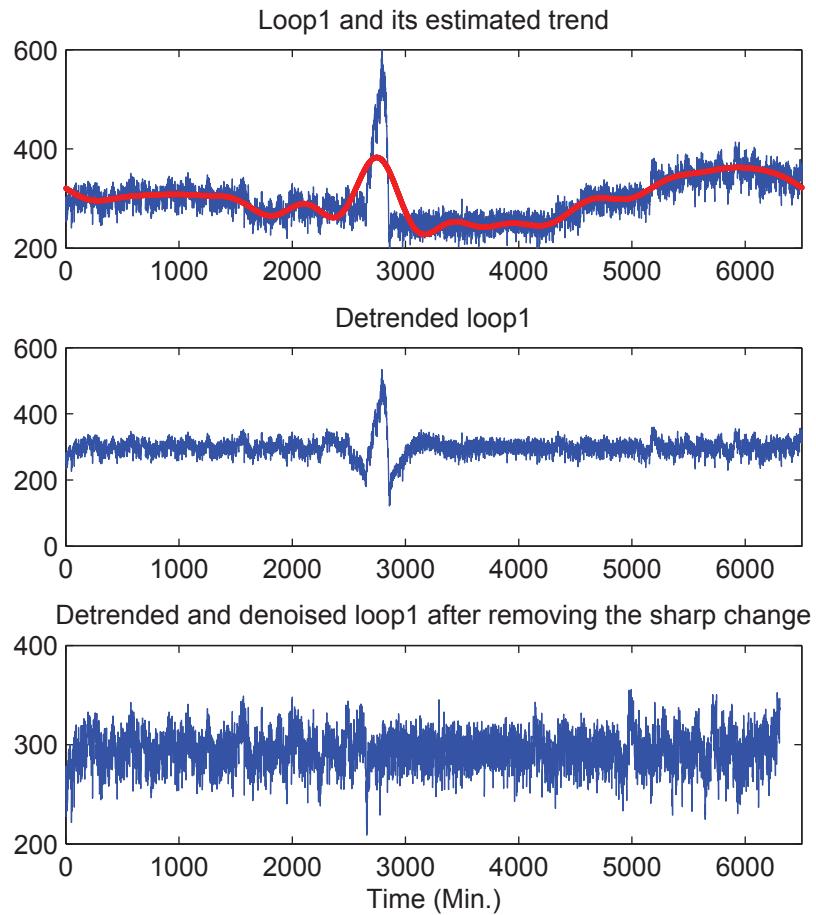


Figure 4.5: Time domain plot of variable loop1

Figure 4.6 plots the estimated ACF of the original variable on the top panel along

with the ACF of its detrended and denoised version on the bottom panel. From the figure we can see that even though the ACF of the original variable does not show signs of the presence of an oscillation, an obvious oscillation is observed in the ACF of the detrended version. Applying the oscillation detection algorithm proposed in the second chapter results in estimating an oscillation with period of 146 samples in the variable. The standard deviation of the estimated period is 23.

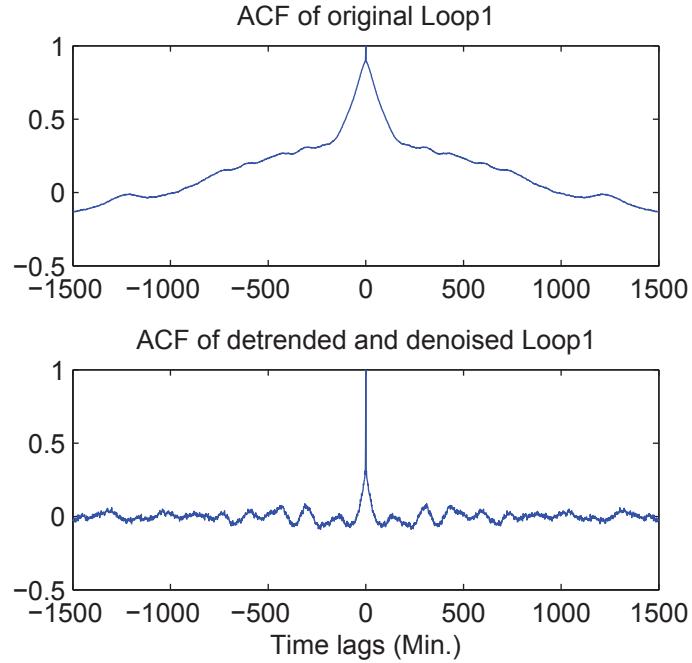


Figure 4.6: ACF of the Loop1 before and after pre-processing

The map of wavelet coefficients of Loop 1 estimated for scales from 1 to 200 is plotted in Figure 4.7. The figure does not indicate which scale contains the highest energy. However, it shows that the energy of the signal approximately from time sample 3000 to 4000 is reduced at all scales.

The scales containing the highest energy are likely to correspond to the periods of oscillations present in the variable. To find the scale containing the highest amount of energy, Figure 4.8 plots the average absolute value of wavelet coefficients at each scale individually. The average is taken over the time samples in order to obtain a general presentation of the signal. It should be mentioned that since in larger scales the signal is averaged over longer time periods, the wavelet coefficients tend to have larger absolute values for very large scales. However, that does not necessarily indicate that the signal is oscillating with a period corresponding to those scales. Only if there is a local peak in the curve of signal's energy versus scale, it is likely that the signal contains an oscillation with a period corresponding to the scale of the

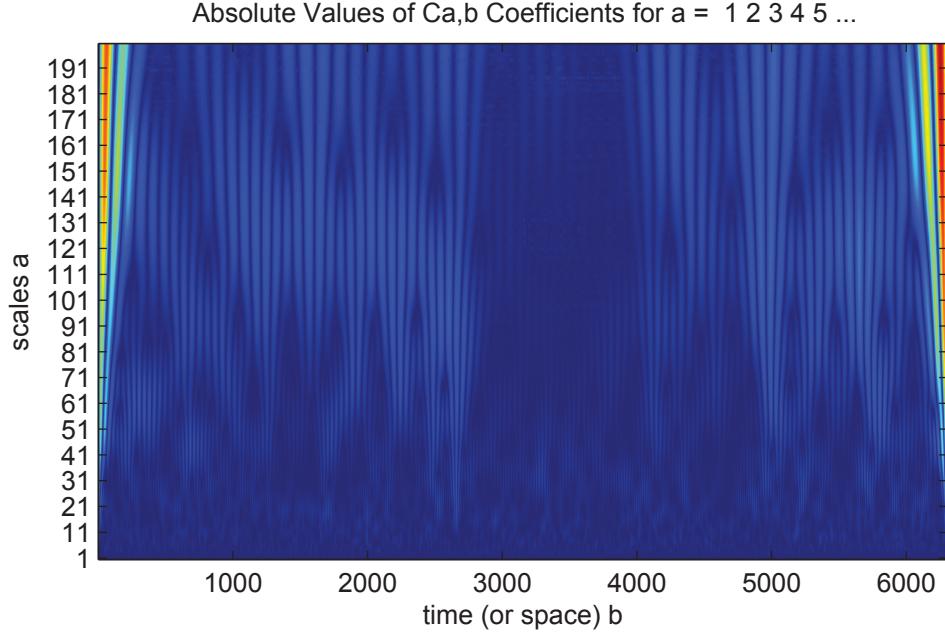


Figure 4.7: Wavelet coefficients of Loop1

local peak. Therefore, only the local peaks are required to be analyzed for presence of oscillations.

Figure 4.8 shows that scales from 120 to 140 (approximately corresponding to oscillation periods from 147 to 172) are on a local peak. Figure 4.9 plots the wavelet coefficients of scale 130. Figure 4.9 also shows the suppression of signal's energy from time sample 2600 to 4000 which requires more investigation in order to find the reason. Since the data mainly consists of noise during this period and does not have the characteristics of the rest of the data, it is better to remove it so that it does not disturb the result of statistical analysis .

The ACF of Loop 1 after removing samples from 2600 to 4000 is plotted in Figure 4.10. As can be seen from the figure, the oscillation in the ACF is more apparent since the influence of the time period when the oscillation does not exist in the data is removed from the estimated ACF. The oscillation period estimated from this ACF equals to 144 with a standard deviation of 13. The period is close to the one estimated previously (149) but the standard deviation is much smaller.

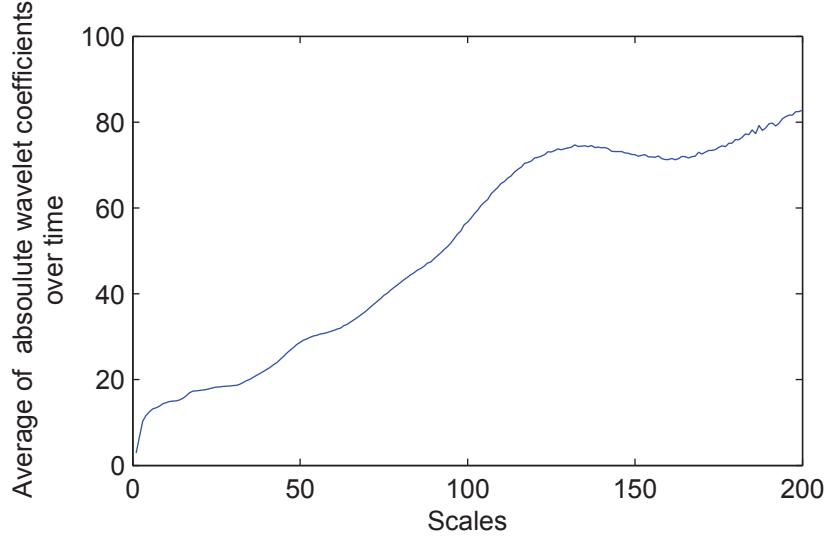


Figure 4.8: Average absolute wavelet coefficients of Loop1

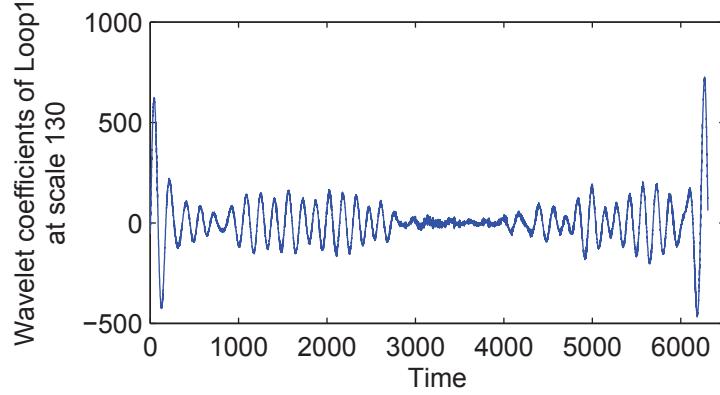


Figure 4.9: Wavelet coefficients of Loop1 at scale 130

## 4.4 Wavelet power spectrum and its application for oscillation diagnosis

Wavelet power spectrum (wps) is utilized in this work for the purpose of oscillation diagnosis. wps of a variable  $x(t)$  at time  $b$  and scale  $a$  is estimated as

$$wps_x(b, a) = |w_x(b, a)|^2 \quad (4.14)$$

The confidence interval for wps of a normally distributed random variable is available in the literature. Zhang et al. [65] mathematically proved that the distribution of the wavelet power spectrum of white noise based on Morlet wavelet function can be obtained as

$$|w_x(b, a)|^2 \Rightarrow \frac{\sigma^2}{2}(1 + e^{-\omega_c^2})X_1^2 + \frac{\sigma^2}{2}(1 - e^{-\omega_c^2})X_2^2 \quad (4.15)$$

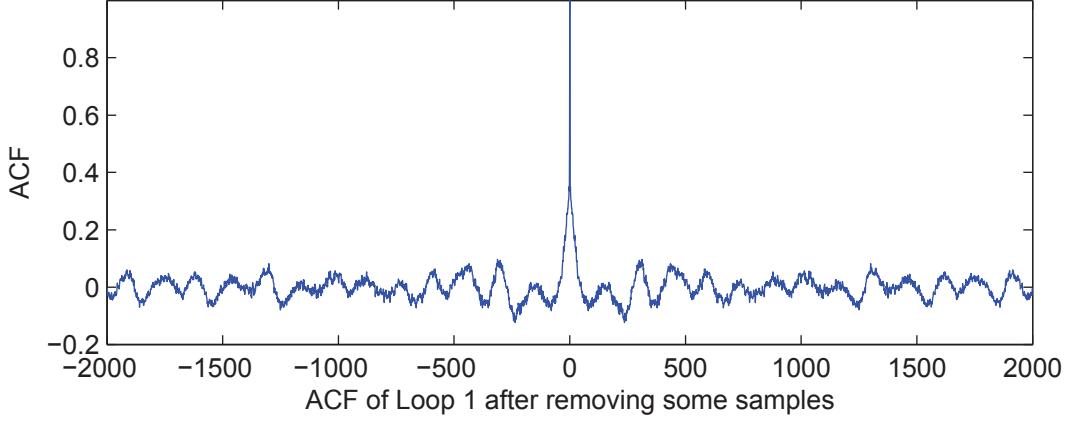


Figure 4.10: ACF of Loop1 after removing the time samples from 2600 to 4000

where  $X_1$  and  $X_2$  are independent standard Gaussian random variables and  $\sigma^2$  is the noise variance. If the wavelet central frequency is sufficiently large ( $\omega_c \geq 6$ ) the exponential terms will be 0 and the distribution simplifies to  $|w_x(b, a)|^2 \Rightarrow \frac{\sigma^2}{2}(X_1^2 + X_2^2)$  or

$$|w_x(b, a)|^2 \Rightarrow \frac{\sigma^2}{2}\chi_2^2 \quad (4.16)$$

where  $\chi_2^2$  has Chi squared distribution with 2 degrees of freedom. It should be noted that if the wavelet function is real with no imaginary part, the Chi squared distribution will have 1 degree of freedom.

Torrance et al. [43] heuristically proposed that the probability distribution of the wavelet spectrum of the response of a general linear process to Gaussian innovations is:

$$|w_x(b, a)|^2 \sim \frac{1}{2}\lambda_k\chi_2^2 \quad (4.17)$$

where  $\lambda_k$  is the power spectrum of the data at frequency  $k$  corresponding to scale  $s$ . A confidence interval for the true value of  $|W_x(b, a)|^2$  is derived based on the estimated value and the distribution in Equation 4.17 as shown in Equation 4.18 [43].

$$\frac{2}{\chi_2^2(p/2)}|w_x(b, a)|^2 \leq |W_x(b, a)|^2 \leq \frac{2}{\chi_2^2(1-p/2)}|w_x(b, a)|^2 \quad (4.18)$$

where  $p$  is the significance level and  $\lambda_k$  in Equation 4.17 is replaced by the true wavelet spectrum value. Even though this confidence interval is utilized a lot in literature (the corresponding paper was highly cited), it is questionable. Since the expected value of wps of colored noise is not 0, the distribution cannot be central Chi squared. There is no proof that the distribution is Chi squared in general, but even if assuming it is Chi squared it has to be non-central Chi squared. However, Equation 4.18 can be used

for Gaussian noise and its use in oscillation diagnosis is described in the following section.

#### 4.4.1 Diagnosis of oscillations between controller tuning and external disturbance based on wavelet power spectrum

This section proposes a hypothesis test based on using wps for automatic diagnosis between oscillations due to controller tuning and external disturbances. The diagnosis is done based on the different properties of the ACFs of oscillations caused by controller tuning compared with sinusoidal disturbances as described in [54]. As an example, Figure 4.11 plots  $y(t) = \frac{1}{1-0.33z^{-1}+0.9z^{-2}}\varepsilon_t + 2\sin(2\pi/25t)$  along with its ACF and wavelet transform of its ACF in the third panel. From the ACF in the second panel, it can be seen that there is a high frequency oscillation present in the variable in addition to the sinusoidal wave. The third panel shows that two scales are coded with brighter colors which indicate presence of oscillations in the variable with periods close to the two scales. The oscillation due to controller tuning has 3.5 samples per period which corresponds to approximately the same scale in the wavelet domain. As can be seen in Figure 4.11, the wavelet coefficients at scales close to 3 and 4 have varying strength which is due to the varying amplitude of the oscillation itself. However, the scales close to the period of 25 samples which correspond to the harmonic in  $y_t$  has the same pattern during the whole time period.

In this work, CWT of the ACF of the signal is analyzed instead of the CWT of the original variable for the diagnosis of oscillations. As was discussed in [54], ACF of a harmonic with added white noise ( $x_t = A\cos(\omega t + \phi) + \varepsilon_t$ ) has a deterministic behavior which can be determined by knowledge of its frequency and the value of one of the peaks in the ACF ( $P_r$ ) as  $\rho_\tau = p_r \frac{N-\tau}{N-\tau_r} \cos(\omega\tau)$  where  $N$  is the sample size and  $\tau$  is the time lag. The method to find the peaks of the ACF and estimate oscillation frequency is presented in [38]. The CWT of this ACF is estimated as the convolution of two deterministic signals (ACF and the wavelet function). Thus, the CWT of the ACF of a harmonic has deterministic values while CWT of the variable itself is random due to the noise effect.

For an illustration on this fact, the wavelet coefficients of the ACF of  $y(t)$  at two scales of Figure 4.11 corresponding to the two different oscillations are plotted in Figure 4.12. As can be seen at the bottom panel of Figure 4.12, the CWT coefficients corresponding to the harmonic part of the signal have a constant frequency oscillation with an amplitude which linearly decays with time as is expected from the ACF of a harmonic itself. Also, the coefficients at the top panel corresponding to the oscillation

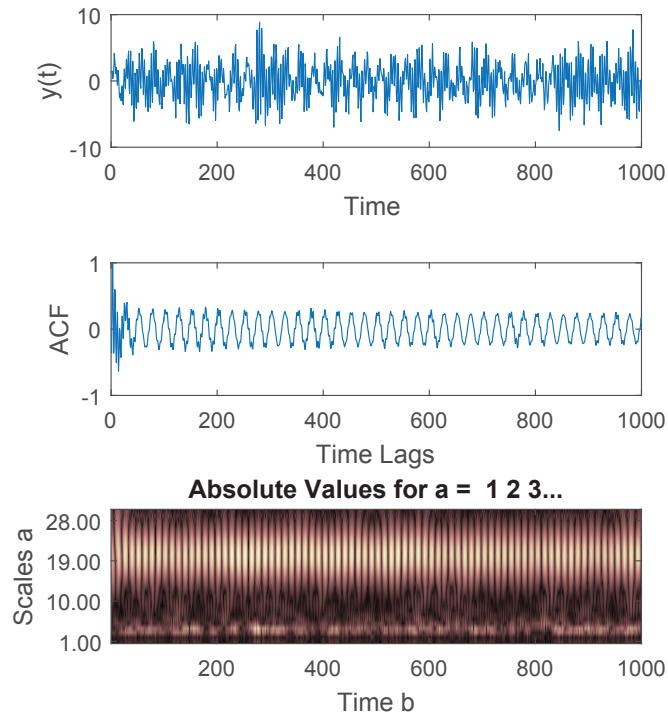


Figure 4.11:  $y(t) = \frac{1}{1-0.33z^{-1}+0.9z^{-2}}\varepsilon_t + 2\sin(2\pi/25t)$  in time along with its ACF and continuous wavelet transform of its ACF

due to the controller tuning show oscillation with varying amplitude as is expected from the respective ACF. The capability of the wavelet transform in isolating different oscillatory components of the signal makes it possible to individually diagnose these two different oscillatory components of the variable.

A hypothesis test needs to be developed for distinction between these two oscillation types. The null hypothesis is that the oscillation is a harmonic process and not due to controller tuning. The viability of the hypothesis test is due to availability of the probability distribution of wps of Gaussian noise. As mentioned, the ACF of a sinusoidal signal with added noise can be written as  $\rho_\tau = p_r \frac{N-\tau}{N-\tau_r} \cos(2\pi f\tau)$ . The exact value of the CWT of this ACF is not of concern here. However, since the CWT at the scale corresponding to  $f$  is just the convolution of the wavelet function with this ACF, it can also be written as

$$p_m \frac{N-t}{N-t_m} \cos(2\pi ft) \quad (4.19)$$

where  $p_m$  is the peak value at time  $t_m$  which is in the middle of the wavelet coefficient's time domain. The peak at the middle time is considered to avoid the uncertainty in

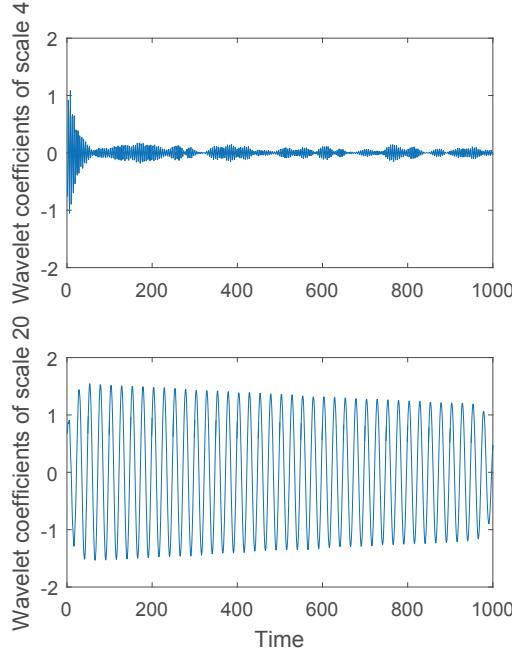


Figure 4.12: Wavelet coefficients of two scales of the transform of the ACF of  $y(t) = \frac{1}{1-0.33z^{-1}+0.9z^{-2}}\varepsilon_t + 2\sin(2\pi/25t)$

the estimated CWT coefficients both at the beginning and end due to the cone of influence phenomena.

It is expected that by subtracting the expected CWT coefficients of the ACF of a harmonic from the estimated values, no residuals will be left. The residuals are obtained as

$$R = w_\rho(b, a) - p_m \frac{N-t}{N-t_m} \cos(2\pi ft) \quad (4.20)$$

where  $E[R] = 0$ .  $R$  is a normally distributed variable with 0 mean value and the sample variance ( $\hat{\sigma}^2$ ) can be calculated. Therefore, the squared value of standardized  $R$  will follow standard Chi squared distribution with 1 degree of freedom.

The hypothesis test can be established based on the variance of the residuals  $R$ . The variance of i.i.d. random variables following  $\chi_d^2$  distribution is known to be  $\sigma^2 = 2d$  where  $d$  is the degrees of freedom. The sample variance ( $\hat{\sigma}^2$ ) is known to have a variance as:

$$\text{Var}[\hat{\sigma}^2] = \frac{1}{N}(\mu_4 - \frac{N-3}{N-1}\sigma^4) \quad (4.21)$$

where  $N$  is the sample size,  $\mu_4$  the forth moment of the distribution and  $\sigma^4$  the squared value of the true variance which is  $2^2$  in this case. The forth moment of Chi square distribution is  $\mu_4 = 12d(d+4)$  [66]. Therefore,  $\text{Var}[\hat{\sigma}^2] = \frac{1}{N}(60 - 4 \times \frac{N-3}{N-1})$  for a random variable distributes as  $\chi_1^2$ .

If the oscillation is truly harmonic, then  $\sigma^2$  should be within the confidence interval of  $\hat{\sigma}^2$ . However, if the oscillation is due to controller tuning, residuals  $R$  will be significantly different from 0 due to the random oscillation amplitude. Thus, the estimated variance  $\hat{\sigma}^2$  will be larger than  $\sigma^2$ . The null hypothesis is rejected when  $\sigma^2$  is below the lower confidence level of  $\hat{\sigma}^2$ .

Another hypothesis test as a visual examination can also be developed based on verifying the ratio between the calculated wavelet power spectrum of the ACF and the expected values. From Equation 4.17 we can derive

$$\frac{|w(n, s)|^2}{|W(n, s)|^2} \sim \frac{1}{2}\chi_1^2 \quad (4.22)$$

In Equation 4.22  $w(n, s)$  is taken as the estimated value of wavelet coefficients while  $W(n, s)$  is taken as the expected value by Equation 4.19 assuming that the oscillation is harmonic.

This hypothesis test is equivalent to assuming a confidence interval for the ratio between the calculated wavelet spectrum and the expected ones. The ratio corresponding to truly harmonic signals should be equal to 1 with a minor error since

$$E\left[\frac{|w_\rho(n, s)|^2}{|p_m \frac{N-t}{N-t_m} \cos(2\pi f t)|^2}\right] = 1 \quad (4.23)$$

If the ratio is significantly different from 1, then it is inferred that the oscillation is caused by the system due to controller tuning. The significance level of the distribution determines the confidence interval.

Figure 4.13 plots this ratio corresponding to the harmonic oscillation in the data plotted in Figure 4.12 while Figure 4.14 plots the ratio corresponding to the oscillation caused by controller tuning in Figure 4.12. Only the peak values are considered in this analysis. As can be seen in Figures 4.13 and 4.14, the ratio corresponding to the harmonic oscillation is close to 1 with small error range while the ratio corresponding to the oscillation due to controller tuning is far from 1.

## 4.5 Wavelet coherency

Coherency is the generalization of power spectrum to two variables. The classical coherency between two variables  $x$  and  $y$  at frequency  $f$  is defined as

$$COH_{x,y}(f) = \frac{S_{x,y}(f)}{[S_{x,x}(f).S_{y,y}(f)]^{1/2}} \quad (4.24)$$

where  $S_{x,y}(f)$  is the cross-spectral density between the two variables  $x$  and  $y$  (i.e. the Fourier transform of the cross-correlation function  $R_{x,y}(\tau)$ ).

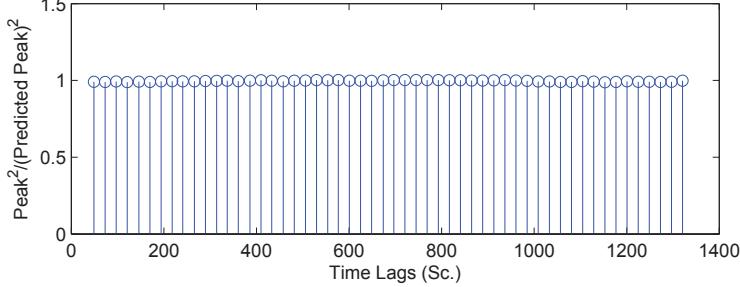


Figure 4.13: Ratio of the calculated wavelet power spectrum to the predicted ones at the peaks corresponding to the harmonic oscillation

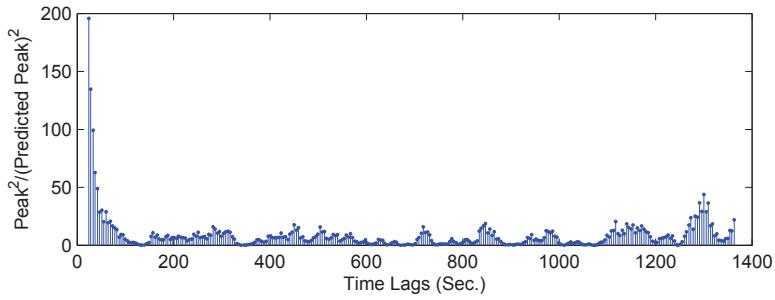


Figure 4.14: Ratio of the calculated wavelet power spectrum to the predicted ones at the peaks corresponding to the oscillation generated by the linear system

The absolute value of  $COH_{x,y}(f)$  is interpreted as the correlation coefficient between the components of  $x$  and  $y$  at frequency  $f$  [67]. If  $|COH_{x,y}(f)|$  is close to 1 for all frequencies, then  $x$  can be approximated by a linear time-invariant transformation of  $y$  and vice versa. The phase of the coherence is also informative and is called as phase coherence.

Coherency in the wavelet domain, which is the time-frequency localized correlation function between the two variables, is estimated as

$$WCOH(t, f) = \frac{SW_{x,y}(t, f)}{[SW_{x,x}(t, f).SW_{y,y}(t, f)]^{1/2}} \quad (4.25)$$

where

$$SW_{x,y}(t, f) = \int_{t-\frac{\delta}{2}}^{t+\frac{\delta}{2}} w_x(\tau, f)w_y^*(\tau, f)d\tau \quad (4.26)$$

It can be proved that the value of wavelet coherency is between 0 and 1 based on the Schwarz inequality. Wavelet coherence between  $x$  and  $y$  can be interpreted as the correlation coefficient between the frequency components of  $x$  and  $y$  around time  $t$ . It can be applied to monitor the variables for detection of a change in the structure of the data generation process. The significance level of wavelet coherence

can be estimated based on the distribution of coherence values estimated for two independent signals.

#### 4.5.1 Bicoherence

Bicoherence generalizes coherency to measure the interaction between two frequencies in one variable or between two variables. The interaction between two frequencies which is also called phase coupling is usually a sign of nonlinearity or existence of some kind of structure in the variable.

Phase coupling occurs when oscillations with two frequencies,  $f_1$  and  $f_2$ , are present in the signal as well as the frequency  $f = f_1 + f_2$ , while the phases of these frequency components also have the relation  $\phi_{f_1} + \phi_{f_2} = \phi_f + const..$  The value of the bicoherence estimated for two frequencies is close to 1 if these conditions are satisfied and is close to 0 otherwise.

To understand bicoherence, it should be mentioned that bicoherence is the normalized version of bispectrum which is the frequency domain counterpart of the third order cumulant  $C_3$ . Third order cumulant of variable  $x(t)$  with 0 mean value is defined as:

$$C_{3x}(m, n) = E[x(t)x(t+m)x(t+n)] \quad (4.27)$$

A linear variable  $x(t)$  (linear in the sense that it can be expanded as  $x(t) = \sum_{s=-\infty}^{\infty} h(t-s)\varepsilon(s)$  where  $\varepsilon(t)$  is i.i.d. innovations), is normally distributed if the distribution of innovations  $\varepsilon(t)$  is normal.  $C_3$  of a linear time series can be non-zero only if the innovations  $\varepsilon(t)$  are not normal distributed and have non-zero third-order cumulant.

When the innovations are normal distributed,  $x(t)$  may have non-zero  $C_3$  only if it is a nonlinear function of the innovations. Therefore, non-zero third order cumulant indicates that the variable does not have a symmetric distribution which could be due to the non-linearity in signal generation process or the non-Gaussian distribution of the innovations.

Bispectrum can be estimated as the Fourier transform of  $C_3$  or directly based on the Fourier transform as:

$$bis(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)] \quad (4.28)$$

where  $X(f)$  is the discrete Fourier transform of time series  $x(t)$  and  $*$  denotes complex conjugate. Bispectrum can indicate if there is an interaction between the two frequencies  $f_1$  and  $f_2$  in signal  $x(t)$ .

The value of bispectrum depends on the signal's energy and in order to achieve a

normalized quantity for comparison, bicoherence is defined as

$$bic(f_1, f_2) = \frac{E[X(f_1)X(f_2)X^*(f_1 + f_2)]}{\sqrt{E[|X(f_1)X(f_2)|^2]E[|X(f_1 + f_2)|^2]}} \quad (4.29)$$

where  $X(f)$  is the discrete Fourier transform of time series  $x(t)$  and  $*$  denotes complex conjugate. Bicoherence itself is a complex quantity while squared bicoherence has a real value between 0 and 1. For more details on the methods for estimation of bicoherence, readers can refer to [68, 69]. It should be noted that bispectrum (or bicoherence) has 12 symmetric regions in the  $f_1, f_2$  plane [70]. The principal domain is defined as the region where  $f_1 \leq f_2$  and  $f_1 + f_2 \leq f_s$ . The principal domain contains all the information and the other regions are redundant.

As an example to examine the estimated bicoherence values for a linear signal, consider  $x(t) = \frac{z^{-2} + 0.3z^{-3}}{z^2 - 0.8z + 0.25}\varepsilon_t$  where  $\varepsilon_t$  is standard normally distributed variable. The bicoherency of  $x(t)$  is plotted in Figure 4.15 which shows that the estimated values are not significantly different from 0. The method to determine a significance level will be described shortly.

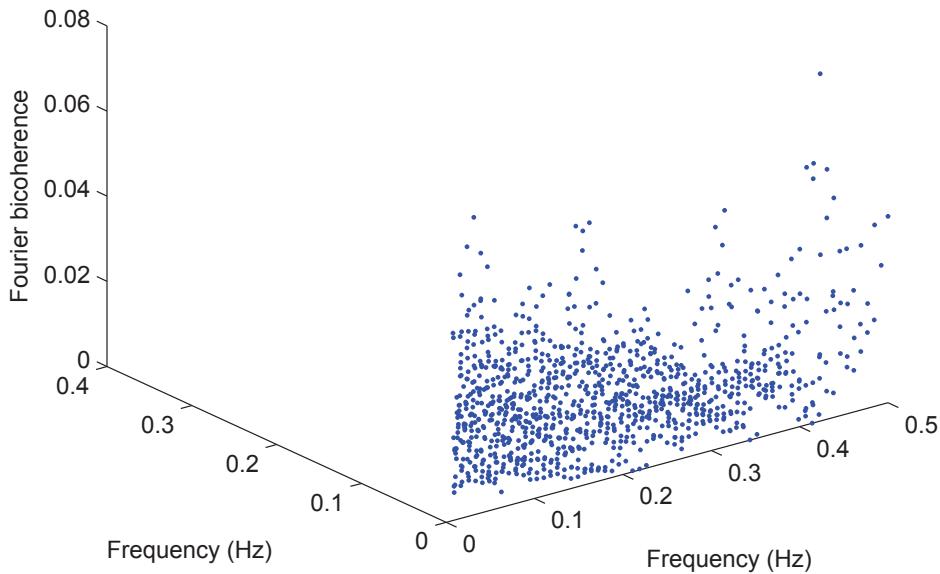


Figure 4.15: Bicoherency of  $x(t) = \frac{z^{-2} + 0.3z^{-3}}{z^2 - 0.8z + 0.25}\varepsilon_t$  in frequency domain

The main application of bicoherence is to detect if the distribution of the variable is non-Gaussian. It is proved in the literature that even though a linear time series with

non-Gaussian innovations will have non-zero bicoherence, the value of bicoherence will be constant and independent of bi-frequencies [70]. On the other hand, if there is non-linearity in the structure of data generation process, bicoherence will have non-zero and non-constant values. These facts have been used in literature to propose methods for detection of non-normality and non-linearity [70, 71].

The hypothesis tests proposed in the literature mainly use the probability distribution of the estimated bicoherence values of white noise to produce a test for Gaussianity and linearity. It is known that the real and imaginary parts of estimated bicoherence values are normally distributed and asymptotically independent of each other. Also the value of bicoherence at each bi-frequency is independent from the neighbouring frequencies. Thus, the squared bicoherence has a  $\chi^2$  distribution with 2 degrees of freedom.

#### 4.5.2 Wavelet bicoherence

Wavelet bicoherence [72] is the generalization of the classical bicoherence. Similar to other quantities based on wavelet transform, wavelet bicoherence is specifically suitable for analysis of non-stationary signals containing temporary behaviors, pulses and abrupt changes. [73] has applied wavelet bicoherence to detect phase coupling and nonlinearities in the variables.

Wavelet bispectrum is defined as:

$$wbis_x(a_1, a_2) = \sum_{k=1}^m w_x(k, a_1)w_x(k, a_2)w_x^*(k, a) \quad (4.30)$$

where  $\frac{1}{a} = \frac{1}{a_1} + \frac{1}{a_2}$  and the summation is over  $m$  samples around the desired time sample. Bispectrum can be estimated between two variables  $x$  and  $y$  as:

$$wbis_{x,y}(a_1, a_2) = \sum_{k=1}^m w_x(k, a_1)w_x(k, a_2)w_y^*(k, a) \quad (4.31)$$

which measures the interaction between the two signals at scales  $a_1$  and  $a_2$ .

The squared wavelet bicoherence between two scales  $a_1$  and  $a_2$  is defined as:

$$wbic_x(a_1, a_2)^2 = \frac{|wbis|^2}{\sum_{k=1}^m |w_x(k, a_1)w_x(k, a_2)|^2 \sum_{k=1}^m |w_x(k, a)|^2} \quad (4.32)$$

$wbic_x(a_1, a_2)^2$  has a value between 0 and 1. Squared bicoherence will be non-zero if the variable does not follow normal distribution.

[73] also proposed total bicoherence as

$$(Wbic_x)^2 = \frac{1}{2} \sum \sum wbic_x(a_1, a_2)^2 \quad (4.33)$$

where the summation is over all the scales  $a_1$  and  $a_2$ .  $S$  is the number of all the summands in the summation. This quantity is again between 0 and 1 and can be used as an overall representation of the signal to detect presence of non-Gaussian distribution. It is also possible to average the estimated bicoherence values over the time domain in order to capture an average representation of the variable during the whole time period. Wavelet bicoherence averaged through the time domain is called total bicoherence here.

There are few papers in literature proposing methods to estimate the significance level of squared wavelet bicoherence. All of the proposed methodologies are based on comparing the estimated wavelet bicoherence with the one which could be estimated from Gaussian noise due to numerical issues. Ge [74] has mathematically derived the probability distribution function (PDF) of the wavelet bicoherence estimated from white Gaussian noise as:

$$f_{wbic^2}(z) = \frac{\sqrt{N_e(f_1, f_2, f)}}{\sqrt{2\pi z}(1-z)} e^{-(\operatorname{arctanh}^2 \sqrt{z})/2[N_e(f_1, f_2, f)]} \quad (4.34)$$

where

$$N_e(f_1, f_2, f) = \frac{m}{f_c f_s \max[\frac{\sqrt{f_1^2 + f_2^2}}{f_1 f_2}, \frac{2}{f_1 + f_2}] \sqrt{ln d}} \quad (4.35)$$

where  $d$  is the allowable correlation value between the estimated wavelet coefficients. The significance at the level of  $\alpha$  ( $D_\alpha$ ) can be obtained as:

$$D_\alpha = \int_{D_\alpha}^{\infty} f_{wbic^2}(z) dz = \alpha \quad (4.36)$$

Squared bicoherence significantly different from 0 is a sign of non-Gaussian distributed variable. Again, non-constant value of squared bicoherence can be taken as a sign of nonlinear structure in the variable.

For illustration purposes consider a simple signal consisting three oscillatory components as  $x(t) = \sin(2\pi \times 0.04 + \frac{\pi}{3}) + \sin(2\pi \times 0.025 + \frac{\pi}{12}) + \sin(2\pi \times 0.065 - \frac{\pi}{4})$ . The components with frequencies 0.04 and 0.025Hz are coupled to each other since an oscillatory component with frequency equal to the summation of the two frequencies exist in the signal while the phases have a constant relation as well. Figure 4.16 plots the total wavelet bicoherence for signal  $x(t)$  where x and y axis are frequencies. There is an obvious peak in the bicoherence plot as can be seen in Figure 4.17. The bicoherence plot is easier to study by plotting it versus the scales instead of frequency. Figure 4.18 plots the same bicoherence values but versus scales.

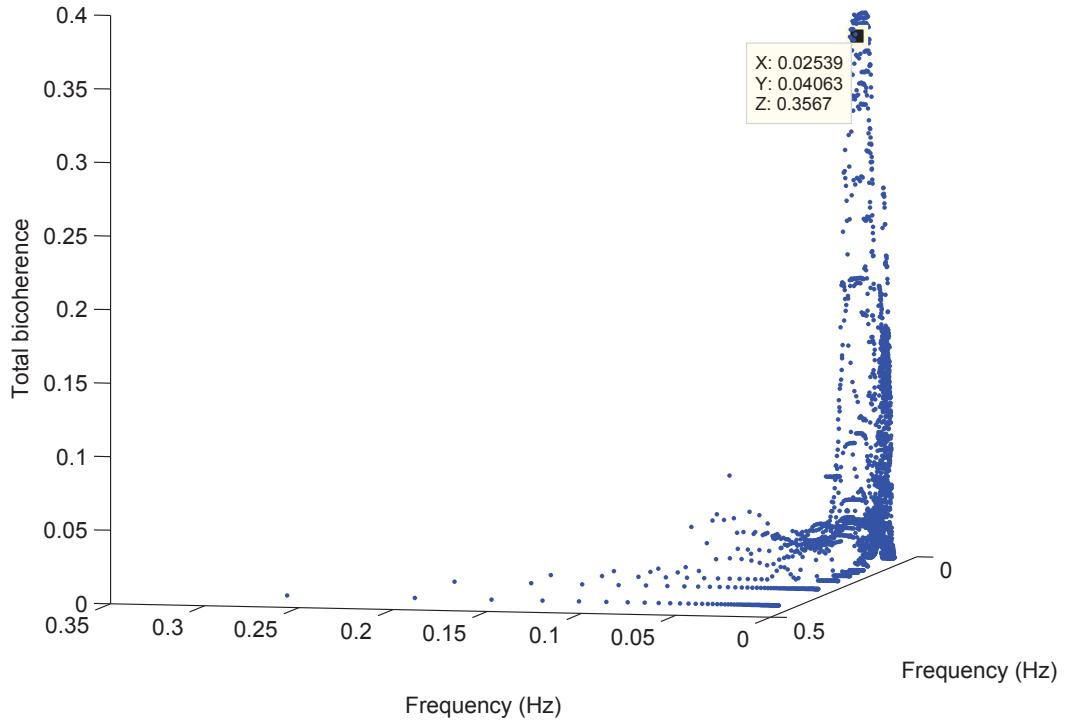


Figure 4.16: Total bicoherence of  $x(t)$  in frequency domain

As can be seen in Figures 4.17 and 4.16 the total bicoherence has large values in frequencies and scales where the coupling exist in the signal. Of course the bicoherence has large values in the neighbouring frequencies and scales as well due to the power leakage. Figure 4.18 also plots the bicoherence of the same signal estimated based on Fourier transform.

Total bicoherence provides a general presentation of the signal which the bicoherence based on Fourier analysis also does. To observe one of the advantages of wavelet bicoherence over Fourier method, the signal has changed to  $y(t) = \sin(2\pi \times 0.04 + \frac{\pi}{3}) + \sin(2\pi \times 0.025 + \frac{\pi}{12}) + u(t - 1000)\sin(2\pi \times 0.065 - \frac{\pi}{4})$  where  $u(t - 1000)$  is a step function with initial value of 1 and final value of 0. The step function eliminates the oscillation with the frequency of 0.065 from the signal from time sample 1000 onward. The bicoherence between scales 32 and 20 of this signal is plotted in Figure 4.19.

From Figure 4.19 the change in the bicoherence at time sample equal to 1000 is easily observable. Also the total bicoherence of the signal is plotted in Figure 4.20. By comparing Figure 4.20 with Figure 4.17, we can see that the value of the total bicoherence has decreased almost by 50%. This decrease is expected since for half

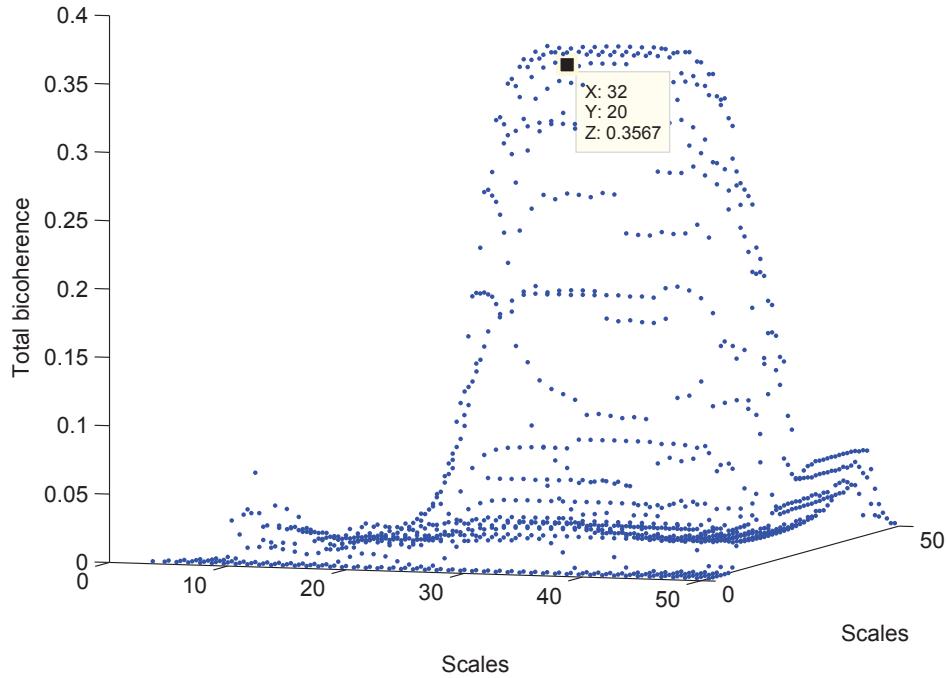


Figure 4.17: Total bicoherence of  $x(t)$  versus scales

of the time the bicoherence has a value close to 0. The bicoherence of the signal obtained by the Fourier method is also plotted in Figure 4.21.

By comparing Figures 4.21 to 4.18, it is observed that the value of the maximum bicoherence has decreased. Since Fourier method can only provide an average representation of the signal, it cannot give any information on the changes of the bicoherence with time. Therefore, wavelet method is a much more suitable tool to learn about the changes in signal's behaviour.

## 4.6 Diagnosis of oscillations due to nonlinearity

The purpose of using bicoherence is to detect if the oscillation is nonlinear which indicates the source of the oscillation to be valve problem. After detection of the oscillation period, it is required to verify if  $wbic_x^2(a_1, a_2)$  estimated for the scale corresponding to the oscillation has a larger value compared to  $wbic_x^2(a_1, a_2)$  estimated for the rest of the scales.

A hypothesis test is required to detect nonlinear oscillations. The null hypothesis is that the oscillation is not nonlinear and therefore, the bicoherence has a constant value at all the scales. In this test, the maximum of  $wbic_x^2(a_1, a_2)$  when  $a_1$  equals to

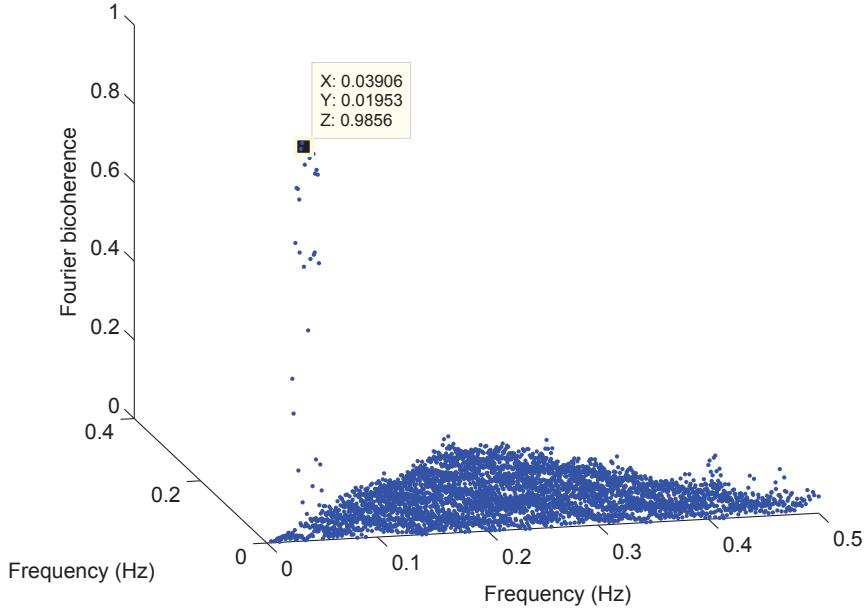


Figure 4.18: Total bicoherence of  $x(t)$  in frequency domain

the oscillation scale  $a_{osc}$  and  $a_2$  changes over the feasible range of scales is taken as

$$wbic_{max}^2(a_{osc}, a_2) = \max_{a_2} (wbic_x^2(a_{osc}, a_2)) \quad (4.37)$$

The average of the  $wbic_x^2(a_1, a_2)$  when neither  $a_1$  nor  $a_2$  is equal to  $a_{osc}$  is also estimated as

$$wbic^2(a \neq a_{osc}) = \frac{1}{N} \sum_{a_1 \neq a_{osc}} \sum_{a_2 \neq a_{osc}} wbic_x^2(a_1, a_2) \quad (4.38)$$

where  $N$  is the number of summands. If  $wbic_{max}^2(a_{osc}, a_2)$  is significantly different from  $wbic^2(a \neq a_{osc})$  then, it is concluded that the oscillation is due to nonlinearity.

Since the probability distribution of  $wbic_x^2(a_1, a_2)$  values is unknown, the hypothesis test is performed based on the Chebyshev's inequality. Chebyshev's inequality states that for a random variable  $x$  with mean  $\mu$  and standard deviation  $\sigma$  the following inequality holds.

$$Pr(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (4.39)$$

In this case  $\mu = wbic^2(a \neq a_{osc})$  and  $\sigma$  can be estimated as the standard deviation of  $wbic^2(a_1 \neq a_{osc}, a_2 \neq a_{osc})$ .  $k$  is a tuning parameter which defines the significance level. If  $wbic_{max}^2(a_{osc}, a_2)$  is significantly larger than  $wbic^2(a \neq a_{osc})$ , it implies the oscillation has nonlinear structure.

In terms of oscillation diagnosis, if a feedback loop is oscillating and no nonlinearity is detected in the oscillation, the cause could be due to a tightly tuned controller,

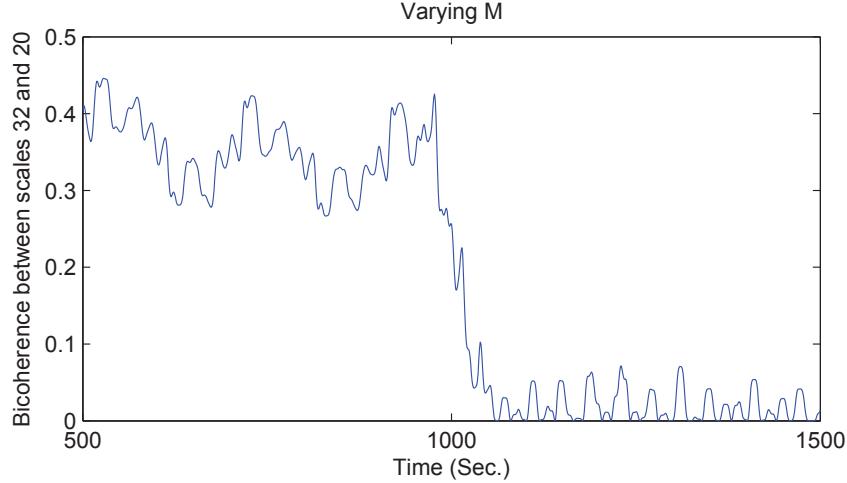


Figure 4.19: Bicoherence of signal  $y(t)$  between scales 32 and 20 in time

external disturbance or due to sensor. The oscillations caused by valve problems such as stiction or backlash are naturally nonlinear. Therefore, bicoherence can diagnose if the fault is caused due to the nonlinear behavior of the valve or the process itself.

The flowchart in Figure 4.22 shows the steps required for oscillation diagnosis. It should be noted that this work diagnoses the oscillation category for each variable individually thanks to the use of wavelet transform. After analysis of all the variables of the process, the ones carrying similar oscillations should be grouped together and further analysis is required to find the loop which propagates the oscillation to other ones.

It is also noteworthy that due to the low pass filtering nature of chemical processes, nonlinear oscillations become more linear getting further away from the oscillation source. The reason is that high frequency harmonics which shape the nonlinear structure of the oscillation get filtered out. This has been used as a method to diagnose which loop has caused the oscillation when several loops carry the same kind of oscillation. Since bicoherence quantifies the nonlinearity, it can be said that the loop with highest bicoherence value is propagating the oscillation to other loops. Diagnosing the source of the oscillation among several loops with similar oscillation requires application of other methodologies if the oscillation is due to controller tuning or due to a sinusoidal disturbance. The appropriate methodologies are being considered in the on-going work.

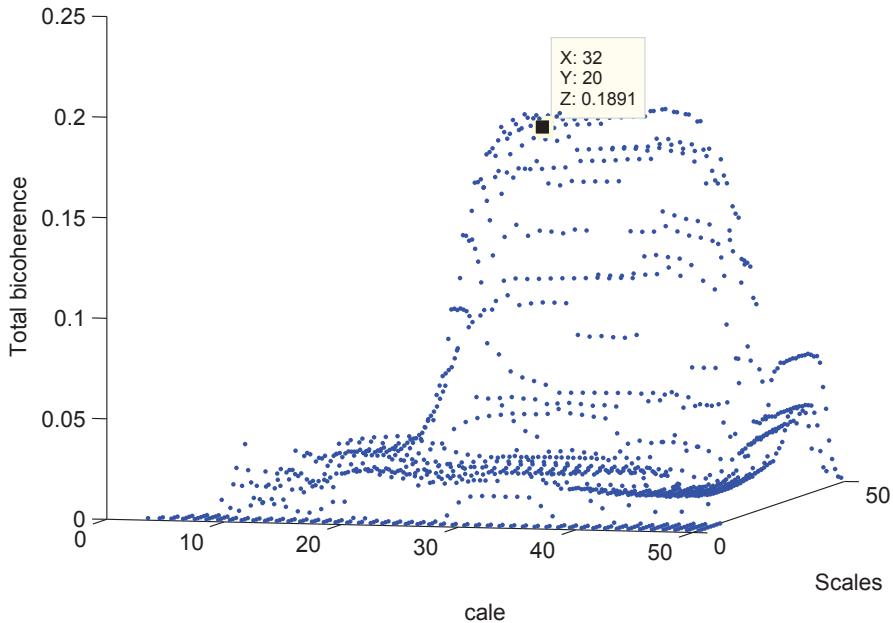


Figure 4.20: Total bicoherence of signal  $y(t)$

#### 4.6.1 A note regarding the selection of the scales for which the bicoherence should be estimated

A correct inference about nonlinearity in the process based on estimated bicoherence values requires careful selection of the scales for which the bicoherence is estimated. Specially estimation of bicoherence for scales larger than what is required may lead to wrong inference about presence of nonlinearity. To illustrate this fact, consider  $x(t) = 2\sin(2\pi \times 0.04 + \frac{\pi}{3}) + 2\sin(2\pi \times 0.025 + \frac{\pi}{12}) + 2\sin(2\pi \times 0.065 - \frac{\pi}{4}) + \varepsilon(t)$  and its bicoherence plotted in Figure 4.23 with consideration of scales up to 80.

From Figure 4.23 it can be observed that the bicoherence value again increases considerably for scales around 60 to 70. These scales approximately correspond to periods between 75 to 85. Periods 75 to 85 are approximate multiples of the two periods present in the signal. As a result, the wavelet coefficients at these scales have larger values compared to neighboring scales leading to a larger bicoherence value between periods 40 (frequency 0.025) and 75 (frequency 0.013). In this case the sinusoidal wave with frequency 0.04 is causing the coupling between these two frequencies.

These large values of bicoherence which can wrongly indicate presence of nonlinearity in the signals can always exist in frequencies which are multiples of the two

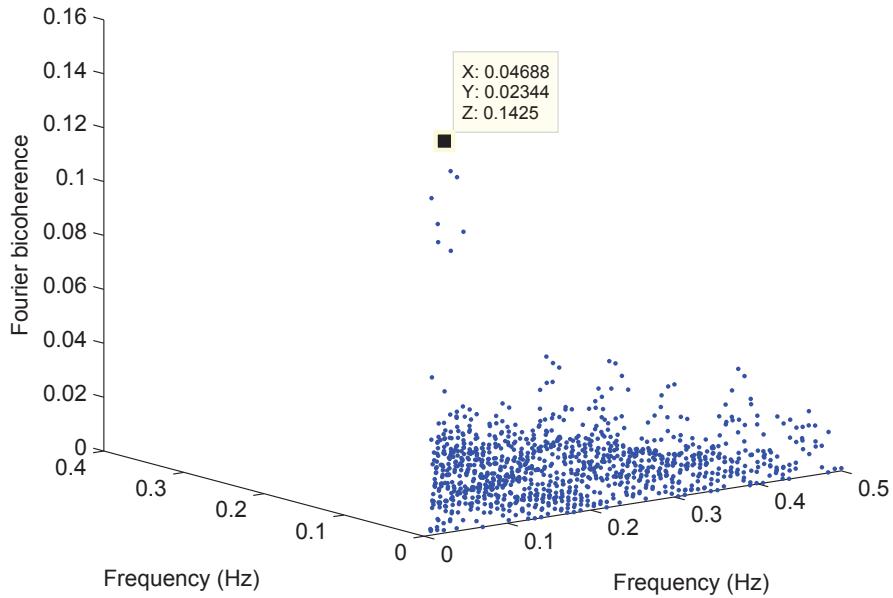


Figure 4.21: Bicoherence of signal  $y(t)$  based on Fourier method

frequencies present in the signal. The same bicoherence is again plotted in Figure 4.24 when the sinusoidal wave with frequency of 0.065 is removed from the signal. Even though there is no real phase coupling in signal  $x(t) = \sin(2\pi \times 0.04 + \frac{\pi}{3}) + \sin(2\pi \times 0.025 + \frac{\pi}{12})$ , high value bicoherency is observed in periods from 75 to 85 similar to Figure 4.23.

To avoid wrong detection of phase coupling in a variable, it is important to avoid estimating the bicoherence for periods which are approximately multiples of the oscillation periods present in the signal. The map of wavelet coefficients as in Figure 4.25 can help learning up to which scale the bicoherency should be estimated. From Figure 4.25 it can be observed that the energy of the signal is mostly concentrated in scales up to 41. If there is phase coupling in the signal it should be between these scales. The bicoherency can be estimated for scales close to 41. If no high value bicoherency is detected at these scales, it means the oscillations are linear. Therefore, it is better to avoid estimating bicoherency for large scales which could be multiples of the main oscillatory components.

## 4.7 Case study

The introduced tools are used for diagnosis of a unit-wide oscillation through an oil sands industrial data set. The analysis consists of detecting oscillatory variables, estimating multiple oscillation frequencies and finally diagnosing the source of the oscillation. The data set contains the controller outputs as well as the process values (PV) of 24 loops. The detailed analysis of 5 of the loops carrying similar oscillations is presented here. The variables are denoted as Loop 1 to Loop 4 and Loop 30. Figure 4.26 plots 4 PVs along with the average of absolute wavelet coefficients at scales from 1 to 300. The average is taken over the time domain in order to compare the strength of different scales in general. Local peaks in these plots correspond to oscillations in variables.

From the right column of Figure 4.26, we can observe that all the 4 variables have similar pattern of wavelet coefficients. Similar pattern indicates that all the 4 loops are carrying similar oscillations and suffer from same problem. The source of the oscillation could be one of these loops or other variables within the process which have similar pattern. The first step in oscillation diagnosis is to categorize the oscillation meaning to detect if the oscillation is due to controller tuning, valve nonlinearity or external disturbance.

In order to categorize the oscillation, we first examine the ACF of the variables which are plotted in Figure 4.27. The ACFs of the variables show presence of an oscillation. Even though the amplitude of the oscillation is very small in the ACFs, it does not decay to 0. The oscillation is steady with almost constant amplitude at all the time lags which is not similar to the ACF of an oscillation induced by controller tuning. Therefore, the oscillation could be due to an external disturbance to these loops or could be due to nonlinearity in one of them. The next step is to examine the presence of nonlinearity in these 4 loops. Figure 4.28 to 4.31 plot the bicoherence of the 4 loops. Since scale 300 is too large for estimation of bicoherence, the variables are down-sampled by a factor of 2. Again, similar pattern in the plot of averaged bicoherence values of the 4 loops is observed. The average is with respect to time.

Figure 4.32 plots the controller output of Loop 30 and its wavelet coefficients. Oscillation detection algorithm finds an oscillation with period of 300 with 0 standard deviation from the ACF of the variable. The averages of absolute wavelet coefficients are also plotted in Figure 4.34 to observe which scales contain the highest energy.

Since the oscillation with 300 samples per period is dominant in the variable, the other oscillations which can be seen in Figure 4.34 are not detected from the ACF. Running the oscillation detection algorithm [38] individually on wavelet coefficients

of scales 126 and 63 results in the following estimations:

<i>Period</i>	<i>STD</i>
150	2.8
75	1.4

All the bicoherence plots of the 5 variables show a peak at the scale around 120 which approximately corresponds to oscillation period of 300 samples. The bicoherence estimated for the controller output of Loop 30, plotted in Figure 4.33, is much larger compared to the other 4 loops. This shows that Loop 30 is closer to the source of the oscillation while the other 4 loops receive the oscillation from the external source (propagated by Loop 30). Also, similar bicoherence value for Loop 1 to Loop 4 indicates that they are influenced by Loop 30 in a similar way.

## 4.8 Summary

In this chapter, a comprehensive algorithm is proposed which is capable of both detection and diagnosis of individual oscillatory components of variables in the presence of multiple oscillations and non-stationary signals. The independent diagnosis of multiple oscillations is viable due to the inherent capability of wavelet transform in decomposing the variables to its components of different frequencies. Two hypothesis tests are proposed based on the properties of wavelet bicoherence and wavelet power spectrum for the purpose of oscillation diagnosis.

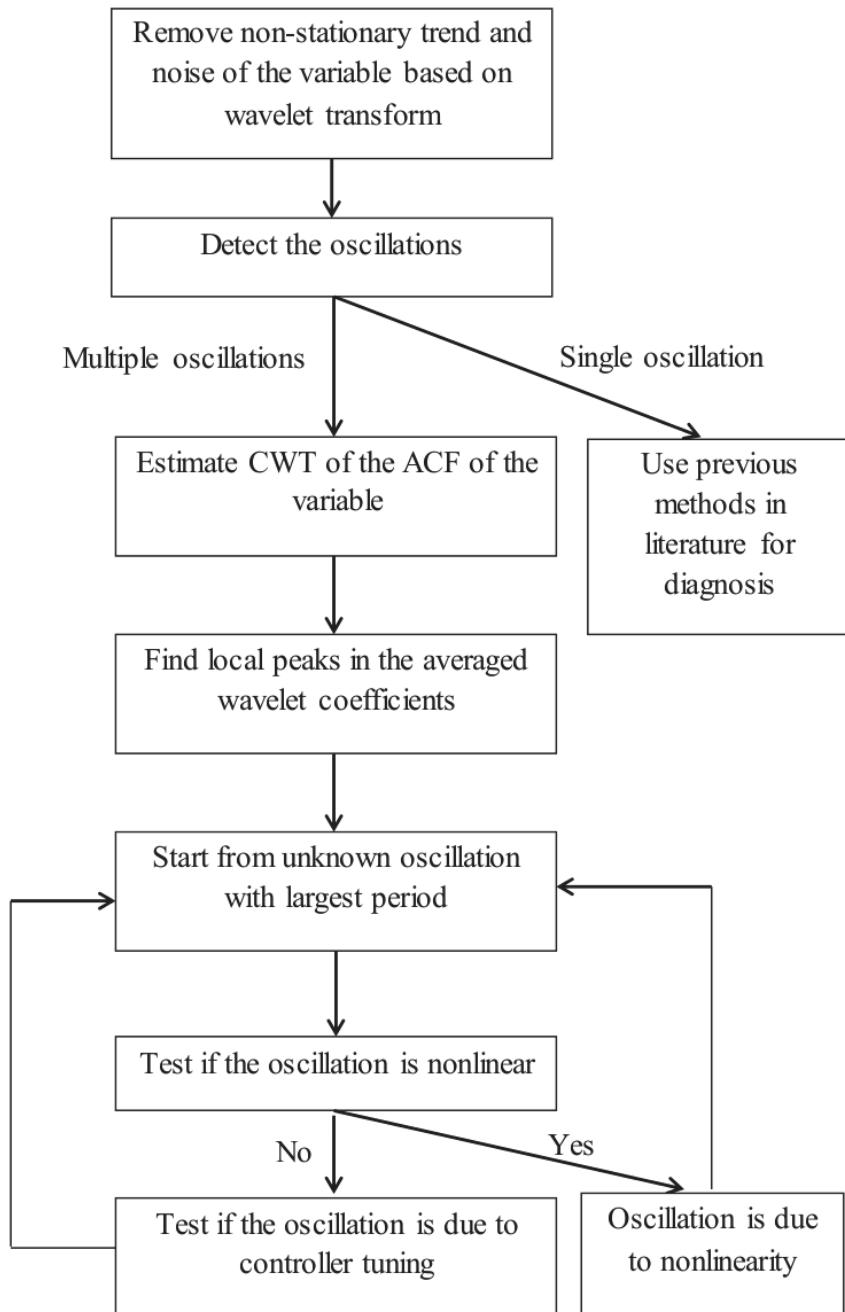


Figure 4.22: Required steps for oscillation diagnosis

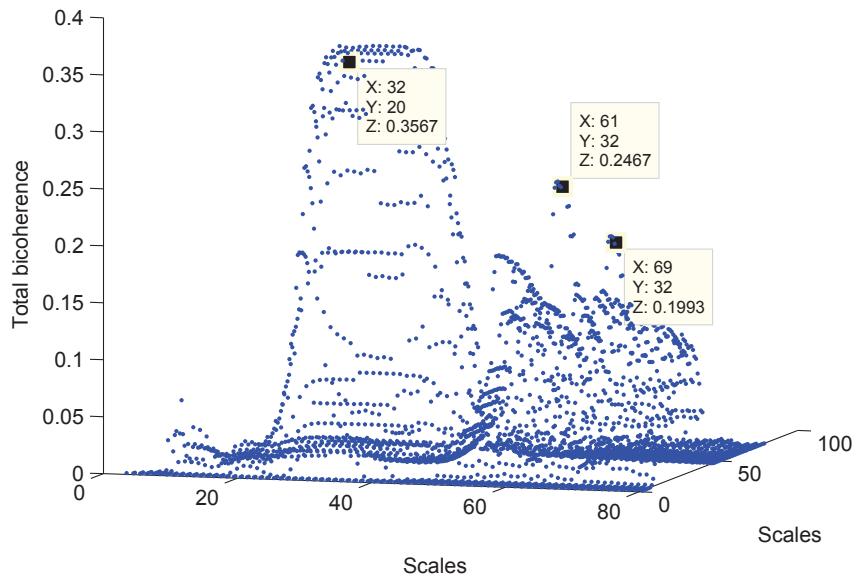


Figure 4.23: Bicoherence of signal  $x(t)$  for scales up to 80

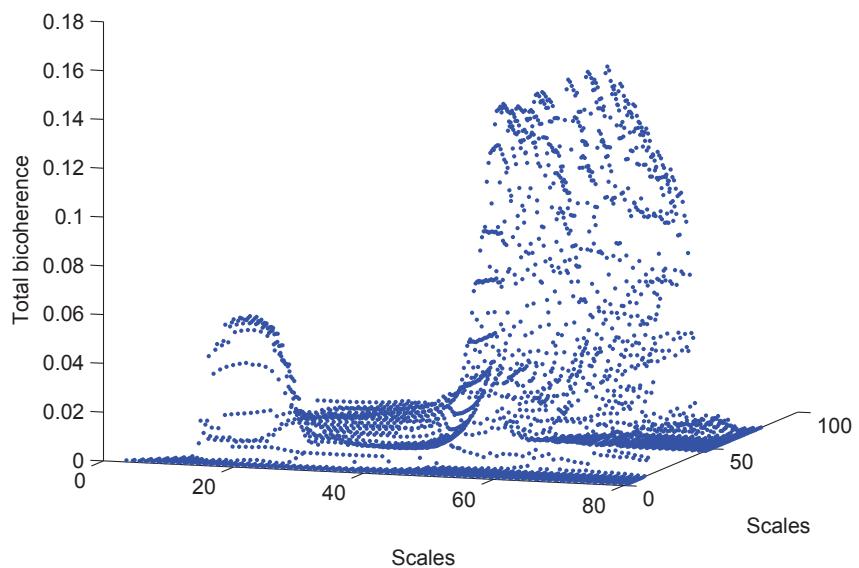


Figure 4.24: Bicoherence of signal  $x(t) = \sin(2\pi \times 0.04 + \frac{\pi}{3}) + \sin(2\pi \times 0.025 + \frac{\pi}{12})$  for scales up to 80

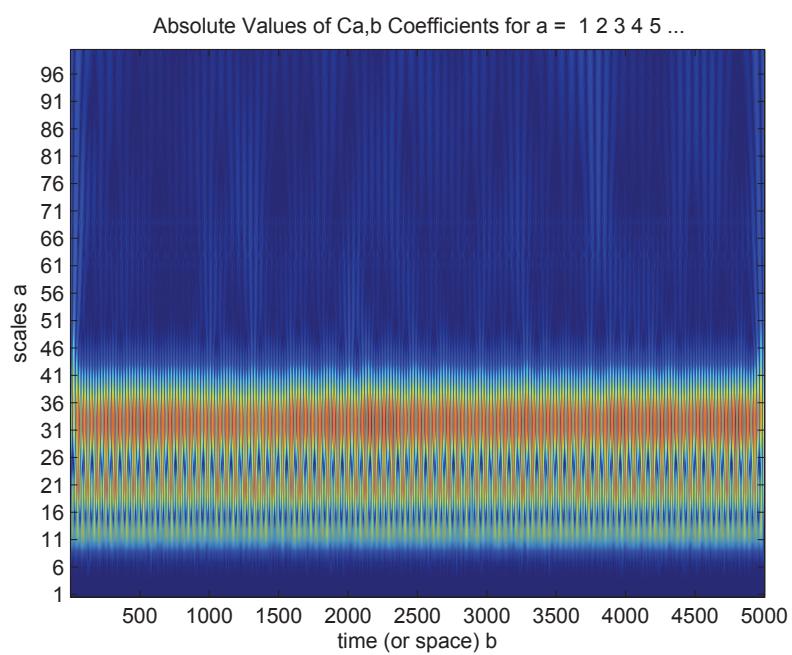


Figure 4.25: Wavelet coefficients of signal  $x(t) = 2\sin(2\pi \times 0.04 + \frac{\pi}{3}) + 2\sin(2\pi \times 0.025 + \frac{\pi}{12}) + 2\sin(2\pi \times 0.065 - \frac{\pi}{4}) + \varepsilon(t)$

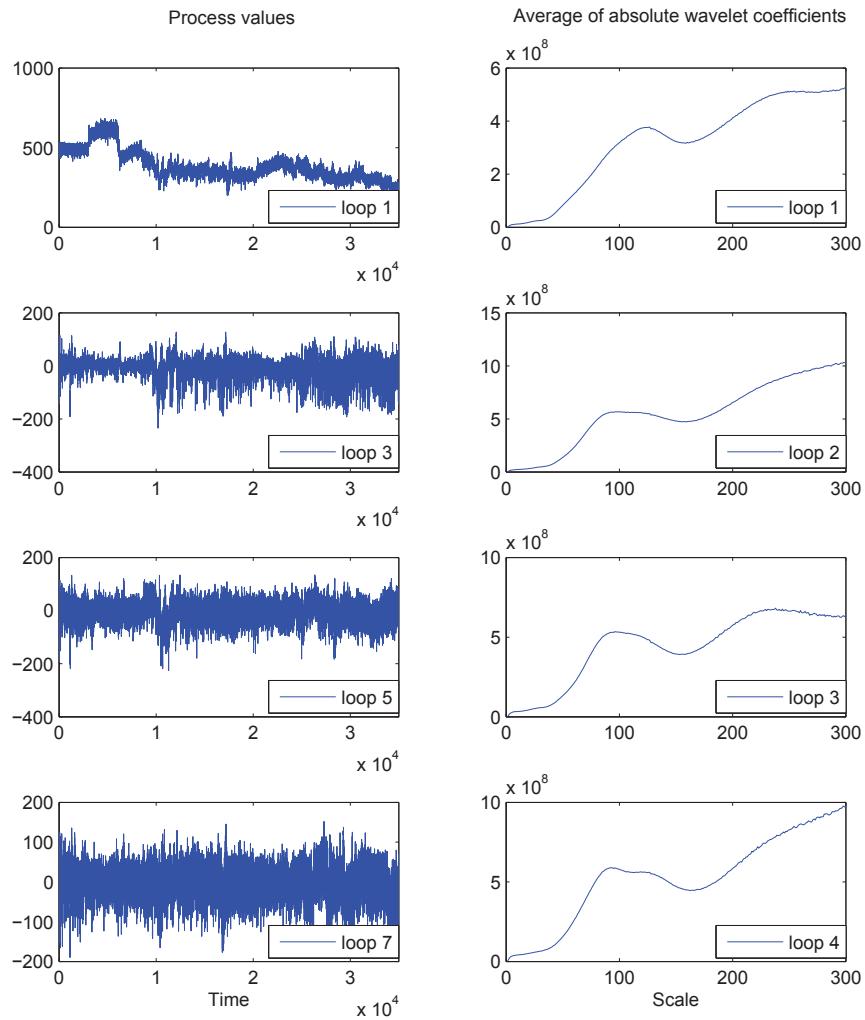


Figure 4.26: 4 industrial variables along with the absolute wavelet coefficients

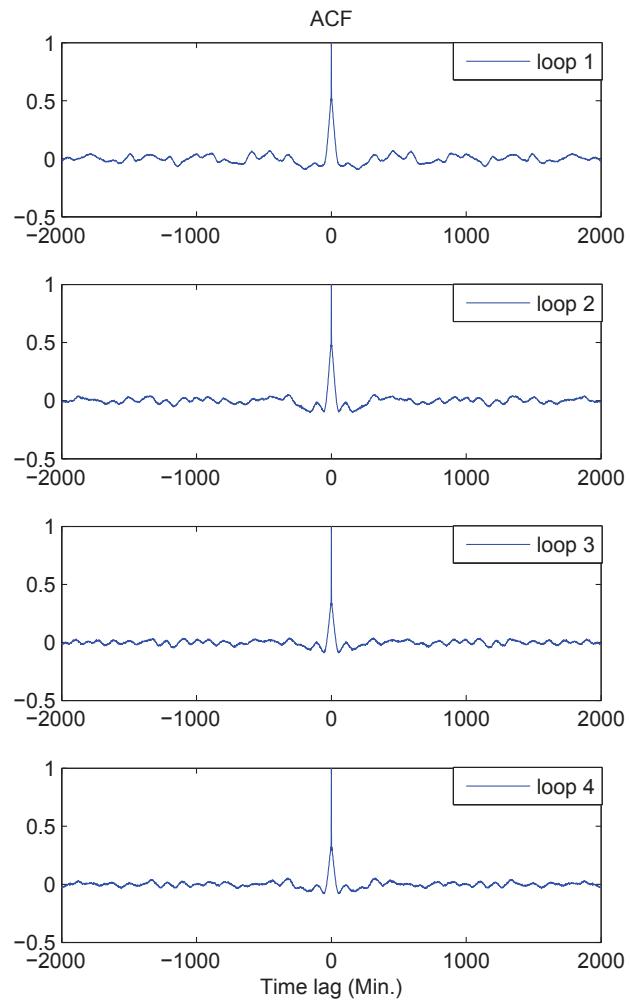


Figure 4.27: ACFs of the 4 industrial variables

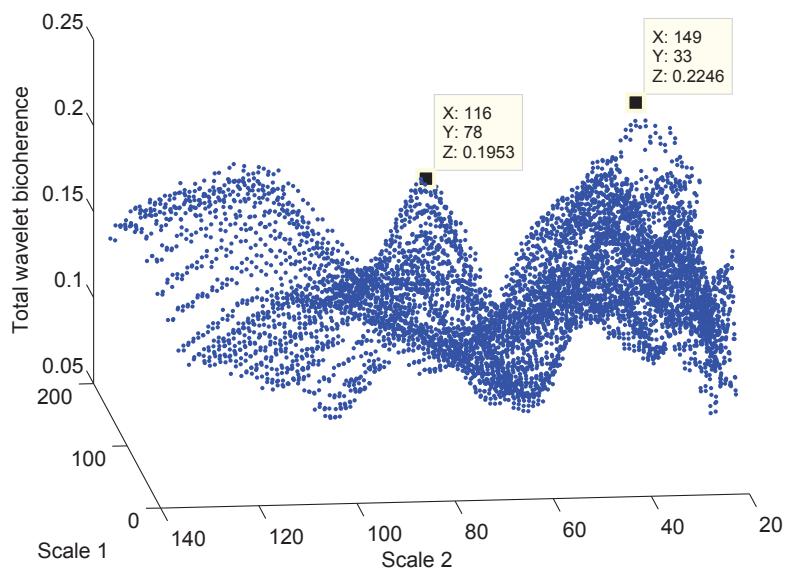


Figure 4.28: Wavelet bicoherence of down-sampled Loop 1 by factor of 2

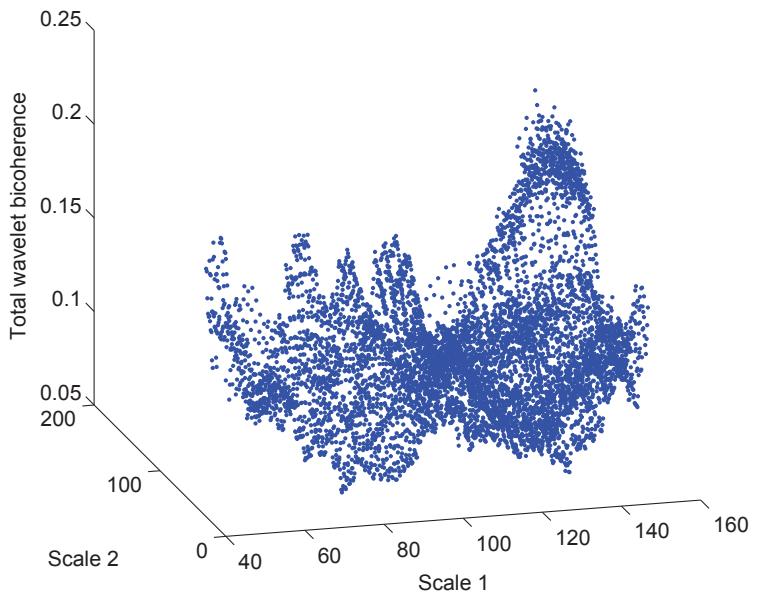


Figure 4.29: Wavelet bicoherence of down-sampled Loop 2 by factor of 2

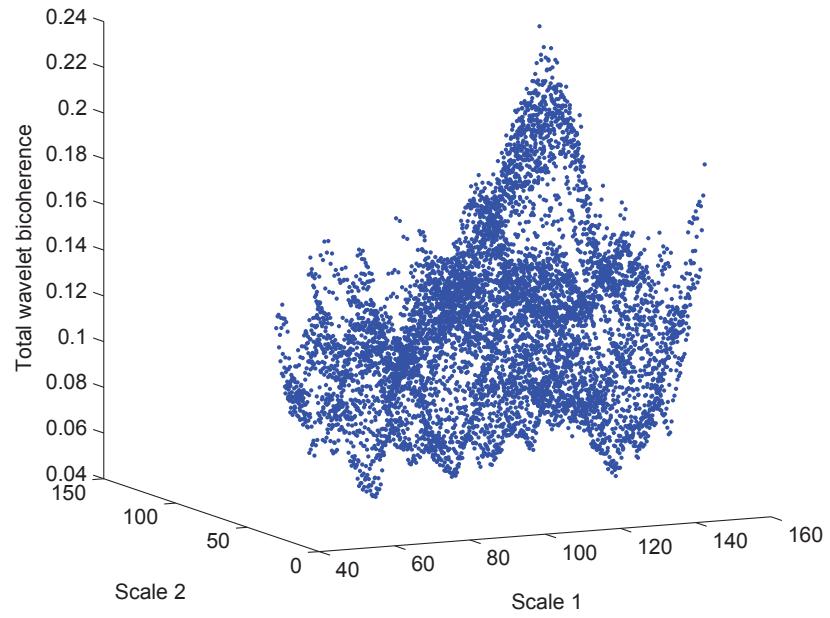


Figure 4.30: Wavelet bicoherence of down-sampled Loop 3 by factor of 2

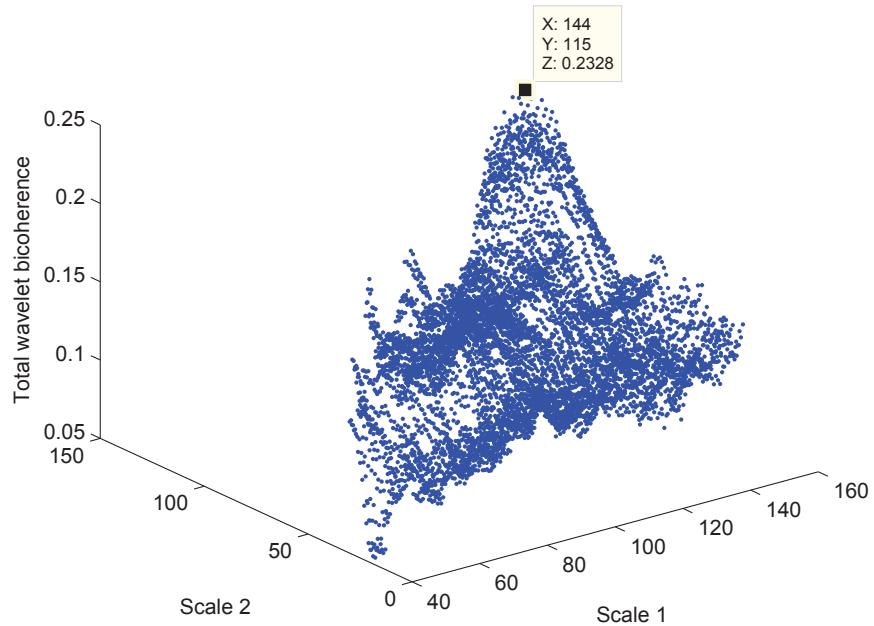


Figure 4.31: Wavelet bicoherence of down-sampled Loop 4 by factor of 2

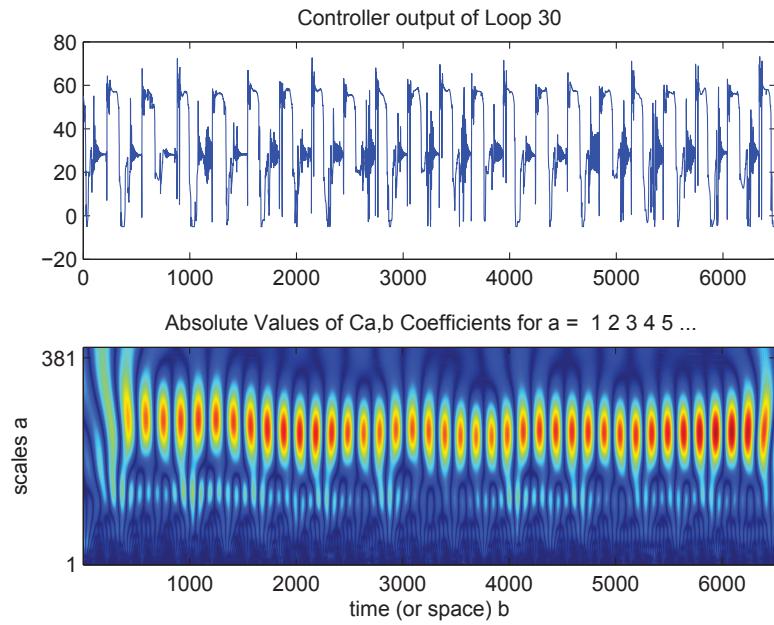


Figure 4.32: Top: Controller output of Loop 30 in time. Bottom: Wavelet coefficients

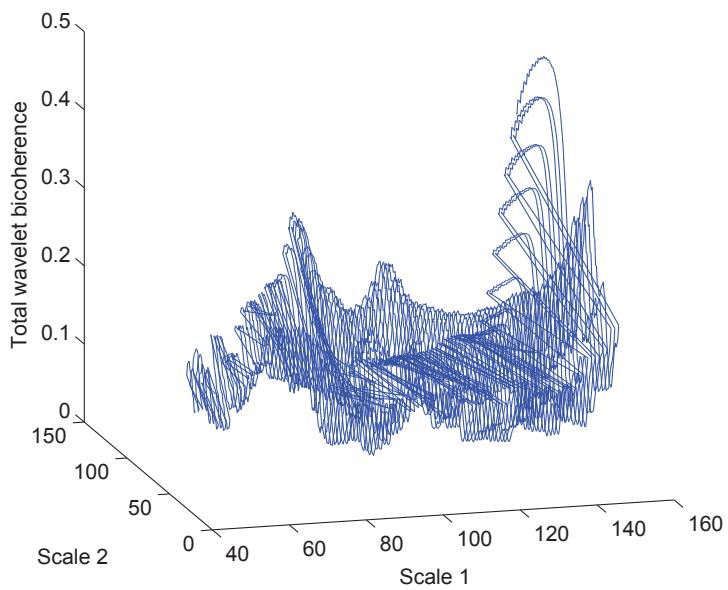


Figure 4.33: Wavelet bicoherence of down-sampled Loop 30 by factor of 2

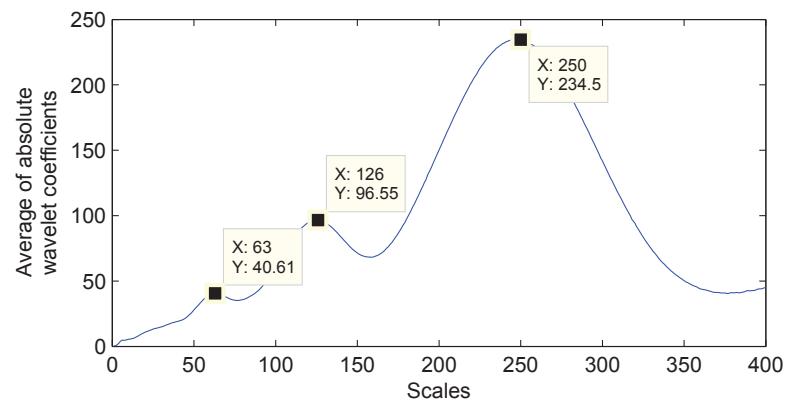


Figure 4.34: Average of absolute wavelet coefficients of controller output of Loop 30

# **Chapter 5**

## **Interaction analysis of multivariate control systems under Bayesian framework**

## 5.1 Abstract

Detection and quantification of interactions between the loops of a multivariate system is of interest for various purposes such as control system design, optimization, fault diagnosis and performance assessment. This chapter proposes a new method for interaction analysis based on decomposing the estimated transfer function between variables in the form of impulse response coefficients. The method not only provides an estimation of the direct (feedback and interaction free) transfer function between the variables, but also provides a measure of strength of all the indirect paths connecting variables together individually. The advantage of the method is that it provides a complete picture of the different paths through which variables can influence each other along with an estimation of the energy transferred through each path independently. The analysis is performed by estimating Structural Vector Autoregressive models under Bayesian framework. Bayesian approach provides certain advantages in terms of dealing with high dimensional variables and over parameterization problem. An appropriate design of the prior probability for the model parameters also better ensures convergence to a physically interpretable model. A procedure to design the prior distribution for the model parameters is presented here.

## 5.2 Introduction

Interaction analysis of multivariable control systems has been extensively studied in the literature mostly for the purpose of input/output pairing as a preliminary stage of control system design. The most well known method which is called relative gain array (RGA) [75] and its more advanced forms have helped engineers in the design of control system. Van de Wal et al. [76] provides a review on the developed interaction analysis methods specifically for the purpose of input/output selection and [77] provides a review on the Gramian interaction analysis methods. These methods are mainly based on the available plant model and in some cases with the assumption of existence of perfect controllers.

Interaction analysis has also been applied in fault diagnosis and hazard analysis where the term causality analysis is mostly used instead. In fault diagnosis it is required to understand how a fault occurred in one loop can propagate through the process and influence other variables. Process connectivity charts developed based on P&IDs (piping and instrumentation diagrams) or knowledge of physical principles governing the process aid the engineers in detecting the interactions within the process [15]. However, causality analysis based on the physical principles or qualitative

knowledge may not be always possible and it does not provide quantitative measures to assess the strength of the interactions between variables. This fact has motivated interaction or causality analysis based on historical data.

The focus of this work is on causality analysis based on routine operational data targeting its application for performance assessment and fault diagnosis. Much work has been done in the literature regarding causality analysis based on historical data [78, 79]. Various data-based methods can be applied to detect causality such as Granger causality, information theory based methods, Bayesian network etc.. Yang et al. [80] provides a review on some of the well known methods of causality analysis specifically applied in engineering. The main difference between data based causality analysis methods and other methods such as RGA is that they analyze routine operation data without any assumption regarding the controller or plant model. Also, these methods are applicable to any kind of variables and are not limited to control system design. Another important difference is the asymmetrical inference regarding the causality between two variables.

Causality between two variables is defined by Granger [18] based on two conditions. The cause variable happens before the effect and has some unique information about it which can improve the prediction of the effect variable. Autoregressive modeling framework is the most suitable modeling structure to detect causality based on Granger's definition. Considering two signals  $x_1$  and  $x_2$ , two models are built to predict  $x_1(t)$  as shown in Equations 5.1 and 5.2. A test statistic is defined in order to quantify the improvement in the prediction of  $x_1(t)$  by considering the history of  $x_2(t)$ , based on comparing the variances of modeling residuals [81].

$$x_1(t+1) = \sum_{j=0}^{\infty} a(j)x_1(t-j) + \varepsilon_1(t) \quad (5.1)$$

$$x_1(t+1) = \sum_{j=0}^{\infty} \alpha_{11}(j)x_1(t-j) + \sum_{i=0}^{\infty} \alpha_{12}(i)x_2(t-i) + \varepsilon_{1|2}(t) \quad (5.2)$$

A test statistic is defined as  $F_{2 \rightarrow 1} = \ln \frac{var(\varepsilon_1(t))}{var(\varepsilon_{1|2}(t))}$  based on variances of  $\varepsilon_{1|2}(t)$  and  $\varepsilon_1(t)$  [81].  $F_{2 \rightarrow 1}$  always has a positive value and follows a chi-squared distribution when there is no causality from  $x_2$  to  $x_1$ . Reverse of the same analysis is required to infer if  $x_1(t)$  is causal to  $x_2(t)$ .

Limitation of this test of Granger causality is that it can analyze only 2 or 3 variables simultaneously. Blinowska et al. [82] has discussed that the application of pair-wise Granger causality in multivariate processes produces spurious results and therefore the application of directed transfer function [83] and directed coherency [84]

is more appropriate. These quantities are based on estimating vector auto-regressive (VAR) models for the variables as defined in Equation 5.3.

$$y_t = y_{t-1}A_1 + \dots + y_{t-p}A_p + e_t \quad (5.3)$$

where  $y_t$  is a  $1 \times m$  matrix of observations ( $m$  is the number of response variables),  $A_l$  is a  $m \times m$  matrix of coefficients and  $e_t$  is  $1 \times m$  matrix of innovations with a diagonal covariance matrix. Existence of direct causality between  $i_{th}$  and  $j_{th}$  variable in vector  $y_t$  can be verified by checking the elements of  $A_l$  matrices. If at least for one  $A_l$  matrix (time lag equal to  $l$ ) the element  $a_{ij}(l)$  is non-zero, there is a direct causal relation from  $y_j$  to  $y_i$  [85]. Zero coefficients in a VAR model is only a sign of non-existence of direct causality and does not tell if there is an indirect causality relation or not. Indirect causality is defined to describe the situation where there is no direct relation between two variables, but the variables are related through a third or even more intermediate variables.

The problem with using VAR models for causality detection is the fact that causality inference depends on the prediction horizon when more than two variables are considered in the analysis [86]. Thus vector moving average (VMA) representation of the VAR model is more suitable to infer causality between two variables in the presence of other variables. The VMA representation of the model in Equation 5.3 can be obtained by inverting the stable VAR model as:

$$y_t = e_t + \phi_1 e_{t-1} + \phi_2 e_{t-2} + \dots \quad (5.4)$$

where  $\phi_l$ s are  $m \times m$  dimensional matrices. Non-zero VMA coefficients between any two variables are a sign of causality. It should be noted that the causality inferred from VMA model represents both direct and indirect causality [85]. Prediction error variance decomposition based on the estimated VMA model or impulse response analysis can quantify the strength of the relations between variables.

As a directionality measure, Gigi et al. [87] proposed the decomposition of the transfer function between variables into direct and indirect terms based on the estimated VAR model. The work is based on frequency domain spectral factorization of the cross spectral density matrix. The total transfer function between two variables  $y_i(t)$  and  $y_j(t)$  is written as

$$h_{ij}(\omega) = h_{D,ij}(\omega) + h_{I,ij}(\omega) \text{ for } i \neq j \quad (5.5)$$

where the total transfer function ( $h_{ij}(\omega)$ ) and the direct transfer function ( $h_{D,ij}(\omega)$ ) are estimated based on the identified VAR model and the indirect transfer function

is estimated as the subtraction of the two. The direct transfer function is devoid of the influence of any feedback loop in the system and estimates the transfer function between the two variables when all the loops are open. The indirect transfer function combines the transfer functions of all the indirect paths from which variable  $j$  can reach variable  $i$  when all the loops are closed.

This work proposes a method to decompose the transfer function between variables to individual transfer functions, each representing a specific path from which the two variables can influence each other. The advantage of the proposed method is that it also decomposes the indirect transfer function between variables to its individual components instead of providing an estimation of the combination of all the indirect transfer functions together. This is performed based on decomposing the transfer functions in the form of impulse responses.

The decomposed impulse responses between variables can be utilized to estimate the energy or variation transferred between variables through different paths. For example if there are 3 different paths for variable  $i$  to influence variable  $j$  (one directly from  $i$  to  $j$ , one through variable  $l$ , and one through variable  $k$ ) the proposed methodology estimates the amount of variance that is transferred directly from  $i$  to  $j$  along with the amount of variance that is transferred through variables  $l$  and  $k$  independently. Therefore, it provides better insight for the engineers about the different paths through which one variable is being influenced by another one. From the performance assessment perspective, it can measure how good the controller is in rejecting the disturbance felt from outside the loop.

The impulse response coefficients are estimated under Bayesian framework which provides the possibility to incorporate prior knowledge in model estimation. As will be described later, proper design of the prior distribution of the model parameters can deal with high dimensionality and over-parametrized models. A proper design of prior probability can also better ensure convergence of the estimated model to the physically interpretable model among several models with similar prediction performance.

The remainder of the chapter is organized as follows. Section 5.3 discusses the appropriate model structure for causality analysis of engineering process variables. Section 5.5 explains the model estimation procedure under Bayesian framework and Section 5.6 presents the proposed method to design prior distribution. Section 5.7 reviews impulse response estimation method along with the method to estimate confidence intervals and Section 5.8 presents the new method in impulse response decomposition along with a simulated example. An experimental case study is presented in Section 5.9. Conclusions are drawn in Section 5.10.

### 5.2.1 Literature review on causality analysis methodologies with the assumption of linearity

Following is a brief review of the methods for multivariate causality analysis assuming linear relationships between variables. The most common approach to examine relations between some recorded variables is correlation analysis. However, correlation is different from causation since existence of correlation cannot indicate if there is any causality relation between the variables. Covariance matrix between variables is utilized in different ways to examine correlation and also causation.

The frequency domain counterpart of the covariance between two variables is called cross spectrum and is usually used to verify the strength of correlation between variables at different frequencies. Assume a multivariate time series matrix as  $X(t) = [X_1(t), X_2(t), \dots, X_k(t)]^T$ . The corresponding power spectral density matrix is defined as

$$S(f) = X(f)X(f)^* \quad (5.6)$$

The off-diagonal elements are cross spectrum and the diagonal elements are called auto spectrum. The ordinary coherence ( $K$ ) between two variables can be obtained from the spectral matrix which is a measure of the similarity between components of the two variables at different frequencies.

$$K(X_i X_j, f) = K_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}} \quad (5.7)$$

Ordinary coherence has the problem to distinguish direct and indirect relations between the variables. Two variables could have a high coherence value due to an intermediate variable. Therefore, partial coherence function is utilized to estimate the direct coherence between two variables while the effect of other variables is removed.

$$C_{ij}(f) = \frac{M_{ij}(f)}{\sqrt{M_{ii}(f)M_{jj}(f)}} \quad (5.8)$$

where  $M_{ij}$  is the determinant of  $S$  when the  $i^{th}$  row and  $j^{th}$  column are removed.  $C_{ij}$  can also be expressed in terms of the inverse of the spectral matrix  $d_{ij} = [S^{-1}]_{ij}$

$$C_{ij}(f) = (-1)^{i+j} \frac{d_{ij}(f)}{\sqrt{d_{ii}(f)d_{jj}(f)}} \quad (5.9)$$

These estimations of partial coherency between variables are based on non-parametric methods which have limitations and uncertainties. Identifying parametric linear models for the data and eventually parametric spectrum has some advantages over non-parametric estimation of power spectrum utilizing Fourier transform. The main disadvantage of a Fourier approach is in its estimation procedure, which assumes the signal is periodic and also multiplies a window to the data which causes distortion.

In parametric modeling, AR models are considered as follows:

$$X(t) = \sum_{j=1}^p \bar{A}(j)X(t-j) + E(t) \quad (5.10)$$

Let  $\bar{A}(0) = I$  and  $A(j) = -\bar{A}(j)$ , the above equation can be written as follows:

$$E(t) = \sum_{j=0}^p A(j)X(t-j) \quad (5.11)$$

By transforming the above equation to the frequency domain, we get

$$\begin{aligned} E(f) &= A(f)X(f) \\ X(f) &= A^{-1}(f)E(f) = H(f)E(f) \end{aligned} \quad (5.12)$$

$$(5.13)$$

The spectral matrix can be written in terms of  $H(f)$  as follows:

$$S(f) = X(f)X^*(f) = H(f)E(f)E^*(f)H^*(f) = H(f)VH^*(f) \quad (5.14)$$

where  $V$  is the input noise covariance matrix.

Directed transfer function (DTF) is defined based on the elements of  $H(f)$  as

$$\gamma_{ij}^2(f) = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^k |H_{im}(f)|^2} \quad (5.15)$$

DTF represents the ratio of the total information passed from variable  $j$  to  $i$  from all paths, to the total energy transferred to  $i$  from all the variables. The assumption is about identity noise covariance matrix.  $|H_{ij}(f)|^2$  contains both direct energy transferred from  $j$  to  $i$ , and also indirect energy through intermediate variables. Therefore, it cannot be utilized to determine if there is any direct relationship between variables. On the other hand, partial directed coherence (PDC) from variable  $j$  to  $i$  is defined as:

$$\pi_{ij}(f) = \frac{\bar{a}_{ij}(f)}{\sqrt{\bar{a}_j^*(f)\Sigma^{-1}\bar{a}_j(f)}} \quad (5.16)$$

PDC can detect direct relationship between variables and thus is capable of detecting the structure of the data. However, it cannot quantify the strength of the relationships between the variables.

Gigi et al. [78] proposed the division of the energy transferred between two variables as direct, indirect and interference energies. The work is based on frequency

domain spectral factorization of the cross spectral density matrix, which contains information about directional interactions between any two pairs of a multivariate process. The jointly stationary multivariate process is represented as  $X = [x_1, x_2, \dots, x_m]$ . The cross power spectral density of  $X$  ( $\Phi_{xx}(f)$ ) can be factored as

$$\Phi_{xx}(f) = H(f)\Sigma_e H^*(f) \quad (5.17)$$

where  $\Sigma_e$  is the covariance matrix of the white noise input to the process and  $H(f)$  is the transfer function matrix in frequency domain.

The elements of the transfer function matrix  $H(f)$  represent a combination of both direct and indirect connections between variables, and therefore cannot be utilized in identifying the structural connections between the variables. Decomposition of  $H(f)$  into direct and indirect terms is possible based on estimating a VAR model for the process.

The total transfer function between two variables  $x_i$  and  $x_j$  can be written as

$$h_{ij}(f) = h_{D,ij}(f) + h_{I,ij}(f) \text{ for } i \neq j \quad (5.18)$$

which is obtained as

$$h_{ij}(f) = \frac{(adj(\bar{A}))_{ij}}{det(\bar{A})} \quad (5.19)$$

while the direct transfer function can be written as [78]

$$h_{D,ij}(f) = \frac{(-1)^{i+j}\bar{a}_{ij}(f)det(\bar{M}_{ji}(f))}{det(\bar{A}(f))} \quad (5.20)$$

$\bar{M}_{ji}(f)$  is the minor of matrix  $\bar{A}(f)$  obtained by eliminating  $i_{th}$  row and  $j_{th}$  column of  $\bar{A}(f)$ . The indirect transfer function can be obtained as the subtraction of the direct transfer function from the total transfer function as

$$h_{I,ij}(f) = h_{ij}(f) - h_{D,ij}(f) \quad (5.21)$$

The total energy transfer between the two variables can be written as in the following equation derived from Equation 5.62.

$$|h_{ij}(f)|^2 = (h_{D,ij}(f) + h_{I,ij}(f))(h_{D,ij}^*(f) + h_{I,ij}^*(f)) = |h_{D,ij}(f)|^2 + |h_{I,ij}(f)|^2 + 2R(h_{D,ij}(f)h_{I,ij}^*(f)) = |h_{D,ij}(f)|^2 + |h_{I,ij}(f)|^2 + |h_{IF,ij}(f)|^2$$

where  $R$  denotes the real part of the term and  $|h_{IF,ij}(f)|^2 = 2|h_{D,ij}(f)||h_{I,ij}(f)||\cos(\phi_D - \phi_I)|$  is the interference term. The interference occurs due to the interaction between the two direct and indirect energies. The following section discusses the appropriate model structure for engineering process variables to be used for causality analysis based on the reviewed methodologies.

## 5.3 Model structure

The most common modeling structure for causality analysis using linear methods is vector auto regressive where the response variables are modeled as a function of the history of themselves along with the history of other predictor variables with an unexplained residual as in Equation 5.3. Causality analysis based on VAR modeling has certain problems as dependence on the prediction horizon. Thus, VMA models are used in order to check if an impulse shock in one variable has any effect on other variables. If there is a reaction in one variable to an impulse in a second variable, then the second one is called causal to the first variable.

If the modeling residuals in an estimated VAR model have simultaneous correlation with each other (noise covariance matrix  $\Sigma_e$  is not diagonal), VMA representation of the VAR model will not be unique and cannot tell about how a shock in one variable affects other variables. The assumption in analyzing impulse responses to infer causality is that it is possible to insert an impulse in only one variable and verify its effect on other variables. However, it requires the variables to have independent innovation terms from each other. In the framework of control systems, instantaneous relations exist due to the feedback loops.

Assuming that  $u_t$  (controller output) and  $y_t$  (process output) belong to a feedback loop,  $u_t$  is usually an instantaneous function of  $y_t$  due to the proportional gain of the controller. However,  $y_t$  is typically not an instantaneous function of other variables in the system. Therefore, the true model structure for the two variables can be written as follows:

$$\begin{aligned} y_t &= ZA_y + e_{y,t} \\ u_t &= y_t k_p + ZA_u + e_{u,t} \end{aligned} \tag{5.22}$$

where  $Z$  matrix includes all the regressors with minimum 1 sample time lag including lagged  $u_t$  and  $y_t$ .  $A_y$  and  $A_u$  are the coefficient matrices.  $e_{u,t}$  and  $e_{y,t}$  are respectively controller output noise and process output noise.  $k_p$  is the proportional gain of the controller with appropriate sign.

Modeling  $y_t$  and  $u_t$  simultaneously in a VAR structure based on historical data (ignoring  $y_t$  in the right hand side of Equation 5.22) will result in correlated modeling residuals ( $\hat{e}_{u,t} = k_p e_{y,t} + e_{u,t}$  will be correlated with  $\hat{e}_{y,t} = e_{y,t}$ ). Therefore, a VAR model including both  $y_t$  and  $u_t$  will have non-diagonal noise covariance matrix. Correlation between the estimated residuals prevents correct detection and quantification of interactions between the variables. The impulse response estimated from this VAR model is not meaningful since the disturbances are linear combinations of individual

disturbances which are affecting  $y_t$  and  $u_t$  independently.

A common solution to the problem of instantaneous correlation between modeling residuals in the literature is to diagonalize the covariance matrix of the modeling residuals through various transformations such as Choleski factorization. The problem with applying transformations to the data is the fact that there is no unique way of transforming the data and the result of causality analysis on the transformed data cannot be interpreted in the original domain.

The other approach to the problem of instantaneous correlation between variables is application of structural vector autoregressive (SVAR) models. SVAR formulation is similar to VAR with an additional  $A_0$  matrix multiplying to Equation 5.3 as is shown in Equation 5.23.

$$y_t A_0 = y_{t-1} A_1^* + \dots + y_{t-p} A_p^* + \varepsilon_t \quad (5.23)$$

where  $A_j^* = A_j A_0$  ( $j = 1, \dots, p$ ) and  $\varepsilon_t = e_t A_0 \sim (0, \Sigma_\varepsilon = A_0' \Sigma_e A_0)$ .  $\Sigma_e$  is the covarince matrix of the reduced form of the model which, in general, is not diagonal due to the simultaneous correlation between variables while,  $\Sigma_\varepsilon$  can be diagonal with a proper choice of  $A_0$  matrix.  $A_0$  matrix is required to be invertible. The diagonal elements of the  $A_0$  matrix could be equal to 1 to allow for different noise variances in different equations or could be scaling factors in order to normalize the residuals to have unity variances.

The advantage of using SVAR framework is the possibility to diagonalize the noise covariance matrix by proper design of the  $A_0$  matrix. However,  $A_0$  and  $A_j$  matrices are not simultaneously identifiable from the data without the help of prior knowledge. The reason for this identification problem is the fact that there are infinite combinations of  $A_0$  and  $A_j$  matrices which result in exactly the same probability distribution of the data. To see this fact, consider multiplying the model in Equation 5.23 by a positive definite matrix  $Q$  which results in  $y_t A_0 Q = y_{t-1} A_1^* Q + \dots + y_{t-p} A_p^* Q + \varepsilon_t Q$ . The reduced form of this model is

$$\begin{aligned} y_t &= y_{t-1} A_1^* Q (A_0 Q)^{-1} + \dots + y_{t-p} A_p^* Q (A_0 Q)^{-1} + \varepsilon_t Q (A_0 Q)^{-1} \\ &= y_{t-1} A_1^* A_0^{-1} + \dots + y_{t-p} A_p^* A_0^{-1} + \varepsilon_t A_0^{-1} \end{aligned} \quad (5.24)$$

which is exactly the same as the reduced form of the model in Equation 5.23. Therefore, the data does not provide any evidence regarding which one of these two models is the true one.

In order to uniquely identify  $A_0$  matrix from data,  $\frac{m(m-1)}{2}$  ( $m$  is number of response variables) restrictions are required which should come from the prior knowledge

or preliminary analysis of the data. Usually  $A_0$  is considered to be lower triangular with  $\frac{m(m - 1)}{2}$  elements equal to 0. However,  $A_0$  matrix can have different patterns. Since in chemical operations, the simultaneous relations are due to the feedback loops, prior knowledge of the variables which are in a loop helps to design the  $A_0$  matrix. It should be noted that  $A_0 = I$  does not mean there is no feedback in the system, but it means the feedbacks are lagged and are not instantaneous.

In a feedback loop, the controller output can be a simultaneous function of the process output but not vice versa. For a multivariate system, outputs of the loops cannot be influenced by other variables with 0 time lag and therefore, the elements in the row of  $A_0$  matrix corresponding to outputs are 0 except for the diagonal element. Controller output of each loop has a term proportional to the current value of the process output due to the proportional gain of the controller. Thus, each row in  $A_0$  matrix corresponding to a controller output has two non-zero elements. The diagonal elements equal to 1 (or scaling factor) and there is an unknown coefficient corresponding to the proportional feedback gain from the respective process output. For example if there are  $k$  loops with controller output noted as  $u_i$  and process output as  $y_i$  ordered as  $[y_1(t)u_1(t)y_2(t)u_2(t)\dots y_k(t)u_k(t)]'$ , the format of the  $A_0$  matrix will be as follows:

$$A_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ \gamma_{11} & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \gamma_{22} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \gamma_{kk} & 1 \end{bmatrix} \quad (5.25)$$

where  $\gamma_{ii}$  is the proportional gain of the  $i_{th}$  loop with appropriate sign.

This design of the  $A_0$  matrix makes the system identifiable since there are sufficient restrictions to uniquely estimate the model parameters and innovations. It is also important to note that the estimated parameters and innovations are physically meaningful. The residuals in each equation are the measurement noise of the corresponding response variable which is naturally independent from other noises.

It should be noted that considering controller outputs in feedback loops in addition to the process outputs in the analysis, provides valuable information specially useful to assess the performance of the controller. However, if the controller output is a perfect function of the process output without any noise, it is better to only consider the process output in the analysis. It should also be noted that persistent excitation in the variables is assumed. Since it is not always known which variables have instantaneous relation and which one is actually leading the other one, the following section proposes

a method to detect instantaneous relations between variables.

## 5.4 Detection of instantaneous feedback between variables

Prior knowledge of the feedback loops in the process is helpful in determining an identifiable model structure with uncorrelated noises. However, this information is not always available in addition to the fact that two variables may have correlated noises simply because of the similar disturbances they receive from other variables. Therefore, a method is required to detect the variables in feedback loops based on the data in a preprocessing step.

This section proposes a method based on examining the covariance matrix of the modeling residuals to detect presence of instantaneous feedback between variables as well as learning which variable is actually a function of the other one. Assume the true model representing a feedback loop is:

$$\begin{aligned} y_t &= G_p(z^{-1})u_t + G_{ey}(z^{-1})e_{y,t} + e_{y,t} \\ u_t &= k_p y_t + G_c(z^{-1})y_t + G_{eu}(z^{-1})e_{u,t} + e_{u,t} \end{aligned} \quad (5.26)$$

where all the transfer functions have at least one sample time delay (the part with 0 time lag is separated). Optimal one step ahead prediction of  $u_t$  without considering  $y_t$  among the predictor variables, is obtained as

$$\begin{aligned} \hat{e}_{u,t} &= u_t - \hat{u}_{t|t-1} = k_p y_t + G_c(z^{-1})y_t + G_{eu}(z^{-1})e_{u,t} + \\ &\quad e_{u,t} - L_1(z^{-1})y_t - L_2(z^{-1})u_t \\ &= k_p[G_p(z^{-1})u_t + G_{ey}(z^{-1})e_{y,t} + e_{y,t}] + \\ &\quad G_c(z^{-1})y_t + G_{eu}(z^{-1})e_{u,t} + e_{u,t} - \\ &\quad L_1(z^{-1})y_t - L_2(z^{-1})u_t \end{aligned} \quad (5.27)$$

By replacing  $e_{y,t}$  and  $e_{u,t}$  with the following equations

$$e_{y,t} = \frac{y_t - G_p(z^{-1})u_t}{1 + G_{ey}(z^{-1})} \quad (5.28)$$

$$e_{u,t} = \frac{u_t - k_p y_t - G_c(z^{-1})y_t}{1 + G_{eu}(z^{-1})} \quad (5.29)$$

$\hat{e}_{u,t}$  will be as:

$$\begin{aligned}
\hat{e}_{u,t} = & [k_p G_p(z^{-1}) - \frac{k_p G_p(z^{-1}) G_{ey}(z^{-1})}{1 + G_{ey}(z^{-1})} + \frac{G_{ey}(z^{-1})}{1 + G_{ey}(z^{-1})}] u_t \\
& + [\frac{k_p G_{ey}(z^{-1})}{1 + G_{ey}(z^{-1})} + G_c(z^{-1}) - \\
& (k_p + G_c(z^{-1})) \frac{G_{eu}(z^{-1})}{1 + G_{eu}(z^{-1})} - L_1(z^{-1})] y_t \\
& + k_p e_{y,t} + e_{u,t}
\end{aligned} \tag{5.30}$$

Considering that all the transfer functions have at least one sample time delay, the minimum residual variance equals to  $k_p^2 var(e_{y,t}) + var(e_{u,t})$ .

Assuming the algorithm converges to the process model when modeling  $y_t$ , the minimum one step ahead prediction will be equal to  $e_{y,t}$ . Thus, the residuals of the modeling of  $u_t$  and  $y_t$  will have non-zero correlation. Correlation between  $\hat{e}_{u,t}$  and  $e_{y,t}$  makes it impossible to interpret the result of impulse response analysis since the disturbances are linear combinations of  $e_{u,t}$  and  $e_{y,t}$ . To remove this correlation,  $y_t$  should be added as a regressor to the model of  $u_t$ . The minimum one step ahead prediction residuals for  $u_t$ , while considering  $y_t$  as a predictor, will be obtained as  $var(e_{u,t})$  which results in zero correlation between the modeling residuals. Based on the knowledge of which two variables are in a loop, it is possible to design the  $A_0$  matrix in a proper way.

It should be considered that in some cases the knowledge of existence of instantaneous feedback between two variables may not be available beforehand. In such cases it is required to first learn the existence of instantaneous feedback between variables in order to be able to perform the analysis. So far it is proved that if it is known which variable is the controller output  $u_t$ , addition of the process output  $y_t$  with no time delay to the model of  $u_t$  makes the modeling residuals to be uncorrelated. Assuming there is no prior knowledge, we investigate the modeling residuals when  $u_t$  with 0 time delay is added to the model of  $y_t$  instead.

Asymptotically, the model of  $y_t$  should converge to the process model with residuals equal to  $e_{y,t}$  even if  $u_t$  is available when estimating the model as a predictor variable. If the model converges to the true process model, availability of the current value of  $u_t$  will not change the estimated model and therefore, the modeling residual will remain equal to  $e_{y,t}$  while  $\hat{e}_{u,t} = k_p e_{y,t} + e_{u,t}$ . So, the correlation between modeling residuals will not be removed in this case.

It is also possible that the identification algorithm based on small sample sizes converges to the inverse of the controller model when estimating the model of  $y_t$ .

Considering Equation 5.26,  $y_t$  can be written as:

$$y_t = \frac{1}{k_p + G_c(z^{-1})} u_t - \frac{1 + G_{eu}(z^{-1})}{k_p + G_c(z^{-1})} e_{u,t} \quad (5.31)$$

$$y_t = \frac{1}{k_p + G_c(z^{-1})} u_t + G_e(z^{-1}) e_{u,t} - \frac{1}{k_p} e_{u,t} \quad (5.32)$$

where  $G_e(z^{-1}) = -\frac{1+G_{eu}(z^{-1})}{k_p+G_c(z^{-1})} + \frac{1}{k_p}$  which has one sample time delay.

One step ahead prediction error in this case can be obtained as:

$$\begin{aligned} e(t) &= y_t - \hat{y}_t = \frac{1}{k_p + G_c(z^{-1})} u_t + G_e(z^{-1}) e_{u,t} - \\ &\quad \frac{1}{k_p} e_{u,t} - L_1(z^{-1}) y_t - K u_t - L_2(z^{-1}) u_t \end{aligned} \quad (5.33)$$

$L_1(z^{-1})$  and  $L_2(z^{-1})$  are assumed to have at least one sample time delay and the part of  $L_2(z^{-1})$  with 0 time delay is separated out as  $K$ .

Substituting  $e_{u,t} = \frac{u_t - k_p y_t - G_c(z^{-1}) y_t}{1 + G_{eu}(z^{-1})}$  results in:

$$\begin{aligned} \epsilon(t) &= \frac{1}{k_p + G_c(z^{-1})} u_t - K u_t - L_2(z^{-1}) u_t + \\ &\quad G_e(z^{-1}) \frac{u_t - k_p y_t - G_c(z^{-1}) y_t}{1 + G_{eu}(z^{-1})} - \frac{1}{k_p} e_{u,t} - L_1(z^{-1}) y_t \\ &= \left[ \frac{1}{k_p + G_c(z^{-1})} - K - L_2(z^{-1}) + \frac{G_e(z^{-1})}{1 + G_{eu}(z^{-1})} \right] u_t - \\ &\quad \left[ \frac{G_e(z^{-1})(k_p + G_c(z^{-1}))}{1 + G_{eu}(z^{-1})} + L_1(z^{-1}) \right] y_t - \frac{1}{k_p} e_{u,t} \end{aligned} \quad (5.34)$$

Since all the transfer functions have at least one sample time delay, the minimum error variance equals to  $\frac{1}{k_p^2} \text{var}(e_{u,t})$  (if  $K$  converges to  $\frac{1}{k_p}$ ). If  $K$  does not converge to  $\frac{1}{k_p}$ , the residual will be a combination of  $e_{u,t}$  and  $e_{y,t}$ . Therefore, even if the identification algorithm converges to the inverse of the controller model instead of the process model (with the availability of  $u_t$  as predictor) the modeling residual will still be correlated with the modeling residual of  $u_t$  ( $k_p^2 \text{var}(e_{y,t}) + \text{var}(e_{u,t})$ ).

This fact can be utilized to distinguish between the process output and the controller output from data. Based on the discussion, it is possible to learn from data if there is instantaneous feedback between two variables and which of the two is the instantaneous function of the other one. Following is the summary of the procedure for detection of instantaneous feedback:

- Build a VAR model including the historical data with minimum one sample time delay.

- Check for non-zero correlation between modeling residuals of individual variables (which can happen if the variables are in a loop)
- Add one of the variables in the loop ( $z_t$ ) as a regressor with 0 time delay to the model of the second variable ( $w_t$ ). Estimate the model of  $w_t$  again.
- Check the correlation between the residuals of the new model of  $w_t$  and the residuals of the model of  $z_t$ .
- If the correlation is 0, then  $w_t$  is the controller output and  $z_t$  is the process output. Otherwise,  $w_t$  with 0 time delay should be added to the model of  $z_t$  to remove the correlation.

It should be mentioned that if the instantaneous correlation is due to a third variable, the correlation between residuals will not be removed in any of the models described above. The appendix presents the method to check the correlation between residuals of a multivariate regression model estimated based on Bayesian method. The following section explains the methodology to estimate the parameters of a SVAR model under Bayesian framework.

## 5.5 Bayesian estimation of SVAR models

There are two main advantages in applying Bayesian framework for model estimation. The first advantage is the possibility of incorporating prior information from various sources in model estimation to ensure the estimated model approximates the true underlying structure of the process. It is well known that several models can be estimated based on the data with similar prediction performance. A proper prior design, which will be discussed in Section 5.6, can better ensure convergence to a physically interpretable model among many models which minimize the modeling residuals.

The other advantage of applying Bayesian framework is that it can deal with the problem of high dimensionality and overfitting in a natural way and again through a proper design of the prior distribution. The main difficulty in application of VAR models is the large number of parameters to be estimated from data even for a medium size VAR model. For example, if there are 10 response variables included in the model with the maximum time lag equal to 15 samples, there will be 150 parameters to estimate from data. Considering that the amount of quality data is usually limited, traditional OLS could result in unreliable estimation of model parameters.

It is proved in literature that Bayesian VARs can deal with the problem of over-parametrization in the presence of a large number of potential predictor variables with collinearity [88]. This can be achieved through appropriate prior design to shrink the parameters toward zero which is called Bayesian shrinkage in literature. Banbura et al [89] showed that applying Bayesian shrinkage to large models is sufficient to prevent over-parameterization if the tightness of the prior distribution is increased when increasing the number of variables. [88] has compared the prediction performance of several models estimated by conventional methods appropriate for large models (many collinear predictors) and has concluded that the Bayesian shrinkage method outperforms other methods. Giacomini and White [90] have shown that when the predictor variables are strongly correlated, Bayesian shrinkage tends to keep those predictors that explain the most variations in response variables.

Considering the fact that the variables in an engineering process are very likely to have strong collinearity, a framework which can deal with the problem of collinearity between variables as well as presence of a large number of variables is valuable. Section 5.6 discusses the appropriate method for design of the prior distribution in order to avoid over-parameterized models and model formulation under Bayesian framework is discussed in the following.

As discussed in the previous section, SVAR models with the form of Equation 5.23 are used for model estimation. The matrix form of the model including all the observations can be written as in Equation 5.35.

$$YA_0 - XA_+ = \varepsilon \quad (5.35)$$

where  $Y$  is  $N \times m$  matrix of observations of  $m$  response variables for sample size  $N$  and  $X$  is  $N \times k$  matrix of regressors.  $A_+$  is  $k \times m$  and  $A_0$  is  $m \times m$ . Each column of  $A_0$  corresponds to one equation.  $\varepsilon$  is a  $N \times m$  matrix of residuals with rows distributed as  $N(0, \Sigma_\varepsilon)$  where  $\Sigma_\varepsilon$  is diagonal. Matrix  $X$  includes the lagged response variables  $y$  as well as a column of 1 to account for the mean value. Exogenous variables can also be considered in  $X$  matrix. Rows of  $X$  are arranged as  $x_t = [y_{t-1}, \dots, y_{t-p}, 1]$  where  $y_t = [y_{1,t}, \dots, y_{m,t}]$ .

### 5.5.1 Estimation of SVAR model parameters in fully recursive form

If it is possible to write the SVAR model equations in a fully recursive form, where each dependent variable can only be a simultaneous function of the previous dependent variables,  $A_0$  will be triangular and  $\Sigma_\varepsilon$  diagonal. The fully recursive form is as

following which includes the form of  $A_0$  matrix explained above as a special case.

$$\begin{aligned} y_1 &= X\beta_1 + \varepsilon_1 \\ y_2 &= y_1\gamma_{21} + X\beta_2 + \varepsilon_2 \\ &\vdots \\ y_m &= y_1\gamma_{m1} + y_2\gamma_{m2} + \dots + y_{m-1}\gamma_{m,m-1} + X\beta_m + \varepsilon_m \end{aligned} \quad (5.36)$$

where  $y_\alpha$  is  $n \times 1$  vector of observations,  $X$  is  $n \times k$  matrix of rank  $k$  of observations,  $\beta_\alpha$  is a  $k \times 1$  vector and  $\varepsilon_\alpha$  is  $n \times 1$  vector of disturbances while  $\gamma_{\alpha,l}$  are scalars. The elements of  $\varepsilon_\alpha$ ,  $\alpha = 1, \dots, m$  are normally distributed with zero mean and covariance matrix  $E[\varepsilon\varepsilon'] = D(\sigma_\alpha^2) \otimes I_n$  where  $D(\sigma_\alpha^2)$  is a diagonal matrix with  $\sigma_1^2, \dots, \sigma_m^2$  on main diagonal. Therefore, the disturbance terms could have different variances.

Since the equations are independent, the likelihood function can be written as:

$$l(\delta, \sigma | Y, Y_0) \propto \prod_{\alpha=1}^m \frac{1}{\sigma_\alpha^n} \exp\left[-\frac{1}{2\sigma_\alpha^n} (y_\alpha - Z_\alpha \delta_\alpha)' (y_\alpha - Z_\alpha \delta_\alpha)\right] \quad (5.37)$$

where  $Y_0$  denotes given initial conditions,  $\sigma' = [\sigma_1, \dots, \sigma_m]$ ,  $Z_\alpha = [y_1, \dots, y_{\alpha-1}; X]$ ,  $\delta'_\alpha = [\gamma'_\alpha; \beta'_\alpha]$  with  $\gamma'_\alpha = [\gamma_{\alpha 1}, \dots, \gamma_{\alpha, \alpha-1}]$  and  $\delta' = [\delta'_1, \dots, \delta'_m]$ .

The prior information is assumed to be non-informative as:

$$\begin{aligned} p(\delta) &\propto \text{constant} \\ p(\sigma) &\propto \prod_{\alpha=1}^m \frac{1}{\sigma_\alpha} \end{aligned} \quad (5.38)$$

The joint posterior of pdf for  $\delta, \sigma$  can be obtained as:

$$p(\sigma, \delta | Y, Y_0) = \prod_{\alpha=1}^m \frac{1}{\sigma_\alpha^{n+1}} \exp\left[-\frac{1}{2\sigma_\alpha^2} (y_\alpha - Z_\alpha \delta_\alpha)' (y_\alpha - Z_\alpha \delta_\alpha)\right] \quad (5.39)$$

Since the equations are independent from each other, the posterior for each equation can be written as

$$\begin{aligned} p(\sigma_\alpha, \delta_\alpha | Y, Y_0) &\propto \frac{1}{\sigma_\alpha^{n+1}} \exp\left[-\frac{1}{2\sigma_\alpha^2} (y_\alpha - Z_\alpha \delta_\alpha)' (y_\alpha - Z_\alpha \delta_\alpha)\right] \\ &\propto \frac{1}{\sigma_\alpha^{n+1}} \exp\left[-\frac{1}{2\sigma_\alpha^2} (\hat{\varepsilon}'_\alpha \hat{\varepsilon}_\alpha + (\delta_\alpha - \hat{\delta}_\alpha)' Z'_\alpha Z_\alpha (\delta_\alpha - \hat{\delta}_\alpha))\right] \end{aligned} \quad (5.40)$$

where  $\hat{\varepsilon}_\alpha = y_\alpha - Z_\alpha \hat{\delta}_\alpha$  and  $\hat{\delta}_\alpha = (Z'_\alpha Z_\alpha)^{-1} Z'_\alpha y_\alpha$ .

Conditional posterior pdf for  $\delta_\alpha$  given  $\sigma_\alpha$  is the multivariate normal form with the mean  $\widehat{\delta}_\alpha$ . The marginal pdf of  $\sigma_\alpha$  is

$$p(\sigma_\alpha|Y, Y_0) \propto \frac{1}{\sigma_\alpha^{n-q_\alpha+1}} \exp\left(-\frac{\widehat{\varepsilon}'_\alpha \widehat{\varepsilon}_\alpha}{2\sigma_\alpha^2}\right) \quad (5.41)$$

where  $q_\alpha$  is the number of elements in  $\delta_\alpha$  which is in the inverted gamma form.

The marginal posterior pdf for  $\delta_\alpha$  can be obtained as:

$$p(\delta_\alpha|Y, Y_0) \propto [\widehat{\varepsilon}'_\alpha \widehat{\varepsilon}_\alpha + (\delta_\alpha - \widehat{\delta}_\alpha)' Z'_\alpha Z_\alpha (\delta_\alpha - \widehat{\delta}_\alpha)]^{-\frac{n}{2}} \quad (5.42)$$

which is the multivariate student t form.

If the prior distribution for  $\delta_\alpha$  is considered as multivariate normal such as

$$p(\delta_\alpha) \propto \exp\left[-\frac{1}{2}(\delta_\alpha - \bar{\delta}_\alpha)' R_\alpha^{-1} (\delta_\alpha - \bar{\delta}_\alpha)\right] \quad (5.43)$$

while the prior for  $\sigma_\alpha$  is  $p(\sigma_\alpha) \propto 1/\sigma_\alpha$ , the conditional posterior of  $\delta_\alpha$  given  $\sigma_\alpha$  can be obtained as

$$\begin{aligned} p(\delta_\alpha|\sigma_\alpha, Y, Y_0) &\propto \exp\left(-\frac{1}{2\sigma_\alpha^2}(\delta_\alpha - \widehat{\delta}_\alpha)' Z'_\alpha Z_\alpha (\delta_\alpha - \widehat{\delta}_\alpha)\right) \times \\ &\exp\left[-\frac{1}{2}(\delta_\alpha - \bar{\delta}_\alpha)' R_\alpha^{-1} (\delta_\alpha - \bar{\delta}_\alpha)\right] \\ &\propto \exp\left[-\frac{1}{2}(\delta_\alpha - \mu_\alpha)' \omega_\alpha^{-1} (\delta_\alpha - \mu_\alpha)\right] \end{aligned} \quad (5.44)$$

where  $\omega_\alpha = (R_\alpha^{-1} + \sigma_\alpha^{-2}(Z'_\alpha Z_\alpha))^{-1}$  and  $\mu_\alpha = \omega_\alpha(\bar{\delta}_\alpha R_\alpha^{-1} + \widehat{\delta}_\alpha \sigma_\alpha^{-2}(Z'_\alpha Z_\alpha))$ .

Marginal distribution of  $\delta_\alpha$  is also obtained as:

$$\begin{aligned} p(\delta_\alpha|Y, Y_0, PI) &\propto [\widehat{\varepsilon}'_\alpha \widehat{\varepsilon}_\alpha + (\delta_\alpha - \widehat{\delta}_\alpha)' Z'_\alpha Z_\alpha (\delta_\alpha - \widehat{\delta}_\alpha)]^{-n/2} \times \\ &\exp\left[-\frac{1}{2}(\delta_\alpha - \bar{\delta}_\alpha)' R_\alpha^{-1} (\delta_\alpha - \bar{\delta}_\alpha)\right] \end{aligned} \quad (5.45)$$

which is normal-t distribution.

Generating random samples from this posterior distribution is not easy since it does not have a standard form. In order to make the distribution to have a standard form, the t distribution can be approximated by normal distribution as proposed by [91]. By expanding the term with the form of t distribution and keeping the first

normal term of approximation, the posterior probability becomes:

$$\begin{aligned}
p(\delta_\alpha | Y) \propto & \exp\left[-\frac{1}{2}(\delta_\alpha - \widehat{\delta}_\alpha)' \overline{\widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha}^{-1} \otimes Z'_\alpha Z_\alpha (\delta_\alpha - \widehat{\delta}_\alpha)\right] \times \\
& \exp\left[-\frac{1}{2}(\delta_\alpha - \overline{\delta}_\alpha) R^{-1} (\delta_\alpha - \overline{\delta}_\alpha)\right] \propto \\
& \exp\left[-\frac{1}{2}(\delta_\alpha - b) F(\delta_\alpha - b)\right]
\end{aligned} \tag{5.46}$$

where  $\overline{\widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha} = \frac{1}{n} \widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha$ ,

$$b = (R^{-1} + \overline{\widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha}^{-1} \otimes Z'_\alpha Z_\alpha)^{-1} (R^{-1} \overline{\delta}_\alpha + \overline{\widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha}^{-1} \otimes Z'_\alpha Z_\alpha \widehat{\delta}_\alpha) \tag{5.47}$$

and

$$F = R^{-1} + \overline{\widehat{\varepsilon}_\alpha \widehat{\varepsilon}_\alpha}^{-1} \otimes Z'_\alpha Z_\alpha \tag{5.48}$$

### 5.5.2 Solving the equations simultaneously

The reduced form of the model is obtained by right-multiplying the equation with  $A_0^{-1}$  as:

$$Y = XB + E \tag{5.49}$$

where covariance of  $E \Sigma_E = A_0'^{-1} \Sigma_\varepsilon A_0^{-1}$  and  $B = A_+ A_0^{-1}$ .  $\Sigma_E$  is not diagonal in general since its elements are linear combinations of structural shocks  $\varepsilon$ .

The likelihood of observations matrix  $Y$  in Equation 5.49 can be written as in the following equation:

$$\begin{aligned}
p(Y|X, B, \Sigma_E) \propto & \\
|\Sigma_E|^{-T} \exp\left[-\frac{1}{2} \text{tr}((Y - XB)' \Sigma_E^{-1} (Y - XB))\right]
\end{aligned} \tag{5.50}$$

Considering least square estimate  $\widehat{B} = (X'X)^{-1} X^T Y$  and  $S = (Y - X\widehat{B})' (Y - X\widehat{B})$ , the likelihood takes the form:

$$\begin{aligned}
p(Y|X, B, \Sigma_E) \propto & |\Sigma_E|^{-(T-k)} \exp\left[-\frac{1}{2} \text{tr}(S \Sigma_E^{-1})\right] \\
|\Sigma_E|^{-k} \exp\left[-\frac{1}{2} \text{tr}((B - \widehat{B})' X' \Sigma_E^{-1} X (B - \widehat{B}))\right]
\end{aligned} \tag{5.51}$$

Likelihood function is quadratic with respect to  $B$  conditioned on a fixed  $\Sigma_E$ . Marginalized posterior of  $\Sigma_E$  has the form of inverted Wishart distribution under uniform prior distribution.

If the model in Equation 5.35 is exactly identified which means there are exactly  $\frac{m(m-1)}{2}$  restrictions on  $A_0$  matrix, there is a unique map from  $\Sigma_E$  to the elements

of  $A_0$  and  $\Sigma_\varepsilon$ . It is commonly assumed that  $A_0$  has a triangular form and  $A_0^{-1}\Sigma_\varepsilon^{1/2}$  can be obtained by Choleski factorization of  $\Sigma_E$ . Therefore, in order to sample from the posterior distribution of  $A_0$ , it is sufficient to map the samples from posterior distribution of  $\Sigma_E$  through a Choleski factorization. The posterior distribution of  $\Sigma_E$  and  $B$  will remain in the same distribution form as the likelihood if a uniform prior is chosen for  $B$  and  $\Sigma_E$ .

When the number of restrictions on  $A_0$  matrix is more than  $\frac{m(m-1)}{2}$ ,  $\Sigma_E$  will be restricted. Mapping the samples from the unrestricted distribution of  $\Sigma_E$  to  $A_0$  is not accurate anymore [92]. Thus, the posterior should be written as a function of  $A_0$  itself. Another advantage of writing the likelihood in terms of  $A_0$  matrix is the possibility of designing prior distribution for its elements.

Following the work of [92], by vectorizing the matrices in Equation 5.51, the likelihood can be written in a multivariate form instead of matrix variate which makes it easier to work with. Defining  $Z = [Y, -X]$  and  $A = [A_0, A_+]'$ , the model in Equation 5.35 can be written as  $ZA = \varepsilon$ .

The likelihood of the data can be written as in Equation 5.53.

$$L(Y|A) \propto |A_0|^T \exp\left[-\frac{1}{2} \text{trace}(ZA)'(ZA)\right] \quad (5.52)$$

The vectorized form of  $A$  is denoted by  $a$  which is obtained by stacking the columns of the matrix  $A$  on top of one another.  $a_0$  and  $a_+$  are also obtained by vectorizing  $A_0$  and  $A_+$  matrices.

$$L(Y|A) \propto |A_0|^T \exp\left[-\frac{1}{2} a'(I \otimes Z'Z)a\right] \quad (5.53)$$

The prior PDF on the parameters of  $a_0$  and  $a_+$  vectors is defined as following.

$$\pi(a) = \pi(a_0)\mathcal{N}(a_+ - \mu, H) \quad (5.54)$$

$\pi(a_0)$  is the prior on the  $a_0$  vector and the prior on the  $a_+$  is a normal distribution with mean  $\mu$  and covariance  $H$ .

The posterior density of model parameters can be obtained by multiplying the likelihood with the prior density as in Equation 5.55.

$$\begin{aligned} p(a) &\propto \pi(a_0)|A_0|^T|H|^{-\frac{1}{2}} \times \\ &\exp\left[-\frac{1}{2}(a_0'(I \otimes Y'Y)a_0 - 2a_+'(I \otimes X'Y)a_0\right. \\ &\quad \left.+ a_+'(I \otimes X'X)a_+ + (a_+ - \mu')H^{-1}(a_+ - \mu)]\right] \end{aligned} \quad (5.55)$$

The posterior distribution of  $a_+$  given a fixed  $a_0$  is Gaussian as in Equation 5.56.

$$p(a_+|a_0) = \mathcal{N}[a_0^*; (I \otimes X'X + H^{-1})^{-1}] \quad (5.56)$$

where  $a_0^* = (I \otimes X'X) + H^{-1})^{-1}((I \otimes X'Y)a_0 + H^{-1}\mu)$ .

Marginal distribution of  $a_0$  can be obtained as:

$$\begin{aligned} p(a_0) &\propto \pi(a_0)|A_0|^T|(I \otimes X'X)H + I|^{-1/2} \\ &\exp[-0.5(a_0'(I \otimes Y'Y)a_0 + \mu'H^{-1}\mu \\ &- a_0^{*'}(I \otimes X'X) + H^{-1})a_0^*] \end{aligned} \quad (5.57)$$

In order to sample from the posterior distribution of model parameters, we need to first sample from the marginal posterior of  $a_0$  followed by sampling from the conditional posterior of  $a_+$ . Sampling from the posterior of  $a_0$  requires application of special methods since the probability density does not have a standard form. [93] has proposed a Gibbs sampler in order to sample from the posterior of over identified SVAR model parameters with exact restrictions on elements of  $a_0$ . The proposed method is the most reliable for over identified models where there are more than  $\frac{m(m-1)}{2}$  exact restrictions on  $A_0$  matrix. However, here it is preferred not to impose exact restrictions on the lower triangular elements of  $A_0$  matrix but to use prior distribution to shape the posterior distribution. There are methods such as importance sampling and Metropolis-Hastings (MH) algorithms which can be used to sample from the distribution in Equation 5.57 [94]. The Metropolis algorithm is used here as explained in following.

Application of Metropolis algorithm requires a proposal distribution ( $q(a_0)$ ) proportional to the true posterior distribution to sample from. Waggoner and Zha [95] considered a multivariate t distribution with zero mean value and a scale equal to the inverse of the Hessian of the log likelihood as  $q(a_0)$ . The new sample ( $a^n$ ) is considered as the summation of the previous sample with the random sample taken from  $q(a_0)$  ( $z$ ) as  $a^n = a^{(n-1)} + z$ .

A problem with using t distribution to approximate the likelihood is that the posterior distribution will have a non-standard form if the prior distribution is not uniform. Since a normal distribution is considered as the prior distribution in this study, it is difficult to sample from the approximate posterior distribution which is the multiplication of the approximate likelihood to a multivariate normal distribution. Therefore, here the likelihood is approximated by a normal distribution to achieve a posterior with a standard form.

Another problem is that since  $q(a_0)$  is considered as the multiplication of the prior to approximate likelihood, the posterior of some of the elements of  $a_0$  will have an exact value of 1 or 0 due to the strict design of the prior distribution as will be described later. Therefore, the summation of the random sample taken from  $q(a_0)$  with the previous sample makes the samples to approach infinity. To resolve this

issue, the new sample equals to the random sample directly taken from  $q(a_0)$  with the difference that the mean of the distribution approximating the likelihood equals to the previous sample instead of 0. This adjustment in addition to the fact that the new sample is accepted if it is more probable compared to the previous sample, makes it more likely to search the whole space covered by the likelihood and converging to the most likely point while the samples do not approach infinity. The covariance matrix of the distribution approximating the likelihood is equal to the inverse of the Hessian of the log likelihood at its mode (Hessian is denoted by  $\mathcal{H}$  and equals to  $I \otimes YY' + I \otimes Y'Y - 2(I \otimes X'Y)'(I \otimes X'X + H^{-1})^{-1}(I \otimes X'Y)$ ). Thus,  $q(a_0)$  is taken as  $\pi(a_0) * \mathcal{N}(a^{(n-1)}, \text{inv}(\mathcal{H}))$  and the following steps should be taken to sample from the posterior of  $a_0$ .

- Initialize arbitrary value  $a^0$ . For  $n = 1, \dots, N$  do the following:
- generate  $z$  from  $q(a_0)$  and  $u$  from uniform distribution  $U(0, 1)$
- compute

$$J(a^{(n-1)}, z) = \min\left\{\frac{p(z)}{p(a^{(n-1)})}, 1\right\} \quad (5.58)$$

- if  $u \leq J(a^{(n-1)}, z)$ , set  $a^n = z$ ; else, set  $a^n = a^{(n-1)}$

### 5.5.3 Imposing exact restrictions on the posterior distribution

Some of the coefficients in each equation might be known that are exactly equal to zero a priori. For example, we may know that the process time delay is surely more than 5 samples and therefore in the equation modeling the process output, the coefficients of the lags within 5 samples should be exactly 0. Estimated values of these coefficients will converge to 0 in large sample since the estimation methods are asymptotically unbiased and consistent. However, estimated values based on small sample sizes could be non-zero. Even though these coefficients are likely to be close to zero even for small sample sizes, the correlation of the corresponding predictor variables with other predictor variables could make the coefficients to have a non-zero value. It is observed that even small valued non-zero coefficients cause a bias in estimated impulse response. Restricting the coefficients which are surely expected to be zero by prior knowledge, can help to achieve a better estimation of non-zero coefficients and therefore a more reliable quantification of interactions.

Restricting the values of specific coefficients in each equation is possible by conditioning the posterior distribution of equations. It should also be mentioned that

since different coefficients are restricted in each equation, it is only possible to impose restrictions individually for each equation which is again possible if the equations are independent.

$\delta_\alpha$  is partitioned as  $[\delta_a \delta_b]'$  where  $\delta_a$  represents non-zero coefficients and  $\delta_b$  includes the coefficients which should be equal to 0.  $Z_\alpha$  is also partitioned accordingly into  $[Z_a Z_b]'$ . The posterior covariance matrix  $\omega_\alpha$  will be partitioned as

$$\begin{bmatrix} A & C \\ C^T & B \end{bmatrix}$$

where  $A$  is the covariance matrix of the predictor variables corresponding to non-zero coefficients,  $B$  is the covariance matrix of zero coefficients and  $C$  is the cross covariance matrix. Inverse of  $\omega_\alpha$  matrix can be written as

$$\begin{aligned} \omega_\alpha^{-1} = & \begin{bmatrix} I & 0 \\ -B^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - C^T B^{-1} C)^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \times \\ & \begin{bmatrix} I & -C^T B^{-1} \\ 0 & I \end{bmatrix} \end{aligned} \quad (5.59)$$

which results in

$$\begin{aligned} p(\delta_a, \delta_b) \propto & \exp\left[-\frac{1}{2} \begin{bmatrix} (\delta_a - \mu_a) - C^T B^{-1}(\delta_b - \mu_b) \\ (\delta_b - \mu_b) \end{bmatrix} \times \right. \\ & \begin{bmatrix} (A - C^T B^{-1} C)^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \times \\ & \left. \begin{bmatrix} (\delta_a - \mu_a) - C^T B^{-1}(\delta_b - \mu_b) \\ (\delta_b - \mu_b) \end{bmatrix} \right] \end{aligned}$$

which can be simplified as:

$$\begin{aligned} p(\delta_a, \delta_b) \propto & \exp\left[-\frac{1}{2}((\delta_a - \mu_a) - C^T B^{-1}(\delta_b - \mu_b))^T \times \right. \\ & (A - C^T B^{-1} C)^{-1}((\delta_a - \mu_a) - C^T B^{-1}(\delta_b - \mu_b))] \times \\ & \left. \exp\left[-\frac{1}{2}(\delta_b - \mu_b)^T B^{-1}(\delta_b - \mu_b)\right]\right] \end{aligned} \quad (5.60)$$

(5.61)

$\delta_b$  equals to zero and  $\mu_b$  is the estimated value of  $\delta_b$  which is a weighted summation of the MLE estimation and the prior  $\bar{\delta}_b$  (which is 0). It is observed that the mean of  $\delta_a$  is modified with the addition of  $C^T B^{-1} \mu_b$  and its covariance is also modified due to non-zero  $C$  matrix.

## 5.6 Design of the prior distribution

Capability of incorporating the prior distribution in estimation of model parameters is an inherent advantage of Bayesian analysis. It is mentioned in the literature that in large systems with many variables and large time lags, the prior distribution on the model parameters help the model identification. However, the prior probability should be based on sound information to ensure convergence of the model parameters to the true values. In modeling engineering process variables, knowledge of the probable time delay between the variables and probable time constant of the system is helpful in designing the prior if available. It is note worthy that if the prior knowledge used in the model estimation is not accurate, it can lead to in-accurate estimation of model parameters.

### 5.6.1 Prior on elements of $A_0$

The  $A_0$  matrix introduced in Section 5.6 makes the system identifiable since there are sufficient restrictions to uniquely estimate the model parameters. Imposing the  $A_0$  matrix or equivalently  $a_0$  vector to the model is possible through the prior distribution. The elements of each column of  $A_0$  matrix which correspond to one equation are considered to be independent of the elements in other columns. Therefore, the prior distribution on the elements of  $A_0$  has a multivariate normal form as:

$$p(A_0) = \prod_1^m \mathcal{N}(\Gamma_m, F_m) \quad (5.62)$$

where  $\Gamma_m$  and  $F_m$  are the prior mean and covariance matrix for the elements of each column. The  $m_{th}$  element of  $\Gamma_m$  equals to 1 and it may have one more non-zero element if the equation corresponds to the controller output of the  $m_{th}$  loop. The covariance matrix  $F_m$  should be designed in such a way that ensures the posterior distribution for  $\Gamma_m$  will have the  $m_{th}$  element equal to 1 and other elements equal to 0 except for  $(m - 1)_{th}$  element in equations corresponding to controller outputs. Choosing a variance very close to 0 for all the elements except for the  $(m - 1)_{th}$  element in equations modeling the controller outputs guarantees that the posterior mean of the elements of  $A_0$  matrix will also be close to the desired values.

The posterior mean of the  $(m - 1)_{th}$  element in the equations modeling the controller outputs should be equal to the proportional gain of the corresponding controller. Therefore, the mean of this element can be chosen based on the priory knowledge on the likely value of controller proportional gain with a large variance to cover the range of uncertain values if the prior is uncertain.

Prior on  $A_0$  matrix can be easily transformed into a prior on  $a_0$  vector by stacking vectors of  $\Gamma_m$  on top of each other to form the vector of mean values. The covariance matrix can also be constructed as  $\text{diag}(F_1, \dots, F_m)$ .

### 5.6.2 Prior on elements of $a_+$

By examining the posterior distribution of  $a_+$ , it is observed that a block diagonal design of the covariance matrix of prior distribution, can make it possible to transform the joint normal distribution to multiplication of  $m$  independent normal distribution, each corresponding to the coefficients of one equation. The term  $(I \otimes X'X) + H^{-1}$  can be written as a block diagonal matrix if  $H$  is block diagonal as is  $I \otimes X'X$ . Block diagonal design of  $H$  matrix means prior independence of the equations. Different covariance matrices can be designed for different equations in the model denoted as  $R$ , and the  $H$  matrix can be constructed as  $\text{diag}(R_1, \dots, R_m)$ . By this design the term  $(I \otimes X'X) + H^{-1}$  can be written as  $\text{diag}(X'X + R_1^{-1}, \dots, X'X + R_m^{-1})$ . Therefore, sampling from the posterior distribution of  $a_+$  can be transformed into sampling from  $m$  normal distributions, each belonging to an equation in the system.

The form of prior distribution for each equation's coefficients is considered as  $\mathcal{N}(\mu_m, R_m)$ . Regardless of the coefficients for which there is a prior knowledge of their values, the common practice in econometric literature to design prior distribution for Bayesian VARs is to follow Litterman's prior. Litterman's method is based on the assumption that the variables are usually in the form of a random walk. Litterman assigns a high prior probability on the lags of the response variable itself and a smaller probability for the lags of the predictor variables in each equation. In order to do that, the elements of  $\mu_m$  are set equal to 0 except for the element corresponding to one sample lagged of the response variable itself in each equation.

The covariance matrix  $R_m$  controls the tightness of the belief in the prior mean.  $R_m$  is usually chosen to be diagonal (assuming independence between the coefficients) with non-zero elements designed as:

$$\frac{\lambda_0 \lambda_1}{\sigma_j l^{\lambda_3}} \quad (5.63)$$

where  $l$  is the time lag of the respective regressor.  $\lambda_0$ ,  $\lambda_1$  and  $\lambda_3$  are the hyper parameters.  $\lambda_0$  controls the general tightness of beliefs on  $\mu_m$ ,  $\lambda_1$  controls the tightness of belief around the random walk assumption and  $\lambda_3$  controls the shrinking rate of the variance.  $\sigma_j$  is the scale factor accounting for the different measurement scales for the variables and different residual variances.  $\sigma_j$  can be selected based on the knowledge of the measurement units of the variables or a preliminary analysis on the

data.

$\lambda_1$  can be chosen equal to 1 for the history of the response variable itself and less than 1 for the predictor variables in each equation. This reduces the probability of the coefficients corresponding to the predictor variables to be different from zero which is desirable based on the random walk assumption.  $\lambda_3$  is also between 0 and 1 and its role is to push the coefficients at larger lags to 0. It is desirable to avoid over-fitting problem and having unnecessary non-zero coefficients. The role of  $\lambda_3$  is similar to the forgetting factor where it is desired to give less weight to the historical data based on their time lag. A value equal to 0.95 is chosen in this work for  $\lambda_3$ .

Since variables in an engineering process do not necessarily follow a random walk behaviour, the prior distribution is defined differently. In each equation, some of the coefficients might be known that should be equal to 0 based on the nature of the response variable or the likely time delays. The elements in  $\mu_m$  corresponding to these coefficients can be chosen as 0 with a variance close to 0 to make the posterior mean of the coefficients approach 0 as well.

Hyper parameter  $\lambda_1$  which gives more weight to the history of the response variable itself rather than other regressors has a different role here since process variables do not necessarily have a random walk behavior. By selecting  $\lambda_1 = 1$  for all the elements in equations corresponding to process outputs, equal importance is given to all the regressors that are desirable.

In equations modeling the controller outputs,  $\lambda_1$  is selected very close to 0 for the history of the response variable itself and also the history of other variables except for the corresponding process outputs. The reason is that the true model for each controller output is only a function of the corresponding process variable. By manipulating  $\lambda_1$ , it is possible to achieve an estimation of the true model parameters.

The role of  $\lambda_3$  is kept the same since it is desirable to assign less importance to regressors with larger time lags. However, it should be noted that if there is a time delay ( $D$ ) from one regressor to the corresponding response variable,  $l - D + 1$  should be used instead of  $l$  in Equation 5.63 so that the variance has its highest value at the time lag corresponding to the time delay.

## 5.7 Impulse response estimation

A stationary stable SVAR( $p$ ) process as defined in Equation 5.23 has a corresponding VMA representation according to Wold's theorem as follows.

$$y_t = \Theta_0 \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2} + \dots \quad (5.64)$$

where  $\Theta_j = \phi_j A_0^{-1}$  and  $\phi_j$ s are  $m \times m$  dimensional matrices recursively obtained by the following equation.

$$\phi_s = \sum_{j=1}^s \phi_{s-j} A_j \quad (5.65)$$

where  $\phi_0 = A_0$ .

Proper design of the  $A_0$  matrix as described in Section 5.3 guarantees that  $\varepsilon_t$  has a diagonal covariance matrix. Therefore, VMA representation of the SVAR model can be uniquely identified.

In order to determine a confidence interval for the estimated impulse responses, the posterior distribution of impulse responses coefficients are needed. The relation between  $\Theta_j$  and the estimated coefficients of SVAR is nonlinear and the analytical relation between their posterior distributions cannot be derived for  $j > p + 1$  where  $p$  is the order of the AR part of the model. [96] presented the analytical form of the impulse response distribution based on the posterior of Bayesian SVAR model up to time horizon  $p + 1$ . Since there is one to one relation between the first  $p + 1$  impulse response coefficients and the  $p + 1$  coefficient matrices in the VAR model, it is possible to analytically derive the probability distribution of the impulse response coefficients up to horizon  $p + 1$ . However, quantification of the energy transfer between the variables requires estimation of the impulse response coefficients till the values approach to 0 and not only the first  $p + 1$  coefficients. Therefore, we have to rely on drawing samples from  $p(\Theta_j | a_+, a_0)$  based on the Monte Carlo method.

In order to draw samples from the posterior distribution of the impulse responses, samples from the posterior distribution of the SVAR model parameters should be transformed into the impulse response form. The method to obtain a confidence interval for the estimated impulse responses is described in the following section.

### 5.7.1 Confidence interval for the impulse responses

Inference on the impulse response requires determining confidence intervals for the estimated impulse response coefficients. [97] has proposed to estimate the posterior probability density of the impulse responses at each time horizon individually and to obtain the standard deviation and skewness measure for each impulse response coefficient independently. The method that is used in this work is based on that of [98] where the dependency between the impulse response coefficients in the time horizon is also taken care of. Lutkepoh et al. [99] presents a comparison between different methods for estimation of confidence intervals for impulse response coefficients based on classical methods.

Since the impulse response is estimated based on the SVAR model, sampling of the impulse response distribution requires sampling from the distributions of the parameters of the SVAR model. For each draw from the posterior distribution of the SVAR model as in Section 5.5, the impulse response is estimated for a suitably large time horizon.

One way to derive the confidence interval for the impulse response coefficients is to find the confidence interval for each  $\Theta_{ij}(t)$  (impulse response from variable  $j$  to  $i$  after time  $t$ ) independently. However, as is discussed in [98], impulse responses  $\Theta_{ij}$ s are not independent of each other over the time horizon. Therefore, this dependence should be considered in determining the confidence intervals. In order to do that, [98] proposed to consider  $\Theta_{ij}(t)|_{t=0}^H$  as a  $H$  dimensional vector with covariance matrix  $\Omega$ .  $\Omega$  can be estimated from the Monte Carlo samples.

The eigenvector decomposition of the covariance matrix  $\Omega$  is utilized as follows.

$$W\Lambda W' = \Omega \quad (5.66)$$

where  $\Lambda$  is diagonal and  $W'W = I$ . Columns of  $W$  are the eigenvectors of  $\Omega$ . The vectors of  $\Theta_{ij}$  can be represented as:

$$\Theta_{ij} = \widehat{\Theta}_{ij} + \sum_{k=1}^H \gamma_k W_{.k} \quad (5.67)$$

where  $\widehat{\Theta}_{ij}$  is the estimated mean of  $\Theta_{ij}$  and  $W_{.k}$  is the  $k_{th}$  column of  $W$ .  $\gamma_k$  causes the uncertainty in the impulse responses and its variance equals to the  $k_{th}$  eigenvalue of  $\Omega$ .

A confidence interval can be estimated as  $\widehat{\Theta}_{ij} \pm \sqrt{\lambda_k} W_{.k}$  for the impulse responses. However, this confidence interval is symmetric while the posterior distribution of the impulse responses might be asymmetric. To take care of this asymmetry, a second pass through the simulated samples is made in order to find the interval for  $\gamma_k$ s.  $\gamma_k$  can be obtained as  $(\Theta_{ij} - \widehat{\Theta}_{ij})'W_{.k}$  for each draw of  $\Theta_{ij}$ . The two functions  $\widehat{\Theta}_{ij} + \gamma_{k,0.16}W_{.k}$  and  $\widehat{\Theta}_{ij} + \gamma_{k,0.84}W_{.k}$  provide the confidence interval for  $\Theta_{ij}$  along the direction of  $W_{.k}$  which could be asymmetric. It is possible to estimate the confidence interval considering all the eigenvectors as  $\widehat{\Theta}_{ij} + \sum_{k=1}^H \gamma_{k,0.16}W_{.k}$  and  $\widehat{\Theta}_{ij} + \sum_{k=1}^H \gamma_{k,0.84}W_{.k}$ . Availability of a confidence interval for the impulse response coefficients is important in order to find the time lag in which the impulse response converges to 0.

## 5.8 Impulse response decomposition

The impulse response obtained in the previous section represents the total transfer function between the variables in the sense that each  $\Theta_{ij}$  is a summation of the impulse

responses of all the different paths from which an impulse at  $\varepsilon_j$  can reach variable  $i$ . The 2-norm of the impulse response is the total energy or variation transferred to variable  $j$  in response to an impulse in variable  $i$ .

If the total impulse response from variable  $j$  to  $i$  is non-zero, there is either a direct or indirect causality from variable  $j$  to  $i$ . However, the total impulse response does not tell about the different paths from which variable  $j$  can reach variable  $i$ .

Gigi et al [87] proposed a way to decompose the total transfer function between variables to direct and indirect relations in the frequency domain. The direct transfer function is defined as the transfer function from innovations in the input directly transferring to the output without going through any loop or any intermediate variable. For example, Figure 5.1 depicts the variables belonging to two loops interacting with each other. The transfer functions representing each path are denoted by  $G_{\cdot,\cdot}$ ,

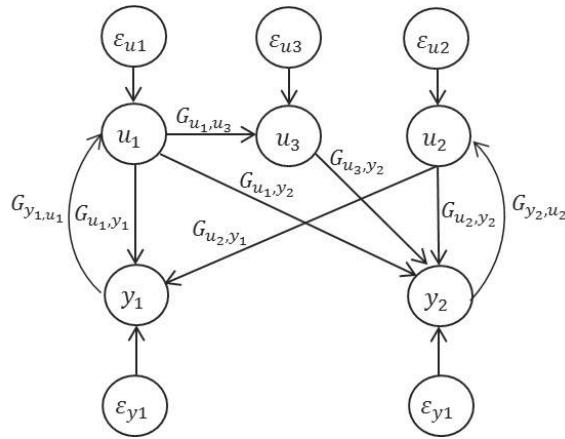


Figure 5.1: Simulated process

The direct transfer function from  $\varepsilon_{u1}$  to  $y_2$  is  $G_{u1,y2}$ .  $\varepsilon_{u1}$  can also reach  $y_2$  from two other paths: one is after going through  $y_1$  and returning to  $u_1$  (with a transfer function as  $G_{u1,y1} \times G_{y1,u1} \times G_{u1,y2}$ ), the second path is through  $u_3$  which again can be decomposed into 2 transfer functions based on inclusion of  $y_1$  ( $G_{u1,u3} \times G_{y2,u3}$  and  $G_{u1,y1} \times G_{y1,u1} \times G_{u1,u3} \times G_{y2,u3}$ ). These paths are indirect since  $\varepsilon_{u1}$  reaches  $y_2$  after passing through other variables. The total transfer function from  $\varepsilon_{u1}$  or  $u_1$  to  $y_2$  is a combination of all these transfer functions. The method proposed by [87] separates the direct transfer function ( $G_{u1,y2}$ ) from the total transfer function but does not separate the remainder of the transfer functions from each other. The method to be proposed below is capable of decomposing the total transfer function into all its individual components.

Decomposition of the total impulse responses can be done by manipulating the estimated SVAR model before transforming it to the VMA model. The VMA repre-

sentation of a properly manipulated SVAR model can represent any desired impulse response.

To determine the direct transfer function between two variables namely  $u_1$  and  $y_2$  in Figure 5.1, we need to only keep the path from which  $u_1$  reaches  $y_2$  without going through any other variable. Considering the coefficient matrix of an estimated SVAR model, we can remove all the indirect paths from which  $u_1$  can reach  $y_2$  by making the coefficients relating  $y_2$  to other variables except for the history of  $y_2$  and  $u_1$  to zero. Also, regarding the equation of  $u_1$  we need to make all the coefficients corresponding to other variables equal to 0 to ensure that  $u_1$  is not influenced by any variable in the system. For example, if variables  $[y_1(t), u_1(t), y_2(t), u_2(t), u_3(t)]'$  are only a function of  $[y_1(t-1), u_1(t-1), y_2(t-1), u_2(t-1), u_3(t-1)]$  through the coefficient matrix shown in the left hand side of Equation 5.68, the coefficient matrix should be manipulated as shown in the right hand side of Equation 5.68 in order to estimate the direct impulse response from  $u_1$  to  $y_2$ .

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0.2 & 0.33 & 0 & 0 & 0 \\ 0 & 1 & 0.4 & 1 & 0.3 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0.33 & 0 & 0 & 0 \\ 0 & 1 & 0.4 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \end{bmatrix} \quad (5.68)$$

Note that if the model is of  $p_{th}$  order instead of  $1_{st}$ , the same elements should be forced to zero in all the coefficient matrices in the model. The impulse response from  $u_1$  to  $y_2$  estimated from the re-arranged model is the direct impulse response. Graphically, the re-arranged model can be represented as in Figure 5.2 where the above mentioned connections are removed by making their corresponding transfer functions equal to 0. Impulse response coefficients will be estimated based on the re-arranged SVAR model where the effect of feedback loops and other variables are removed from the relation between  $u_1$  and  $y_2$ .

The transfer function from  $u_1$  to  $y_2$  through  $u_3$  can also be estimated in a similar way. In this case, since  $y_2$  should only be a function of  $u_3$ , all the coefficients in the equation of  $y_2$  should be zero other than the coefficients corresponding to lagged  $y_2$  and  $u_3$ .  $u_3$  should also be only a function of  $u_1$  by making all the coefficients equal to 0 except for coefficients corresponding to the history of  $u_3$  and  $u_1$ . Again  $u_1$  itself should not be a function of any other variable in estimation of this transfer function. The graphical representation of the re-arranged model appropriate for estimation of this desired transfer function is shown in Figure 5.3.

By forcing appropriate coefficients in the estimated SVAR model to be 0, any specific transfer function can be individually estimated in the form of impulse re-

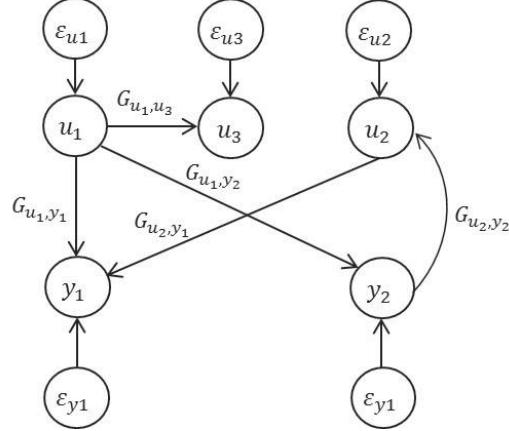


Figure 5.2: Removing the indirect connections from  $u_1$  to  $y_2$

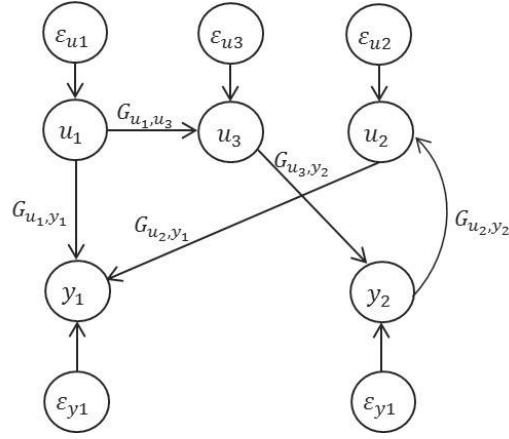


Figure 5.3: The connection between  $u_1$  and  $y_1$  through  $u_3$

sponse. For example, if it is desired to estimate the transfer function from  $u_t$  to  $y_t$  going through  $z_t$  and  $w_t$ , we need to make each one of the variables only a function of their own lags and the lags of their respective input variable. The impulse response estimated from the properly manipulated SVAR is the desired transfer function.

The procedure presented in the chapter is utilized in order to decompose the variations transferred between the 5 variables in the simulated model shown in Figure 5.4. Only the details of decomposing the impulse response transferred from  $u_2$  to  $y_1$  is presented here in details. The total transfer function from  $u_2$  to  $y_1$  in terms of impulse response is plotted in Figure 5.5 along with the estimated 90% confidence intervals. It can be observed that the uncertainty in the estimated impulse response generally increases with time horizon. The impulse response converges to 0 for time horizon larger than 15 samples. The total energy transferred from an impulse in  $u_2$  to  $y_1$  can be estimated as 5.65.

In order to estimate the interaction-free energy transferred from  $u_2$  to  $y_1$ , the

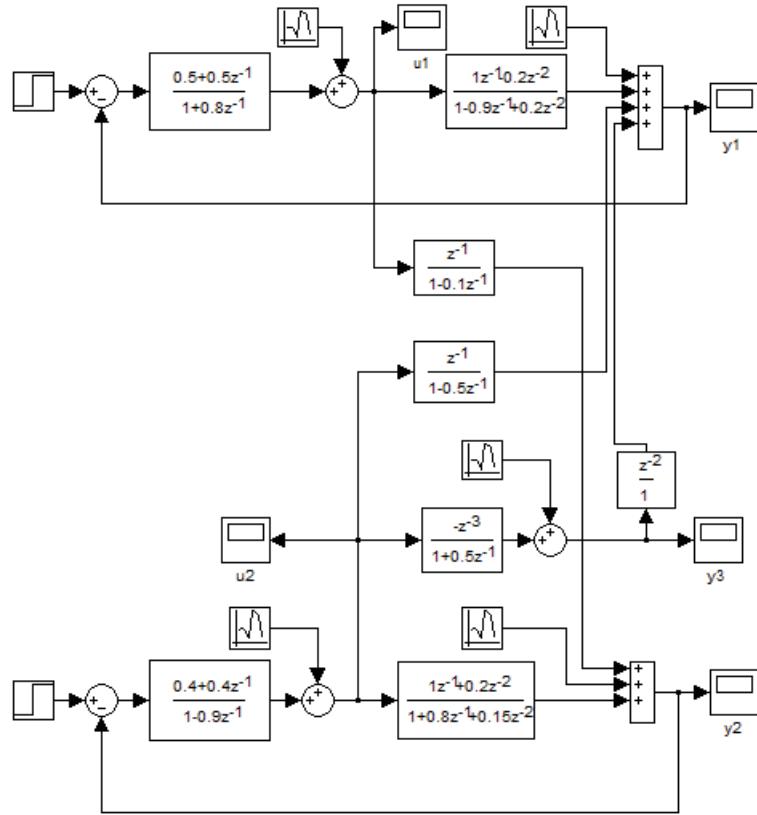


Figure 5.4: Simulated process

direct impulse response is estimated by making all the coefficients in the equation of  $y_1$  equal to 0 except for coefficients corresponding to lagged  $y_1$  and  $u_2$ .  $u_2$  is also a function of itself. Figure 5.6 plots the direct impulse response from  $u_2$  to  $y_1$  along with the confidence intervals. The estimated direct energy transfer is 1.3 which is devoid of the effect of the two feedback loops and variable  $y_3$ .

To check the effect of closing the feedback loop of  $y_1$  on the impulse response transferred from  $u_2$  to  $y_1$ , we can leave  $y_1$  to be a function of both its own history and the history of  $u_1$  in addition to  $u_2$ . In the equation of  $u_1$  also the coefficients corresponding to the history of  $u_2$ ,  $y_1$  and  $u_1$  should remain unchanged while the coefficients corresponding to any other variable should be forced to 0.  $u_2$  itself should not be a function of any other variable to ensure that there is no influence of its own loop in the impulse response to be estimated. The impulse response estimated from this re-arranged model from  $u_2$  to  $y_1$  is plotted in Figure 5.7 which is different from the direct impulse response due to the effect of the first feedback loop. The comparison between these two impulse responses can indicate the performance of the controller in rejecting the disturbance caused by  $u_2$ . It can be seen that closing the

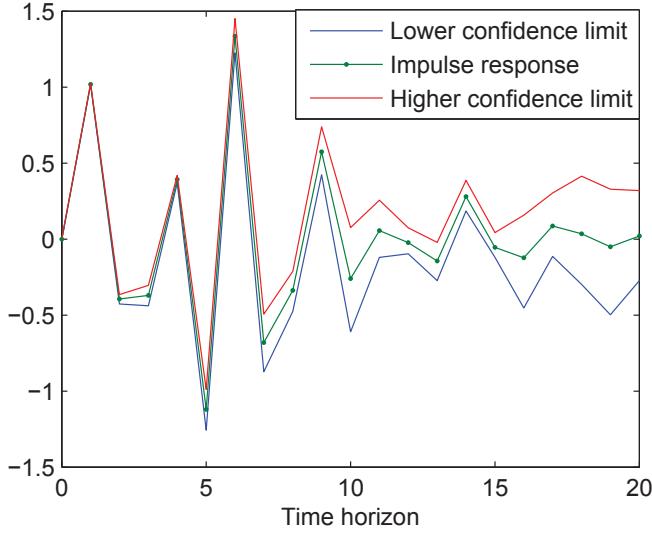


Figure 5.5: Total impulse response from  $u_2$  to  $y_1$  along with 90% confidence interval

loop has reduced the energy transferred to  $y_1$  due to an impulse in  $u_2$  to 1.12.

It is also possible to check the influence of closing the second loop in the transferred impulse response. Figure 5.8 shows the impulse response transferred from  $u_2$  to  $y_1$  while both the loops are closed. Comparing Figure 5.7 with 5.8 shows that closing the second loop increases the energy transferred to  $y_1$  to 1.56. This energy only excludes the energy transferred through variable  $y_3$  from estimation.

The variation in  $y_1$  due to an impulse in  $u_2$  which is independently transferred through  $y_3$  can also be estimated. The impulse response of  $y_1$  to  $u_2$  through  $y_3$  is plotted in Figure 5.9 and the transferred energy equals to 1.26.

The possibility to estimate how much energy is transferred from each path as well as verifying the performance of the controller, is valuable when trying to reduce the variation in  $y_1$  due to a change in  $u_2$ . It should be noted that as explained in [100], summation of the estimated impulse responses is not equal to the total impulse response. The reason is that the impulse responses interfere with each other and may compensate the influence of each other.

## 5.9 Case study

A multiple tank system as shown in Figure 5.10 is considered for interaction analysis. The equipment is composed of three tanks and the levels of the top two tanks are controlled with PI controllers in feedback loops. The level trajectories of the three tanks are plotted in Figure 5.11 with small perturbations around the steady state

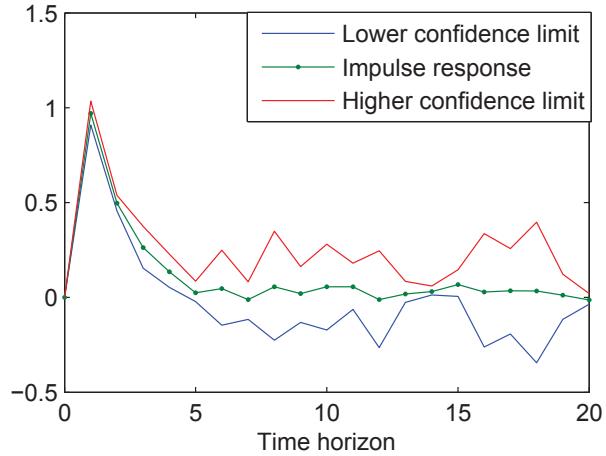


Figure 5.6: Direct impulse response from  $u_2$  to  $y_1$  along with 90% confidence intervals

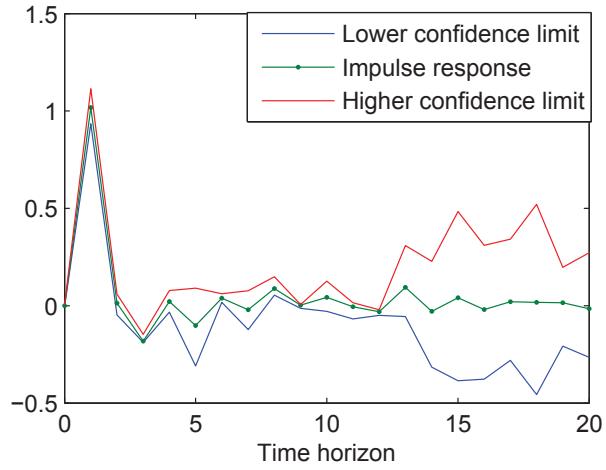


Figure 5.7: Impulse response from  $u_2$  to  $y_1$  along with 90% confidence intervals while only the first loop is closed

operation.

Figure 5.12 plots the estimated impulse responses from the top and bottom tanks to the middle one respectively. As is expected, the impulse response coefficients from the top tank to the middle one are significantly different from 0. On the other hand, impulse response coefficients from the bottom tank to the middle tank demonstrate that there is no influence from the bottom tank on the middle one.

Figure 5.13 plots the total impulse responses from the top and middle tanks to the bottom tank. There are some impulse response coefficients significantly different from 0, indicating the influence of the middle tank on the bottom one. The third panel of the figure also plots the direct influence of the top tank on the bottom one. By comparing the total and direct impulse responses of the top tank to the bottom

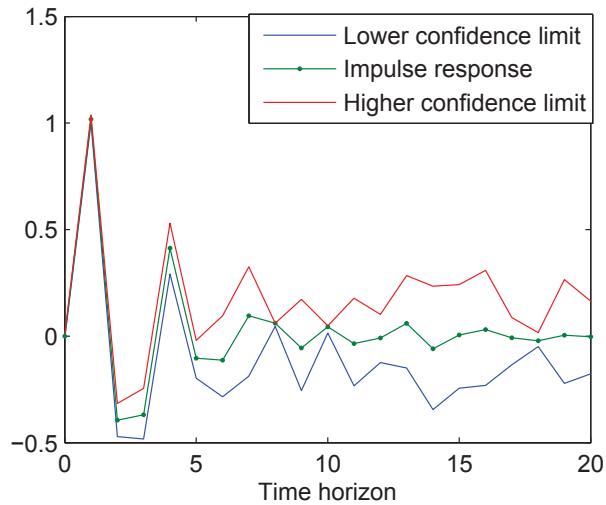


Figure 5.8: Impulse response from  $u_2$  to  $y_1$  while both the loops are closed

one, it can be inferred that that top tank can reach the bottom tank only through the middle tank. Thus, the above data-based analysis does match the physical setup. This illustrates the capability of the proposed method in capturing asymmetrical causality relations between variables as well as decomposing impulse responses in order to infer the actual relation between process variables from data.

## 5.10 Summary

A novel method is proposed in this chapter in order to decompose the estimated transfer function between variables to independent transfer functions, each corresponding to a specific path along which the input can influence the output. The advantage of the proposed method is that it can detect the different path along which variables can influence each other and it provides an estimation of the strength of the different connections between the variables. It also provides a way to check the performance of the controller in disturbance rejection. Estimation of the transfer functions is performed under Bayesian framework. A method to properly design the prior distribution for the parameters to avoid over-fitting as well as the method to sample from the posterior distribution of parameters is presented in this chapter. In addition, a methodology to detect instantaneous relations between the variables only from the data is proposed in the last section of the chapter.

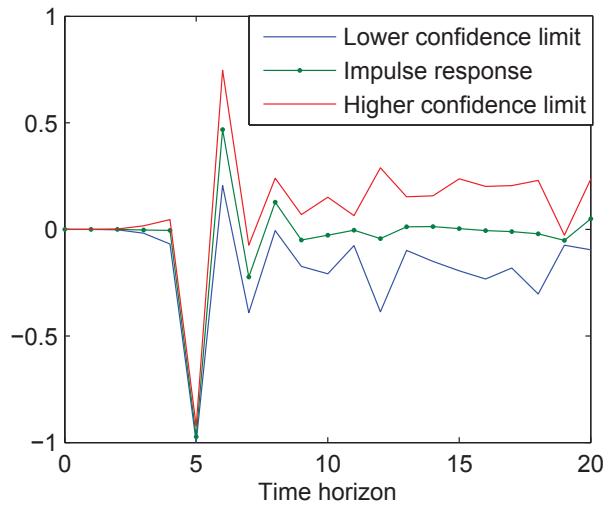


Figure 5.9: Impulse response from  $u_2$  to  $y_1$  through  $y_3$

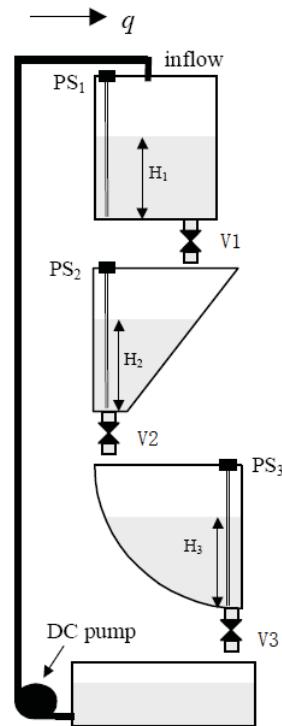


Figure 5.10: Schematic of the multi-tank system under study

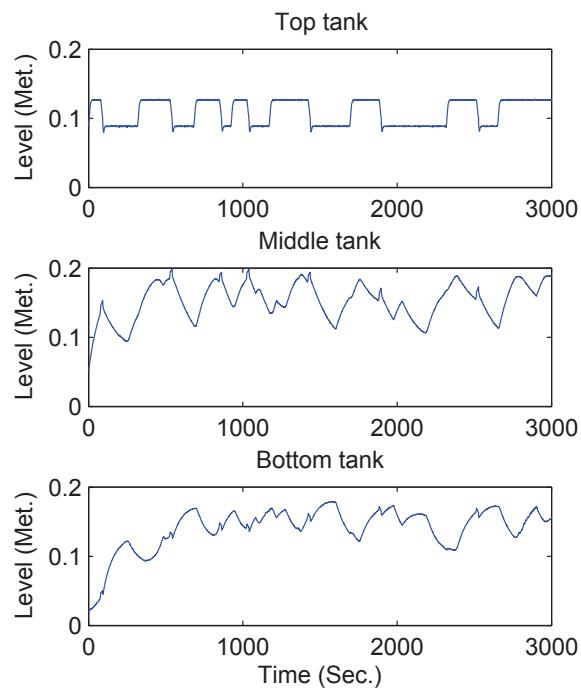


Figure 5.11: Level measurement of the three tanks under study

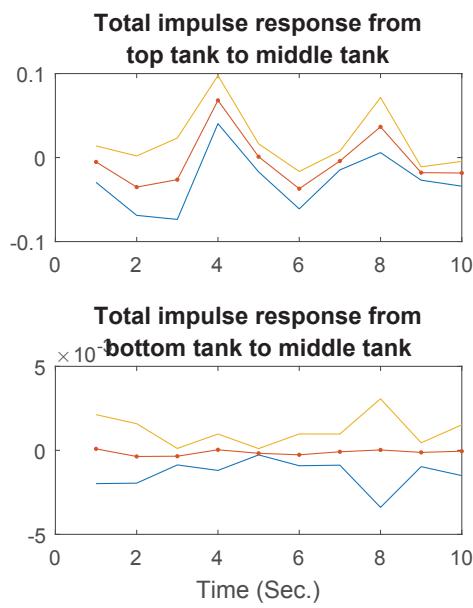


Figure 5.12: Impulse responses to the middle tank

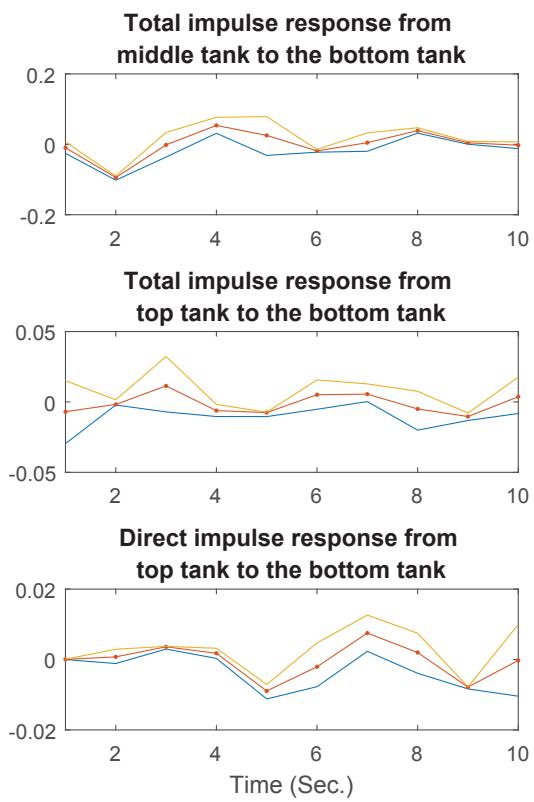


Figure 5.13: Impulse responses to the bottom tank

# **Chapter 6**

## **Information transfer methods in causality analysis of process variables**

## 6.1 Abstract

This chapter studies mutual information and transfer entropy for detection of cause and effect relationships between industrial process variables. Mutual information quantifies the amount of dependency between process variables, while transfer entropy detects the direction of information flow between the variables. The chapter overviews the existing definition and limitations of these two quantities and proposes an algorithm by combining and extending these two quantities for more reliable identification of causal relationship between process measurements. Detection of causal relationships between plant variables is useful for diagnosis of the root cause of a distributed fault in the process. It also helps predicting the effect variables. The proposed method is illustrated through an industrial case study.

## 6.2 Introduction

Causality analysis has been of great interest in various fields of science, engineering and economics. For a common understanding about the cause and effect relationship, the Wiener-Granger definition for causality is quoted here. "For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one". There are also two conditions: the cause should occur before the effect and the cause should contain some information about the effect that cannot be found in any other variable (including the effect variable itself).

The motivation of causality analysis is to identify the variables that are affecting the values or variations of another variable of interest. It provides the basis to make a good prediction for the desired variable or to control it by controlling its cause variables. From the control engineering perspective, the desired variable might be the quality of the product, an important variable considering safety issues or any other variable that should be maintained in a specific operation range.

The other motivation for causality analysis is to develop a causal map for a plant by analysing the relationships between all the variables of the plant. A causal map is helpful in root cause diagnosis of plant wide disturbances. In the cases in which a number of plant variables are out of the operation range or carry the same type of fault, a causal map can be used to detect the propagation path of the fault through the process.

Building causal maps for fault diagnosis has been studied in the literature from different perspectives. Some use process knowledge such as process flow sheet and

control configuration to develop a digraph for the process [101, 102]. Bauer et al. [103] has applied the transfer entropy method for developing a causal map based on the data.

Some other methods for root cause diagnosis of plant wide disturbances are based on principal component analysis in the frequency domain [104] and spectral envelope method to detect the oscillatory variables and identify the most likely root cause [24]. Stockmanna et al.[105] used the K nearest neighbor imputation method to estimate the time delay between the variables and eventually identify the source of the disturbance.

This work proposes a probabilistic method using the concept of information transfer between process variables to detect the cause and effect relationships. Transfer entropy provides an asymmetric measure of the information flow between two variables. It quantifies the improvement in the prediction of a variable by considering the history of a second variable in addition to the history of itself. Transfer entropy, if implemented appropriately, is a reliable method for detection of cause and effect relationships between the variables.

The problem with implementing transfer entropy for industrial data is that its reliable implementation is computationally impossible due to the reasons that will be discussed in Section 3. It is shown that using mutual information along with differential mutual information as proposed in this work, provides information about the type of the relationship between process variables. This information can be used to tune the parameters of the transfer entropy for a more reliable result with reduced processing time.

The chapter is organized as follows. Section 6.3 provides the definition and history of mutual information and transfer entropy. Section 6.4 discusses the method for implementing transfer entropy and its associated problems in application. This section also includes application of mutual information and its differential version for causality analysis. Subsection 6.4.5 investigates parametrising the transfer entropy for a more reliable result. Section 4 presents an industrial case study and Section 5 concludes the chapter.

## 6.3 Mutual information and transfer entropy

Transfer entropy was proposed by [106] as a special version of Kullback entropy. Kullback entropy [107] is defined as:

$$K_I = \sum_i p(i) \log \frac{p(i)}{q(i)} \quad (6.1)$$

where  $p(i)$  is the true underlying probability distribution function (PDF) of random variable  $I$  and  $q(i)$  is assumed probability distribution function of the variable. The summation is over all possible states of variable  $I$ .

Kullback entropy measures the added uncertainty due to the difference between the assumed PDF and the real PDF of the variable  $I$ . More illustration on the Kullback entropy needs the Shannon entropy that is defined by the following formula:

$$H_I = \sum_i p(i) \log \frac{1}{p(i)} \quad (6.2)$$

Shannon entropy was proposed by [108] and quantifies the uncertainty of variable  $I$  based on its probability distribution  $p(i)$ . Shannon entropy was originally introduced as the average number of bits that is needed to encode the value of variable  $I$  with a known probability. If the variable is a deterministic value and there is no randomness in the process,  $p(i)$  equals to one and there is no need for encoding. As a result Shannon entropy will be zero. By increasing the number of possible values of the variable  $I$ , the entropy of the variable increases.

Kullback entropy ( $K_I$ ) is the difference between Shannon entropies of two probability distributions  $p(i)$  and  $q(i)$ . If the real probability distribution of variable  $I$  is known as  $p(i)$ ,  $K_I$  measures how much uncertainty is added by assuming another probability distribution  $q(i)$  instead of  $p(i)$ . Kullback entropy has a non negative value and is zero in the case that  $q(i) = p(i)$ .

A special kind of Kullback entropy, which is called mutual information, measures the amount of dependency between two random variables  $I$  and  $J$  and is defined as in Equation 6.3 [109].

$$M_{IJ} = \sum_{i,j} p(i,j) \log \frac{p(i,j)}{p(i)p(j)} \quad (6.3)$$

Here, the assumption of independency of variables  $I$  and  $J$  is compared with their joint probability distribution. If  $I$  and  $J$  are independent variables, then  $p(i,j) = p(i)p(j)$  and the mutual information will be zero. In the case that variable  $J$  is completely determined by variable  $I$ , then, the mutual information equals to the uncertainty of variable  $I$ . The reason is that  $p(i,j) = p(j|i)p(i) = p(i)$ . So, Equation 6.3 simplifies to  $\sum_{i,j} p(i) \log \frac{1}{p(j)} = \sum_i p(i) \log \frac{1}{p(i)}$  which is the Shannon entropy of variable  $I$ .

Mutual information is a symmetric quantity and its value does not change by reordering the  $I$  and  $J$  in Equation 6.3. It provides a measure of dependency between the two variables. However, it does not provide any information about the direction of information flow between the variables.

It is possible to gain a sense of directionality from the mutual information by incorporating a time lag in one of the variables. The so called time lagged mutual

information has the following definition:

$$M_{IJ}(h) = \sum_{i,j} p(i_t, j_{t-h}) \log \frac{p(i_t, j_{t-h})}{p(i_t)p(j_{t-h})} \quad (6.4)$$

where  $h$  is the time lag.

By trying different values for  $h$ , we can find the time lag for which there is the maximum dependency between variable  $I$  and lagged variable  $J$ . In this time lag, variable  $J$  has the most influence on the future value of variable  $I$  (when  $h > 0$ ). This time lag can be assumed as the time delay from variable  $J$  to  $I$ .

The main advantage of time lagged mutual information is that it is not symmetric between variables  $J$  and  $I$ . By time lagging variable  $I$  instead of  $J$ , one can find the time lag at which variable  $I$  has its most influence on the future value of variable  $J$ .

Figure 6.1 shows the mutual information (called dependency hereafter) between a variable and its past. In Figure 6.1, there is a strong peak at time lag zero, which means that the variable is mostly dependent to itself when there is no time lag, which is expected.

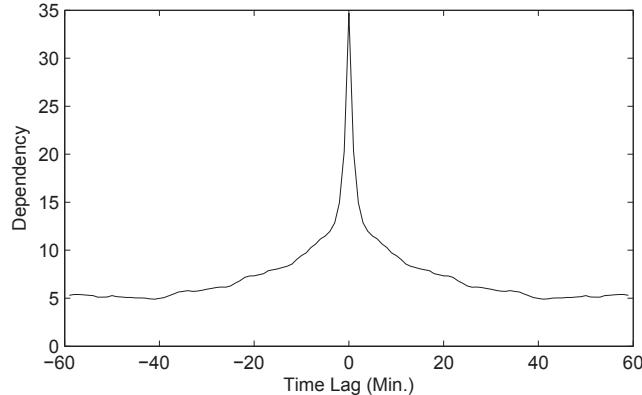


Figure 6.1: Time lagged mutual information between a variable by itself

Figure 6.2 shows the dependency between input and output of an industrial process that will be further elaborated in Section 4. Positive time lags represent the case that the input is lagged while negative time lags correspond to the case that the output is lagged.

As can be seen in Figure 6.2, there is a high amount of dependency between time lagged input and current value of the output variable. This means that input variable can be used in prediction of the output. However, the output signal cannot help in prediction of the input, which is as expected (the input is the product of the previous plant).

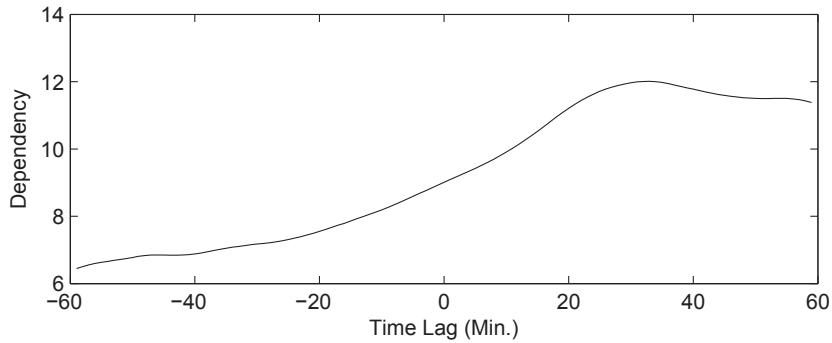


Figure 6.2: Time lagged dependency between input and output of an industrial process

Another point that can be inferred from Figure 6.2 is that the input is the most dependent on the output at a time lag of about 30 minutes. The time lag corresponding to the highest dependency may be taken as the time delay between the two variables if there is no feedback relation between the two.

Problem of using time lagged mutual information for detecting the direction of information flow between the variables is that it does not take into account the influence of the history of a variable itself on its own future value. Consider the case that two variables are linearly correlated to each other and the first variable is just a multiple of the second one. For example, Figure 6.3 shows trajectories of two variables. The second variable is just 0.5 multiplied by the first variable. Time lagged mutual information curve between the second variable and the history of the first variable is depicted in Figure 6.4.

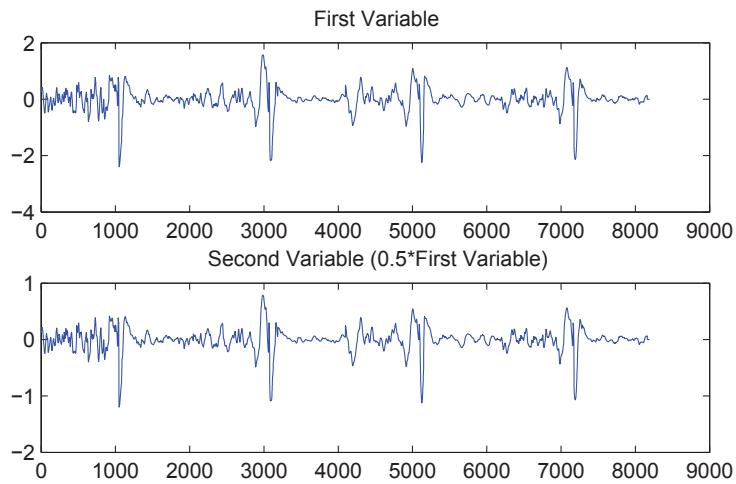


Figure 6.3: Example of two variables that are linearly corelated to each other

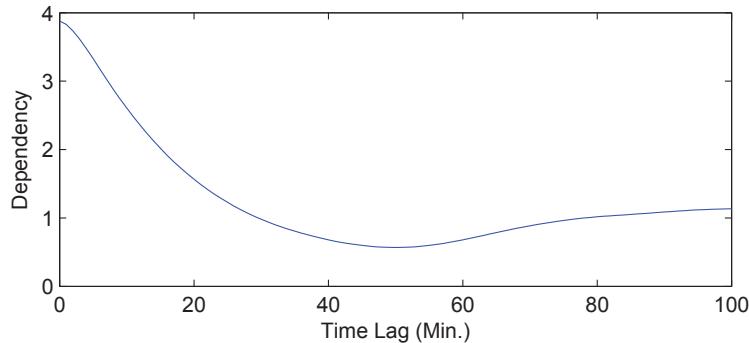


Figure 6.4: Time lagged dependency between the second variable with the history of the first variable in Figure 6.3

Figure 6.4 tells us that there is a high amount of information flow from the first variable to the second one. So, it can mistakenly be inferred that the first variable is the cause to the second variable. The point is that even though there is information flow from the first variable to the second one, but this information is contained in the history of both variables. Figure 6.5 shows the dependency of the second variable on its own history. By comparing Figures 6.4 and 6.5 , it can be seen that the dependency curve between the second variable and the history of the first variable is the same as its dependency with the history of itself. This implies that the history of the first variable does not contain any additional information compared to the history of the second variable. This contradicts with the condition of causality that was mentioned in the introduction.

A high dependency value does not imply causality by itself. It is necessary to make sure that the cause variable has unique information about the effect variable that is not contained in any other variable including the history of the effect variable itself.

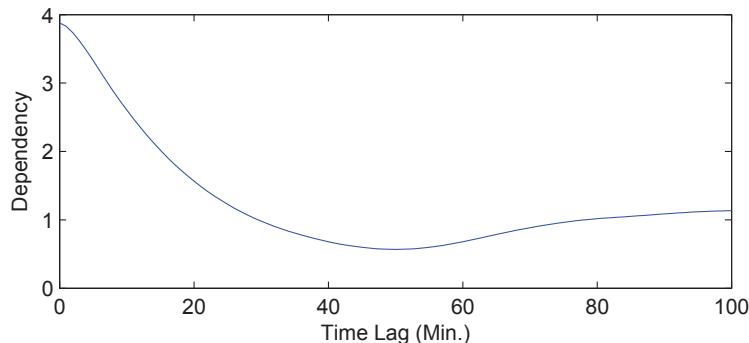


Figure 6.5: Time lagged dependency between the second variable with its own history

Previous discussion illustrates the motivation for development of transfer entropy.

Transfer entropy takes the history of a variable itself into account when performing causality detection.

Transfer entropy is defined as a special type of Kullback entropy that takes care of the dynamics of information flow between two variables and is defined by the following equation [106]:

$$T_{J \rightarrow I} = \sum p(i_t, i_{t-1}^k, j_{t-1}^l) \log \frac{p(i_t | i_{t-1}^k, j_{t-1}^l)}{p(i_t | i_{t-1}^k)} \quad (6.5)$$

where  $k$  and  $l$  are the length of the considered history of variables  $I$  and  $J$ ,  $i_{t-1}^k = (i_{t-1}, \dots, i_{t-k})$  and  $j_{t-1}^l = (j_{t-1}, \dots, j_{t-l})$

Parameter  $l$  at least equals to 1. The parameter  $k$  is the order to model variable  $I$  as a Markov process. In mathematical form the  $k$ th-order Markov process implies:  $P(i_{t+1} | i_t, \dots, i_{t-k+1}, i_{t-k}) = P(i_{t+1} | i_t, \dots, i_{t-k+1})$ .

This means that the state of variable  $I$  at time  $t - k$  or before does not have any influence on its future state value given  $i_t, \dots, i_{t-k+1}$ .

For computation purposes, transition probabilities in Equation 6.5 can be replaced by joint probabilities. This results in the following equation for transfer entropy:

$$T_{J \rightarrow I} = \sum p(i_t, i_{t-1}^k, j_{t-1}^l) \log \frac{p(i_t, i_{t-1}^k, j_{t-1}^l) p(i_{t-1}^k)}{p(i_t, i_{t-1}^k) p(i_{t-1}^k, j_{t-1}^l)} \quad (6.6)$$

Transfer entropy measures the effect of the second variable  $J$  on the future state of variable  $I$ . If the  $l$  past values of  $J$  have no effect on the future values of the variable  $I$ , given  $k$  past values of  $I$ , then  $p(i_t | i_{t-1}^k, j_{t-1}^l) = p(i_t | i_{t-1}^k)$  and transfer entropy will be zero. In the case that variable  $J$  influences variable  $I$ , transfer entropy has a positive value.

Transfer entropy has two characteristics. The first one is that it is asymmetric between  $I$  and  $J$ . Changing the ordering of  $I$  and  $J$  will change the direction of the influence.

The second characteristic is that transfer entropy omits the common information contained in both  $I$  and  $J$  due to correlation between the two variables. If the dependency between  $I$  and  $J$  is due to correlation, the information in the history of variable  $J$  is also contained in the history of variable  $I$ . Inclusion of the history of variable  $I$  in the calculation of transfer entropy prevents mistaking the common information contained in the history of both  $I$  and  $J$  as information transferred from  $J$  to  $I$ .

## 6.4 Practical issues and solutions

### 6.4.1 Computation method

Application of transfer entropy for analysis of real industrial data has many challenges. The first one is that the data values have a fine resolution and it is necessary to bin the data to have finite number of possibilities for estimation of the probability densities.

As is mentioned in [110], using Kernel estimators is a better way in comparison with histograms to estimate the probabilities for a reliable estimation of transfer entropy. Following is a description of the method used in this work. The first step in density estimation is binning the data. Wand [111] made some suggestions on the selection of the number of bins for the data and the method of binning. As is discussed in [111] linear binning of the data generates less error compared to simple binning.

For illustration of linear binning, consider a sample of a univariate random variable  $X$ . The value of  $x$  is between the values of two bins  $a$  and  $b$  such that  $a < x < b$ . In simple binning, all the weight of  $x$  goes to the bin nearest to it. In linear binning both of the bins  $a$  and  $b$  get a weight according to their distance from  $x$ . The weight added to bin  $a$  is  $\frac{b-x}{b-a}$  and the weight added to bin  $b$  equals to  $\frac{x-a}{b-a}$ . For two dimensional data, all the four bins surrounding a sample of data get a proportional weight and the same is for higher dimensional data.

After the data is binned, the Kernel method can be used to estimate the probability density. The Kernel estimator for a random variable  $X$  with observations  $(x_1, \dots, x_n)$  is defined by Equation 6.7.

$$\hat{f}_\sigma(X) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - x_i}{\sigma}\right) \quad (6.7)$$

where  $n$  is the number of data samples,  $\sigma$  is the kernel width and  $K$  is the Kernel function. Selection of Kernel width is widely discussed in the literature [112].

The Kernel function taken in this work is the Gaussian function with the following definition:

$$K[x - x_i] = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right) \quad (6.8)$$

where  $\sigma$  is the Kernel width. For higher dimensional data, the Kernel function is a multiplication of the individual Kernels. For example, for two dimensional data  $x, y$ , Kernel function is as:

$$K[(x - x_i), (y - y_j)] = \frac{1}{2\sigma^2\pi} \exp\left(-\frac{(x - x_i)^2 + (y - y_j)^2}{2\sigma^2}\right) \quad (6.9)$$

The Kernel width can be constant for all dimensions of the data or can be variable. In this work, the data is normalized to have the same standard deviation for all the variables. Therefore, the Kernel width is constant.

A good estimation of probability density needs a large amount of data to be considered in estimation. Larger amount of data generates less finite sample error in probability density estimation. The other important point is that the data should contain sufficient information in it. In other words it should represent all possible values for the variables.

### 6.4.2 Confidence limit

As was discussed above, using finite sample data for estimation of probabilities generates random error in calculations. Dependency and transfer entropy values calculated by this method even for two completely random and independent variables will not be zero due to the finite samples used for calculation. Therefore, we need a method to assess the uncertainty of the dependency or transfer entropy estimation.

For this reason Marschinski et al. [113] proposed the concept of effective transfer entropy as the difference between the usual transfer entropy calculated from the data and the transfer entropy calculated after randomly shuffling the data. By shuffling the data, the dependency relationships are supposedly destroyed and the transfer entropy calculated after the data shuffling is due to the numerical error. Palus et al. [114] proposed the usage of surrogate time series to establish a critical test value in order to examine the reliability of measured information transfer.

In general, determination of uncertainty for the calculated transfer entropy is difficult due to complexity of transfer entropy calculation itself. That is the reason to use Monte Carlo method to assess the uncertainty of the estimations. In this work generation of random data with the same sample size as the original data is considered instead of surrogate or randomly shuffled data. The advantage of using complete randomly generated data is the certainty of non-existence of any relationship between two random variables. Since the random data pairs are generated independently, any dependency and transfer entropy values calculated from them are due to the error in calculation such as finite samples.

The method that is used here is to generate a number of random variables with the same sample size as the original data and perform the same analysis that is done for the original time series for each pair of the random data. The result of the analysis will be some random numbers that can be attributed to computation errors. A hypothesis test can be developed using the mean value ( $\mu_{random}$ ) and standard deviation ( $\sigma_{random}$ )

of the obtained results. The null hypothesis is that there is no dependency between the real time series and the calculated value is due to the estimation error. If the dependency value is higher than  $\mu_{random} + 3\sigma_{random}$ , then the dependency between the time series is confirmed.

### 6.4.3 Computational load

The most restriction in computation of transfer entropy is the need to estimate  $k+l+1$  dimensional probability density as shown in Equation 6.6. Estimation of a joint PDF with  $k+l+1$  dimension needs  $N \times (2\sigma + 1)^{k+l+1}$  summations where  $N$  is the number of data samples and  $\sigma$  is the Kernel width. Estimation of all the PDFs in Equation 6.6 requires  $N(2\sigma + 1)^k \times [(2\sigma + 1)^{l+1} + (2\sigma + 1)^l + 2\sigma + 2]$  summations and calculation of transfer entropy needs  $n^{k+l+1}$  summation at the end ( $n$  is the number of bins in each dimension).

As was mentioned in Section 2, parameter  $k$  is the order of the Markov model for the process  $I$ , and  $l$  at least equals to 1. Considering that many chemical processes involve chemical reactions and heat exchanges, the dominant settling time for the process can be quite long. So, parameter  $k$  will have a large value on average (assuming a small enough sampling time). It makes it impractical to calculate the transfer entropy as defined in Equation 6.6, since it needs estimation of a high dimensional probability density.

The second restriction in implementing transfer entropy method is the effect of finite sample data in estimation of probability densities. As mentioned in Section 3.1, use of larger set of data reduces the error in estimation of probability densities. Also, data should be representative of different possible values of the variables of interest. Therefore, a large amount of data needs to be processed in estimation of transfer entropy. This is another factor that significantly increases the computation.

With the two problems mentioned above, a full implementation of the transfer entropy is not practically feasible. Alternative approach must be found, and we need to reduce the dimension of the probability density while tuning the parameters in a way to get a reliable result. The next section illustrates the algorithm to use transfer entropy in an efficient manner.

### 6.4.4 Application of dependency and differential dependency curves

As discussed in Section 2, time lagged mutual information also contains directionality in it. The main difference between time lagged mutual information and transfer

entropy is that transfer entropy takes into account the influence of history of the effect variable itself on its transition to a new state. This prevents mistaking the information that is in the history of both variables as the information flow from the second variable.

This characteristic is most important in the case that a third variable influences both  $I$  and  $J$  simultaneously without difference in time delays. The effect of the third variable exists in the history of both  $I$  and  $J$ . In this case, each variable has the dependence on the history of the other one, even if there is no information flow between the two.

If the dependency between the two variables is only due to the simultaneous effect of other variables, then the dependency curve usually has an approximately flat and symmetrical shape. This is the case when two variables are linearly correlated to each other due to other variables. For example, Figure 6.6 depicts the time lagged dependency between variables plotted in Figure 6.3. Positive time lags represent the case that the first variable is lagged, while negative time lags are the case that the second variable is lagged. As is seen in Figure 6.6, the two variables have the same values of dependency on the history of the other variable. This is because both variables carry the same information in their history. Therefore, mutual information for negative and positive lags will have similar values.

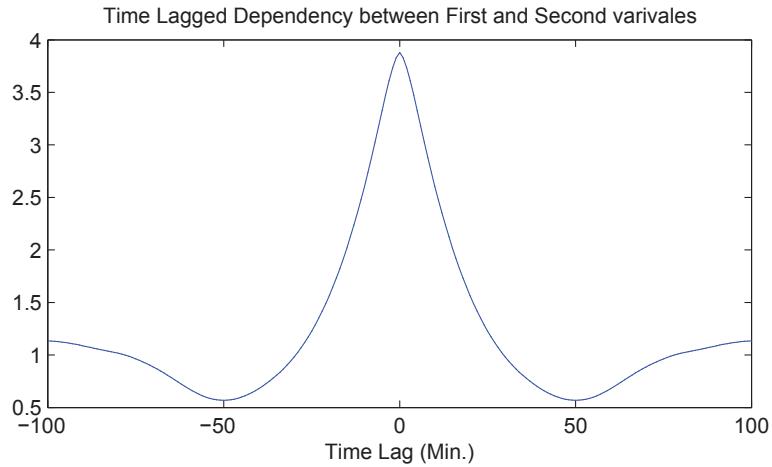


Figure 6.6: Time lagged dependency between first and second variables in Figure 6.3

Flat or symmetrical dependency curve implies dependency due to a third variable or no dependency at all (if the dependency values are less than the confidence limit). A flat dependency curve can also mean that the variables are strongly correlated to each other but there is no information flow between them. Therefore flat or symmetrical

dependency curve is a sign of no causal relationship even if there exists a strong correlation.

Studying the shape of the time lagged dependency between two variables, helps in detection of the probable type of the relationship between two variables and their respective time delays. In addition to the dependency between the original variables, dependency between their differenced version also provides useful information. Differenced time series are generated by calculating the differences between adjacent values of each time series ( $\Delta x = x_{i+1} - x_i$ ).

Time lagged dependency between differenced time series, which will be called differential dependency hereafter, measures the amount of dependency between the rates of variations of two process variables. By considering the differential dependency, we may also determine if variations of a process variable is transferred to other process variables and causes variations in them. Time lagged differential mutual information between two variables  $X$  and  $Y$  is estimated by the following equation:

$$DM_{XY}(h) = \sum_{\Delta x, \Delta y} p(\Delta x_t, \Delta y_{t-h}) \log \frac{p(\Delta x_t, \Delta y_{t-h})}{p(\Delta x_t)p(\Delta y_{t-h})} \quad (6.10)$$

As an example for illustration of the role of differential dependency, consider a simple relationship between two variables  $y_t = \sum_{k=0}^{t-5} x_k$ . Figure 6.7 plots the input ( $x$ ) and output ( $y$ ) data.

Figure 6.8 depicts the time lagged dependency and differential dependency between the input and output. As can be seen, the dependency curve does not provide any information about how these two variables are related. The differential dependency curve shows a peak at time lag of 5 and 6, which means that the change in the output is dependent on the change in the input at 5 samples earlier. This is confirmed by knowing that  $\Delta y_t = y_t - y_{t-1} = x_{t-5} = \Delta x_{t-5} + \Delta x_{t-6} + \dots$ .

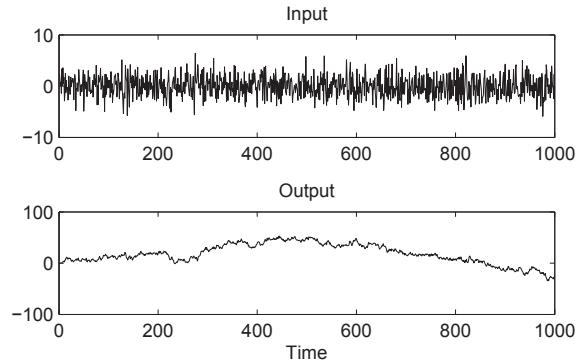


Figure 6.7: Input and output data of the summation model

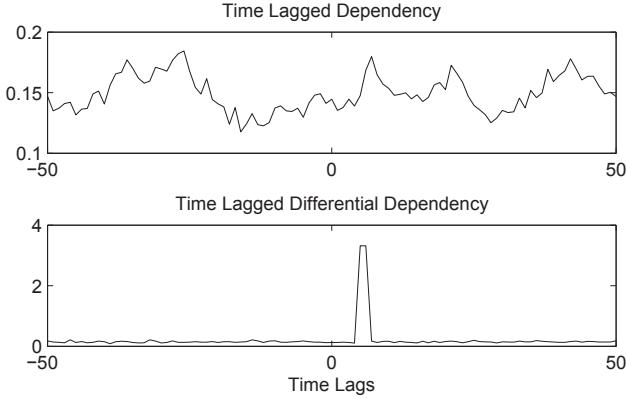


Figure 6.8: Dependency and differential dependency between data in Figure 6.7

When there is an integrator in the process, differential dependency is more useful than original dependency. It is because the actual values of the effect variable (the integration unit output) do not follow the actual values of the input variable. The input variable is mainly causing the fluctuations in the effect variable. Cases like this may appear while analyzing real industrial data in situation when there is some kind of integration unit in the process, like filling out a tank. For processes involving more than one integration unit, higher order of differencing might be necessary to identify causality.

Differencing removes the trend in the data and makes it more stationary. As a result, the differential dependency usually only has a peak value corresponding to the time delay between the two variables and is otherwise less than the confidence limit. Therefore, differential dependency could be a better alternative method for estimation of the time delay between variables.

Use of both dependency and differential dependency curves between two variables usually provides sufficient information to detect probable causality between the variables. To perform causality analysis on a set of variables, the first step is to plot the time lagged dependency and differential dependency curves between each pair of the variables. Cases that have non-flat dependency curves or non zero differential dependency value, should be considered for additional analysis. Transfer entropy is used afterwards to examine the certainty of the analysis.

#### 6.4.5 Parametrising the transfer entropy

As mentioned in Section 3.3, full implementation of the transfer entropy as in Equation 6.5 is not practical due to huge processing time. Only limited number of terms of the history of variables  $I$  and  $J$  can be used in the estimation of transfer entropy

instead of  $i_{t-1}^k$  and  $j_{t-1}^l$ . On the other hand, the result of the estimation is very sensitive to the choice of the time lags of variables  $I$  and  $J$  that are used in the transfer entropy estimation. Therefore, we need a method to find the important terms of the history of variables  $I$  and  $J$  to use for a reliable estimation of the transfer entropy.

One way to overcome this problem is to estimate the transfer entropy considering different parameters and analyse all the results to determine which set of parameters are generating a reliable result [103]. The problem with this method is that it takes a huge processing time to find out a reliable set of parameters for each set of data and the result will still be based on trial and error.

The method that is proposed in this work is based on utilizing the time lagged dependency and differential dependency curves. With these curves, we can find out the most important time lags in the history of variables  $I$  and  $J$ , and use this information to parametrise the transfer entropy. Estimation of the dependency curves requires  $N(2\sigma+1)(2\sigma+3)$  summations for estimation of probability densities in addition to  $n^2$  summations for calculation of the dependency value. The computation load is much less compared to the case of estimation of the transfer entropy, which was mentioned in Section 6.4.3. The curves provide reliable information about the probable type of the relationship between the variables and the most important time lags.

In the transfer entropy equation (Equation 6.5),

$j_{t-1}^l = (j_{t-1}, \dots, j_{t-l})$  can be replaced by  $j_{t-D_1}$ ;  $D_1$  is the lag corresponding to the maximum dependency between lagged variable  $J$  and  $I$ . This time lag corresponds to the time it takes until the variations of  $J$  is transferred to  $I$ . To better illustrate the importance of time lag selection, consider the relation between input and output of the case-study process that was shown in Figure 6.2. It takes about 30 minutes until the effect of the input is seen in the output. So, using input with, for example, 5 minutes time lag in the transfer entropy equation, yields the conclusion that the input does not have a significant effect on the future value of the output. This justifies the need for some priori information about the relationship between the variables for a more accurate transfer entropy estimation.

$i_{t-1}^k = (i_{t-1}, \dots, i_{t-k})$  also needs to be replaced with a few terms of the history of variable  $I$ . Selection of the time lag for variable  $I$  is more critical than variable  $J$ . The rationale behind inclusion of the history of variable  $I$  in the transfer entropy is to ensure that the information in  $j_{t-D_1}$  cannot be found in the history of  $I$  itself. So, one restriction is the need to include a term of the history of  $I$  with a time lag greater than or equal to  $D_1$ . There are two possible outcomes of the transfer entropy estimation considering  $j_{t-D_1}$  and a term of  $I$  with larger time delay. One is that the information of  $j_{t-D_1}$  is already contained in  $i_{t-D_1-h}$  for some non negative value

of  $h$ . So, transfer entropy estimation by considering  $j_{t-D_1}$  and  $i_{t-D_1-h}$  will be zero. The other possibility is that the information of  $j_{t-D_1}$  cannot be found in any term of variable  $I$  with time lag greater than or equal to  $D_1$ . This means that  $j_{t-D_1}$  has some unique information about the future value of  $I$  in addition to the history of  $I$  itself. In this case, even if the information of  $j_{t-D_1}$  can be found in a term of  $I$  with less time delay than  $D_1$ ,  $J$  will still be identified as the cause of the variable  $I$ . The reason is that the information is present in  $J$  earlier than in  $I$ .

Another reason for not including terms of variable  $I$  with time delays less than  $D_1$  in estimation of transfer entropy is the usual slow variations of the variables. The variables of chemical processes usually have low-frequency variations. For example, the value of a variable at time  $t$  is close to its value at time  $t - 1$ . Therefore, with the knowledge of the value of the variable at time  $t - 1$ , it is very likely to make a good prediction for its value at time  $t$  without additional information. This is the effect of near neighbourhoods. By including near neighbours of  $i_t$  in Equation 6.5, it is very likely that the knowledge of the history of  $J$  is not adding any improvement to the prediction of  $i_t$ . This prevents identification of the actual sources of the variations of variable  $I$ .

The last reason for not including terms of  $I$  with less time delay than  $D_1$  is that  $D_1$  is most probably the time delay from variable  $J$  to  $I$ . It is not possible that  $(i_{t-1}, \dots, i_{t-D_1+1})$  contain the information of  $j_{t-D_1}$ . Therefore, inclusion of these terms in the calculation of transfer entropy only increases the processing time without adding more accuracy to the result.

A reasonable choice for a term of the history of  $I$  to be included in the transfer entropy calculation is the term that has the most influence on  $j_{t-D_1}$ . The time lag corresponding to the highest influence of the history of  $I$  on  $J$  can be found using the dependency and differential dependency curves between different time lags of variable  $I$  and variable  $J$ . If the time delay from variable  $I$  to  $J$  is  $D_2$ , then the most reliable term of the history of  $I$  to be included in the estimation of transfer entropy is  $i_{t-D_1-D_2}$ . It is reasonable to include other terms of  $I$  with time delays more than  $t - D_1 - D_2$  for a more reliable result.

This method reduces the computation load while the reliability is increased by considering the most effective terms in the history of variables  $J$  and  $I$  in computation. As a result, the dimension of probability density is reduced significantly while the reliability of estimation is preserved.

To summarize, Equation 6.11 is the general formula for estimating the independent effect of a variable  $J$  on a variable  $I$ .

$$IE_{J \rightarrow I} = \sum p(i_t | j_{t-D_{JI}}, i_{t-(D_{IJ}+D_{JI})}) \\ \times \log \frac{p(i_t | j_{t-D_{JI}}, i_{t-(D_{IJ}+D_{JI})})}{p(i_t | i_{t-(D_{IJ}+D_{JI})})} \quad (6.11)$$

where  $D_{IJ}$  is the time delay from variable  $I$  to variable  $J$  and  $D_{JI}$  is the time delay from variable  $J$  to variable  $I$ .

The meaning of independent effect is that even though we know that variable  $J$  affects variable  $I$  after  $D_{JI}$  sample times, this effect might be due to the previous effect from variable  $I$  to  $J$  but not  $J$  itself. In other words, all of the information transferred from  $J$  to  $I$  might be found in the history of  $I$  itself. Equation 6.11 quantifies the amount of information that variable  $J$  transfers to variable  $I$  excluding the information that is transferred from  $I$  at  $t - D_{IJ} - D_{JI}$  to  $J$ . Following is the path diagram illustrating the relationship between  $I$  and  $J$ .  $i_{t-D_{IJ}-D_{JI}} \rightarrow j_{t-D_{JI}} \rightarrow i_t$

As was mentioned in Section 6.3, transfer entropy is an asymmetric quantity. By reversing the ordering of  $I$  and  $J$ , we can measure the amount of the independent effect of variable  $I$  on  $J$ . The equation for the independent effect of  $I$  on  $J$  is derived in the similar way as  $IE_{J \rightarrow I}$  and is defined as:

$$IE_{I \rightarrow J} = \sum p(j_t | i_{t-D_{IJ}}, j_{t-(D_{IJ}+D_{JI})}) \\ \times \log \frac{p(j_t | i_{t-D_{IJ}}, j_{t-(D_{IJ}+D_{JI})})}{p(j_t | j_{t-(D_{IJ}+D_{JI})})} \quad (6.12)$$

Again, in this case variable  $I$  has the most influence on  $j_t$  at time  $t - D_{IJ}$  and  $i_{t-D_{IJ}}$  gets the most influence from  $j_{t-(D_{IJ}+D_{JI})}$ , as is shown in the following path diagram:

$j_{t-(D_{IJ}+D_{JI})} \rightarrow i_{t-D_{IJ}} \rightarrow j_t$ . Equation 6.12 measures the amount of information added by considering  $i_{t-D_{IJ}}$  in addition to the history of  $J$  itself.

There is no need to include other terms of variable  $J$  with time lags less than  $D_{IJ} + D_{JI}$  in Equation 6.12. As was mentioned before, including other terms of variable  $J$  with time lags more than  $D_{IJ} + D_{JI}$  helps to get a more reliable result.

The variable with more independent effect can be determined by comparing Equations 6.12 and 6.11. The cause variable is the one with larger independent effect. This method is applicable in analysis of the relationship between any two variables.

#### 6.4.6 Case of two variables with common parent

In some cases, variables  $I$  and  $J$  have an identified common source variable  $Z$  as shown in Figure 6.9. Variable  $Z$  firstly affects variable  $I$  and after some time delay

it affects variable  $J$ . As an example, variable  $Z$  is the flow to a continuously stirred and heated tank, variable  $I$  is the bottom pressure of the tank and variable  $J$  is the temperature of the bottom fluid in the tank. Figure 6.10 depicts the schematic of the tank. The feed flow instantly affects the bottom pressure while it takes some time until it influences the temperature of the bottom fluid in the tank.

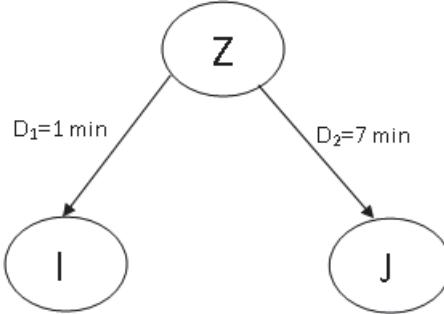


Figure 6.9: An example of two variables with a common source

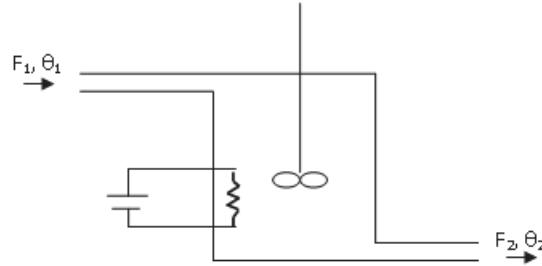


Figure 6.10: A continuously stirred tank

Using transfer entropy for variables  $I$  and  $J$  detects variable  $I$  as the cause to the variable  $J$ . The reason is that the variations of variable  $Z$  is transferred to variable  $I$ , several minutes earlier than to variable  $J$ . So, the history of variable  $I$  is helpful in prediction of variable  $J$ , but variable  $J$  cannot improve the prediction of a future value of  $I$ . At the end, a causality graph as in Figure 6.11 is obtained by using transfer entropy between pairs of variables  $Z$ ,  $I$  and  $J$ .

The problem with the causality graph in Figure 6.11 is that corresponding to the definition of causality, variable  $I$  might be the source of the variable  $J$  only if it contains additional information about variable  $J$  that cannot be found in variable  $Z$ . In this case estimation of a transfer entropy including all 3 variables  $Z$ ,  $I$  and  $J$  is required for correct causality detection. At least one term of the common source (variable  $Z$ ) should be included in the transfer entropy equation as in Equation 6.13.

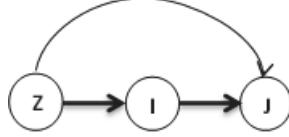


Figure 6.11: Causality graph for the example in Figure 6.9

$$\begin{aligned} IE_{I \rightarrow J} &= \sum p(j_t, j_{t-D_{JI}-D_{IJ}}, i_{t-D_{IJ}}, z_{t-D_{ZI}-D_{IJ}}) \\ &\times \log \frac{p(j_t | j_{t-D_{JI}-D_{IJ}}, i_{t-D_{IJ}}, z_{t-D_{ZI}-D_{IJ}})}{p(j_t | j_{t-D_{JI}-D_{IJ}}, z_{t-D_{ZI}-D_{IJ}})} \end{aligned} \quad (6.13)$$

Equation 6.13 verifies if  $i_{t-D_{IJ}}$  has some information about  $j_t$  in addition to  $z_{t-D_{ZI}-D_{IJ}}$ . Here, again the selection of the time lags is crucial for generating a reliable result. The time lag  $D_{ZI}$  is the time delay between variable  $Z$  and  $I$ .  $D_{IJ}$  is the time delay from variable  $I$  to  $J$  and  $D_{JI}$  is the time delay from variable  $J$  to  $I$ . The reason for considering  $i_{t-D_{IJ}}$  and  $j_{t-D_{JI}-D_{IJ}}$  has been described in Section 6.4.5. By considering  $z_{t-D_{ZI}-D_{IJ}}$  in the equation, we verify if the information in  $i_{t-D_{IJ}}$  is already contained in  $Z$ . For that reason, the term of the history of  $Z$  with the highest influence on  $i_{t-D_{IJ}}$  should be used which is  $z_{t-D_{ZI}-D_{IJ}}$ .

For the example in Figure 6.9, there is no way that variable  $J$  can affect variable  $I$  (the bottom temperature cannot affect the pressure). Therefore, it is possible to remove  $j_{t-D_{JI}-D_{IJ}}$  from the equation without loosing accuracy. The equation to estimate the independent effect of  $I$  on  $J$  is:

$$IE_{I \rightarrow J} = \sum p(j_t, i_{t-6}, z_{t-7}) \log \frac{p(j_t | i_{t-6}, z_{t-7})}{p(j_t | z_{t-7})} \quad (6.14)$$

Equation 6.14 quantifies the highest amount of information transferred from  $I$  to  $J$  excluding the effect of  $Z$ . If the independent effect of  $I$  on  $J$  is zero, then  $I$  cannot be a cause variable to  $J$ . In that case the causality graph is like Figure 6.9. If the independent effect is positive, then variable  $I$  has some unique information about  $J$  which is not found in  $Z$ . The causality graph in this case is as depicted in Figure 6.11. An example of this case is presented in the industrial application section.

An important point to be considered in such cases is that it is not right to remove the connections between variables in Figure 6.11 based on their strength or to simply remove the short cut from  $Z$  to  $J$  for a simplified graph (which is usually done in the literature). By removing the short cut, the causality graph simplifies to  $Z \rightarrow I \rightarrow J$ . This causality graph implies that the variations of the input flow cause the variations

in bottom pressure and then the bottom pressure changes the bottom temperature, which is physically impossible. In cases that two variables have common parent variables, we need to use Equation 6.13 to verify if there is any information transfer between the effect variables. Removing the short cuts with no further analysis is very likely to generate meaningless result. This can be generalized to the cases of more than one parent variable by simply including at least one term of all the common parent variables in Equation 6.13.

Inclusion of intermediate variables in estimation of transfer entropy between two variables has also been considered very recently by an independent study [115], where the authors include the history of the intermediate variables in the transfer entropy estimation to detect direct and indirect causality relations. However, their method of selection of the time lags of the variables to include in the transfer entropy estimation is still based on a rule of thumb with trial and error.

## 6.5 Industrial application

In the following, various possible types of relationships between variables are discussed through examples of an industrial case study. Figure 6.12 shows the schematic of the industrial process. It is an oil sands separation vessel used for bitumen extraction. The input to the process contains bitumen, water and minerals. There are also two water streams to the vessel for the purpose of helping the separation and controlling the process. The three outputs are froth, middlings and tailings.

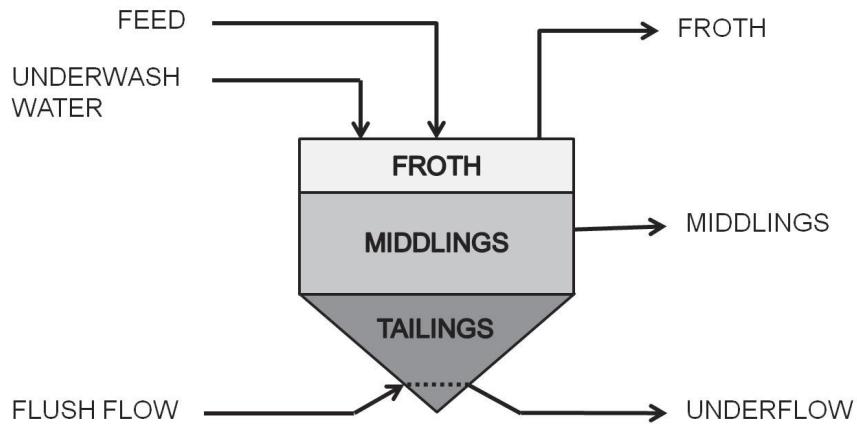


Figure 6.12: Separation vessel schematic

Froth, which is on the top layer, is supposed to only contain bitumen and water. The middling layer contains both bitumen and minerals and is recycled back to the process to separate the remaining bitumen from it. Tailings mostly contain minerals.

The process has about 20 measurements with one minute sampling time and their causal relationships is studied here.

Since the settling time of the process is about one hour, the mutual information is calculated from zero time lags to 60 minutes. By exchanging the order of the variables, the time lagged dependency in both directions is measured. Positive time lags correspond to the case that the first variable is lagged and negative time lags represent the case that the second variable is lagged. 15,000 samples of each variable were used for this analysis and the 5 sigma confidence limit is obtained as 0.035 by generating 1200 pairs of random variables with the same length.

First example is the case of apparent one directional dependency between two variables. This kind of dependency is observed between feed flow to the plant and its density. Figure 6.13 shows the time lagged dependencies between original time series and their differenced version. There is an obvious peak at time lag +3 in both the dependency curves. It implies that density of the feed can be best predicted by feed flow at 3 minutes earlier. This dependency is introduced by the feedback in the upstream plant which provides the feed to the plant under study, where density is being controlled by flow rates.

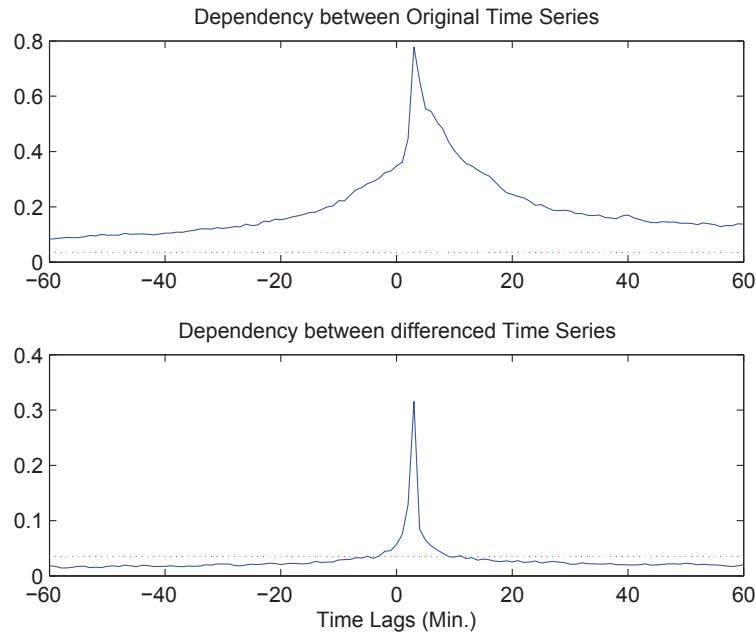


Figure 6.13: Dependency between feed flow and feed density

The second example is the dependency between feed density and underflow density as is depicted in Figure 6.14. Mutual information for negative lags has lower value compared to the positive lags. This means that the underflow density is dependent

on the feed density as expected, but the current value of feed density is not dependent on the history of underflow density. This type of dependency curve, confirms one directional information flow from the first variable (feed density) to the second variable (underflow density).

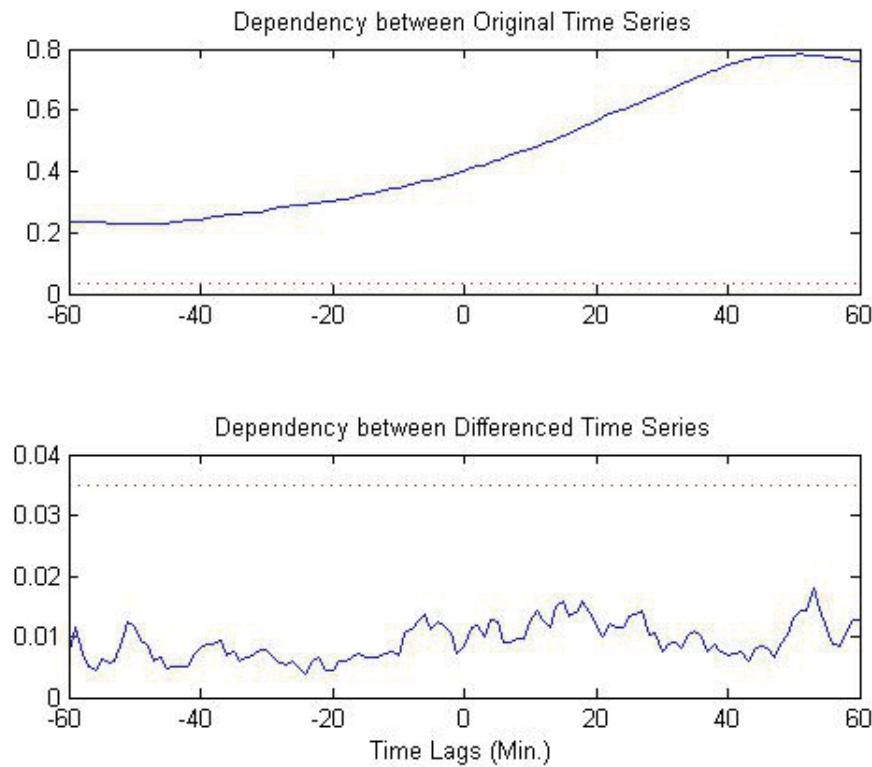


Figure 6.14: Dependency between feed density and underflow density

Differential dependency curve does not show any dependency between the fast variations of these two variables. The reason is that this process is a controlled process with more than half an hour settling time. As a result, the high-frequency variations of the input have been damped out before reaching to the underflow density. Therefore, although values of underflow density are dependent on feed density values, high-frequency variations of the feed density are not directly transferred to the underflow density. We cannot predict the high-frequency variations of the underflow density from the high-frequency variations in the feed density, but we can use the current feed density to predict the underflow density.

A conclusion of the above discussion is that when the value of a variable is highly dependent on another variable, but there is no dependency in their differenced version, then the implication is that the variables are either physically far from each

other or there is a low pass filtering effect between them. The high-frequency variations of the source variable is dumped out in the process before reaching to the subsequent variable. The other possibility is that the measurements are noisy and their differenced version mainly contains noise with no information in it.

Another common type of relationship is the correlation between two variables due to the effect of other variables. An example of this relationship is between middlings density and underflow density. Figure 6.15 shows the dependency between these two variables for -60 to +60 time lags.

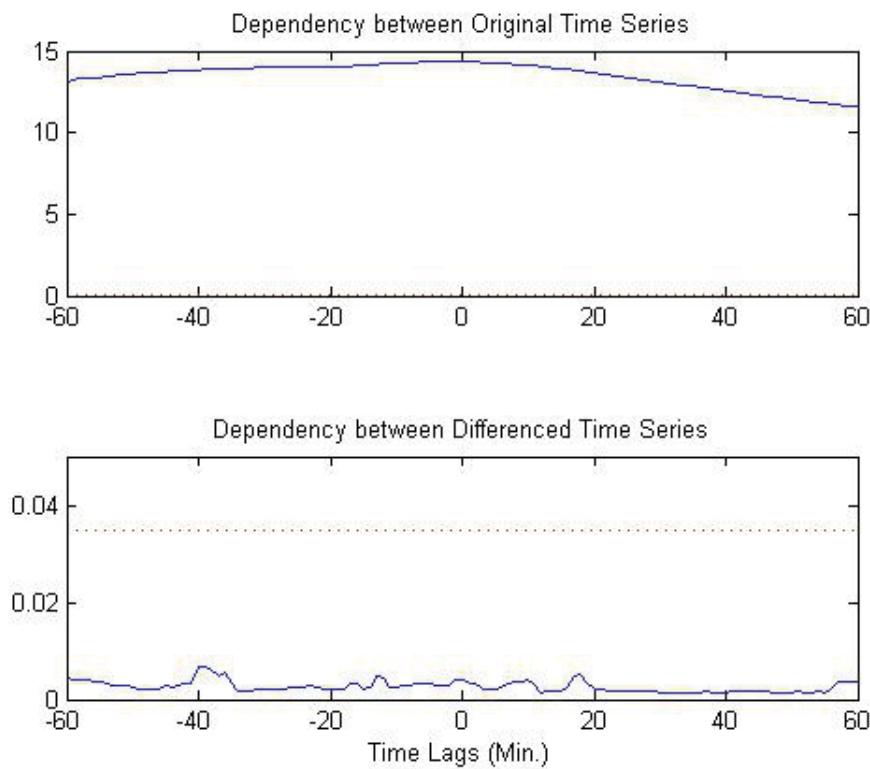


Figure 6.15: Time lagged dependency between underflow density and middlings density

Figure 6.15 shows that there is a high dependency between these two variables at both negative and positive lags. Although there is a decrease in the dependency towards positive lags, it is not substantial and the curve is almost flat and symmetric. This implies that there is no causal relationship between these two variables.

Another example is the dependency between the feed density and underflow pressure. Figure 6.16 shows the dependency between the original time series as well as the differenced versions. The peak at time lag -3 in both curves implies that underflow

pressure influences feed density after 3 minutes. In Figure 6.13 it is seen that the feed flow affects feed density after 3 minutes. Figure 6.17 shows the dependency between feed flow and underflow pressure. Figure 6.17 implies that variations of feed flow is instantly transferred to the underflow pressure. Therefore, underflow pressure is affected by feed flow with zero time delay while feed density is affected after 3 minutes. This is the case of two variables with common parent that was described in Section 6.4.6.

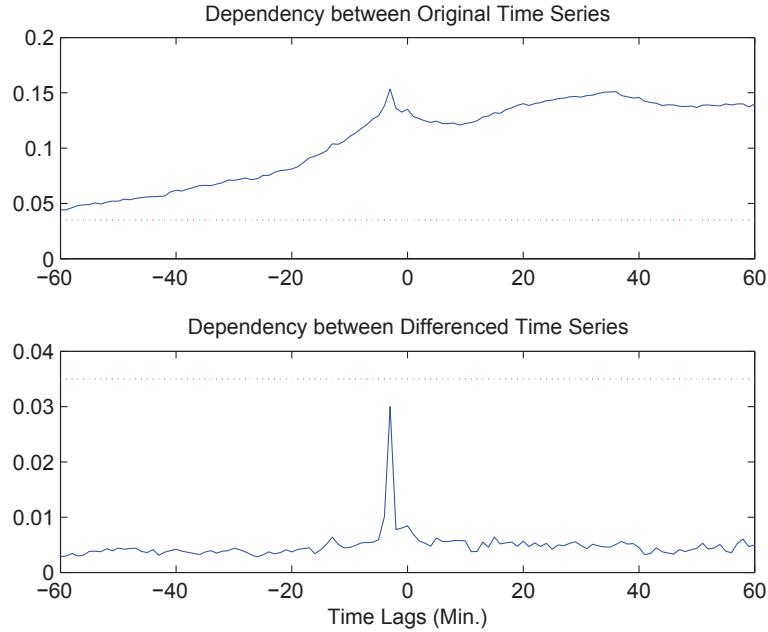


Figure 6.16: Time lagged dependency between feed density and underflow pressure

By using the transfer entropy as in Equation 6.13 it can be verified if any information is transferred from underflow pressure to the feed density when the effect of feed flow is excluded. Independent effect of underflow pressure on feed density considering the current value of feed density and values of feed flow and underflow pressure at 3 minutes earlier is obtained as 0. In the mathematical form:  $P(F.D._n|F.F._{n-3}, U.Press._{n-3}) = P(F.D._n|F.F._{n-3})$  where  $F.D.$  is feed density,  $F.F.$  is feed flow and  $U.Press.$  is underflow pressure. This means that there is actually no information transferred from underflow pressure to the feed density.

Previous discussion shows that when two variables have a common source, it is necessary to consider the effect of the source variable on both of the effect variables to correctly identify the type of the relationship between the two effect variables. In these cases transfer entropy can identify the type of the relationship and dependency curves should be used to determine which terms of the history of the variables should

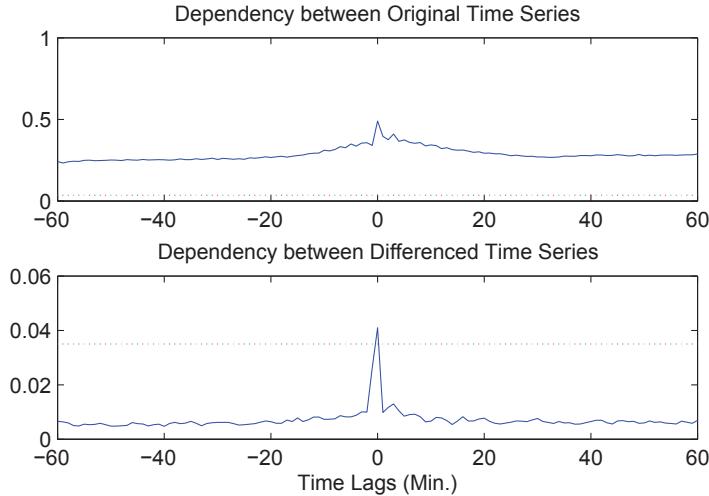


Figure 6.17: Time lagged dependency between feed flow and underflow pressure

be included in the transfer entropy equation for a reliable result.

The other point is that since feed density is highly dependent on the feed flow, we can consider them as one variable. The relationships between feed density with other process variables is similar to the relationships of feed flow with other variables. The difference is simply due to the time delay between feed density and feed flow.

The next example investigates the effect of feedback in the process. As an example, Figure 6.18 shows the dependency between the cone flush flow with underflow density.

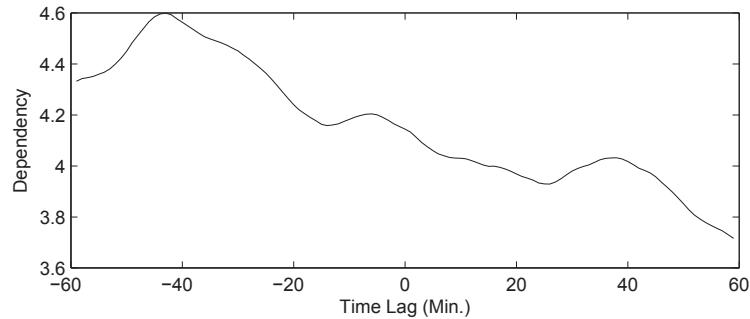


Figure 6.18: Time lagged dependency between cone flush flow and underflow density

Cone flush flow is used to prevent sanding in the vessel and gets a feedback signal from the underflow density. The dependency curve shows a high dependency in negative lags compared to the positive lags, which implies that underflow density is causing cone flush flow. This conclusion is counter intuitive as we know that the operators use cone flush flow to manually control the underflow density.

By estimating the independent effect of the cone flush flow on the underflow density and vice versa it is concluded that the cone flush flow causes the underflow density.

Figure 6.19 shows the result of analysing the important variables of the separation vessel (D. is abbreviation of density). Feed flow and feed density are placed in a box because of the strong dependency between feed density and feed flow. Underflow pressure, Cyclopak pressure and GSW DP tags are highly dependent on each other with zero time delay. Figure 6.20 shows the dependency curves between these three variables. This dependency is not just due to the effect of feed flow which is a parent for all these three tags. Using the transfer entropy shows that there is information flow among these three variables regardless the effect of feed flow. So, they can also be considered as one variable when examining their relationship with the other process variables. These three variables are used to control the separation cell level.

Underflow density and Cyclopak densities are highly correlated with zero time delay and are affected by the feed density, underflow pressure and cone flush flow. All of process variables have correlation with each other and the graph shows the ones that have strong relationships.

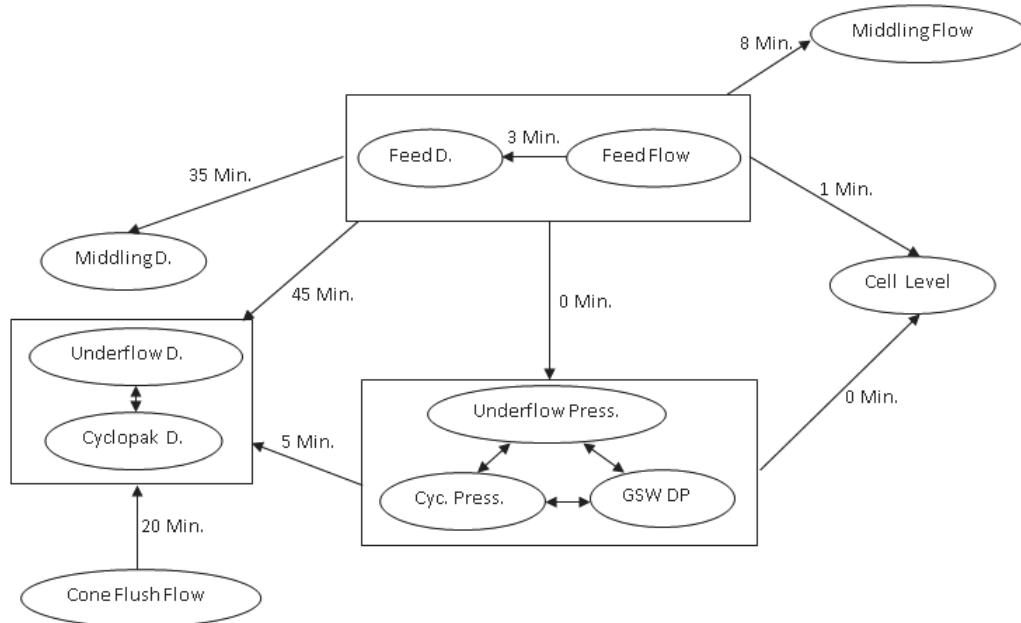


Figure 6.19: Causality relationships in the separation process

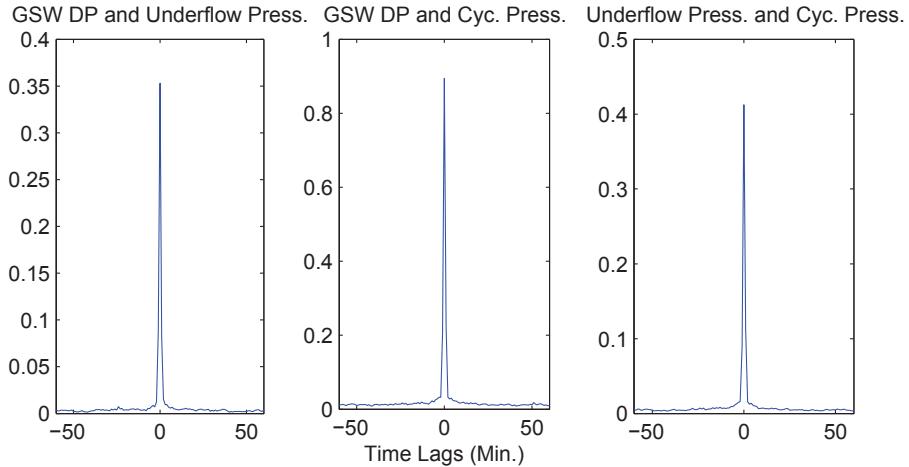


Figure 6.20: Dependency between three tags

### 6.5.1 Summary of the algorithm for causality analysis

The first step in causality analysis is to obtain the time lagged dependency curves between the original time series and their differenced versions. Variables that might have causality relationships can be identified by checking the dependency curves.

Those variables that have small dependency values are omitted from the analysis. Also variables with flat dependency curves without any significant differential dependency value are excluded from the rest of the analysis. There is no cause and effect relationship between these variables even if there is a high correlation between them.

Variables that have differential dependency values higher than the confidence limit should be checked with transfer entropy. Also, the variables that have a non symmetric dependency curve (regardless their differential dependency curve) should be considered for further analysis.

After selection of the variables with probable causality relationship, transfer entropy is used to determine if there is a one directional information flow. Time lagged dependency and differential dependency curves can be used to find the most probable time delay between the measurements. The time lag corresponding to the highest dependency value is taken as the time delay and is used in tuning the parameters of the transfer entropy.

There are two important points to consider at this stage. The first one is the existence of common parent variables. In the transfer entropy parametrization, we need to include at least one term of the common parent variable to ensure the information of the parent variable is not mistaken by the information flow between the variables.

The other point is that the variables that have a strong dependency with an

obvious time delay usually have the same type of relationships with other process variables. It can cause some no meaningful results in causality analysis from the physical point of view. However, this case could also be resolved by using a well tuned transfer entropy.

## 6.6 Summary

Transfer entropy is a promising method for detection of causal relationships between variables if implemented according to Equation 6.5. The problem in implementation is in the computation load of estimating high dimensional probability density as well as a subsequent computation of the transfer entropy.

It was shown that time lagged mutual information and differential mutual information can potentially identify the variables with probable causality relationships. After identification of the variables with important dependencies, transfer entropy is used to determine the direction of information flow. Time lagged dependency curves help in tuning the parameters of the transfer entropy in this stage. Proposed algorithm provides information about the type of the relationships between process variables and an estimation of the time delay between them.

# Chapter 7

## Concluding remarks

## 7.1 Concluding remarks

The direction of this research was toward development of an automatic abnormal event management tool which requires detection, categorization of the fault and finally root cause diagnosis. The contribution of the thesis covers the three stages in general, proposing more advanced methodologies to handle specific issues at each stage. Following is a brief summary of the work presented in each chapter.

Chapter 2 considered the problem of detection and characterization of oscillations in process variables based on routine operational data. Detection of oscillation even though seems to be a trivial task, is a challenge when trying to detect and estimate the frequencies of oscillations in variables which could be hidden within noise and non-stationary trends. Also, many existing procedures require human interference or involve application of complex mathematical tools which is hardly possible in industrial environment. Therefore, first chapter proposed a more reliable methodology to automatically detect and estimate oscillation frequencies in the presence of noise and multiple oscillations. Since the proposed methodology hardly requires human interference it is more appropriate in practice.

The third chapter considered the problem of diagnosis between oscillations due to controller tuning and external oscillatory disturbances. Since oscillations require different troubleshooting procedures based on the element of the loop that has caused the oscillation, it is important to diagnose if the oscillation is due to controller tuning, valve nonlinearity or is a sinusoidal disturbance. The chapter studied the unique characteristics of controller tuning induced oscillations and proposed methodologies to distinguish this type of oscillations from the rest.

The fourth chapter proposes a comprehensive methodology based on wavelet transform for detection and diagnosis of oscillations which is capable of handling non-stationary trends in variables, noise and presence of multiple oscillations. The chapter benefits from the methodologies developed in the previous two chapters while extending them. The proposed methodology is capable of independent diagnosis of different oscillatory components present in variables. It can also distinguish if the oscillation is due to nonlinearity in the process from the other types of the oscillations.

The fifth chapter considers the issue of causality or interaction analysis based on historical data. The chapter first proposes an appropriate model structure for causality analysis using linear methods. Since, instantaneous correlation between variables, which happens in industrial systems due to feedback loops, causes a challenge in causality analysis, a methodology is proposed to detect instantaneous relations between variables from data as a preliminary step of the analysis. The chapter proposes

estimating the model parameters based on Bayesian framework which has several advantages compared to traditional methods. Proper design of prior probability distribution of model parameters helps convergence to a physically interpretable model. Also, it makes it possible to deal with the issue of over-parametrization as well as handling large number of variables. The chapter presents a way to design the prior probability and proposes a method to meaningfully decompose the impulse responses estimated from the model at the end.

In some cases, using linear modeling methodologies is not appropriate for causality analysis due to nonlinear nature of data generation process. In such cases, examining conditional probability distributions and using information transfer concepts such as mutual information and transfer entropy is an alternative. The sixth chapter proposes using transfer entropy in a more reliable way to identify causality relations between variables. Lagged mutual information as well as differential mutual information help in parameterizing the transfer entropy in a way that can correctly identify relations between variables.

## 7.2 Future work

The work presented in this thesis can be continued in several directions. The first three chapters proposed methodologies for detection and categorization of oscillations in variables. The contributions help in selecting a proper root cause diagnosis algorithm to find the loop which is propagating the oscillations to the rest of the process. For root cause diagnosis, causality analysis methods can be utilized. However, due to the nature of oscillation, not all of the causality analysis procedures can be applied. Therefore, the work can be continued in the direction of extending the current causality analysis procedures to be capable of handling oscillations in variables. A promising approach for this purpose is to do causality analysis in the wavelet domain which can potentially have several advantageous.

Regarding the fifth chapter on causality analysis under Bayesian framework, more work is required specially in order to generalize the model structure. Although most of the literature on causality analysis based on linear methods consider a form of vector autoregressive model for the analysis, it is advantageous to generalize the model specifically to be able to handle colored noise in the data. One underlying assumption in using VAR structure is that the noise does not have dynamics and is white which may not be always the case.

Another issue regarding causality analysis methods in general is the usual assumption that all the relevant variables are measured and available. When there are

hidden variables influencing some of the variables under study, the inference about the relations between the measured variables could be falsified. For example, it is known that a correlation between two independent variables can be observed if there is a third variable influencing both of them. The proposed methods in the thesis and most of the literature can work reliably when the third variable is also measured and available. However, when the third variable is hidden, spurious results regarding the relation between the two observed variables can be inferred. Using methodologies such as expectation maximization (EM) for model estimation can potentially facilitate dealing with the challenge of hidden variables.

Regarding the last chapter on nonlinear causality analysis method, still much work needs to be done. One issue is to detect the presence of instantaneous feedback between the variables and how to deal with it based on information transfer concepts. The other one is the fact that these methods can determine the type of the relation between variables but it is not possible to compare the strength of the relations based on the proposed procedure. This has the potential to be a valuable contribution. Also, since nonparametric methods are used to estimate the probability distributions, more work regarding developing faster as well as more accurate methods for estimating the probability densities is valuable.

# Bibliography

- [1] T. J. Harris, C. T. Seppala, L. D. Desborough, A review of performance monitoring and assessment techniques for univariate and multivariate control systems, *Journal of Process Control* 19 (1) (1999) 1–17.
- [2] I. Nimmo, Adequately address abnormal operations, *Chemical Engineering Progress* 91 (9) (1995) 1395–1412.
- [3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. N. Kavuri, A review of process fault detection and diagnosis: Part I: Quantitative model-based methods, *Computers and chemical engineering* 27 (3) (2003) 293–311.
- [4] B. Huang, S. L. Shah, *Performance assessment of control loops: theory and applications*, Springer, 1999.
- [5] Y. Yamashita, An automatic method for detection of valve stiction in process control loops, *Control Engineering Practice* 14 (5) (2006) 503–510.
- [6] M. A. A. Choudhury, S. L. Shah, N. F. Thornhill, D. S. Shook, Automatic detection and quantification of stiction in control valves, *Control Engineering Practice* 14 (12) (2006) 1395–1412.
- [7] M. A. Mohammad, B. Huang, Compensation of control valve stiction through controller tuning, *Journal of Process Control* 22 (2012) 1800–1819.
- [8] F. Qi, B. Huang, Estimation of distribution function for control valve stiction estimation, *Journal of Process Control* 21 (8) (2011) 1208–1216.
- [9] M. Jelali, An overview of control performance assessment technology and industrial applications, *Control Engineering Practice* 14 (5) (2006) 441–466.
- [10] F. Qi, B. Huang, Bayesian methods for control loop diagnosis in the presence of temporal dependent evidences, *Automatica* 47 (7) (2011) 1349–1356.
- [11] B. Huang, Bayesian methods for control loop monitoring and diagnosis, *Journal of Process Control* 18 (9) (2008) 829–838.
- [12] M. Iri, K. Aoki, E. O’Shima, H. Matsuyama, An algorithm for diagnosis of system failures in the chemical process, *Computers and Chemical Engineering* 3 (1) (1979) 489–493.
- [13] C. T. Chang, H. C. Huang, New developments of the digraph-based techniques for fault-tree synthesis, *Industrial and engineering chemistry research* 31 (6) (1992) 1490–1502.

- [14] R. E. Neapolitan, Learning bayesian networks, Upper Saddle River: Pearson Prentice Hall, 2004.
- [15] M. R. Maurya, R. Rengaswamy, V. Venkatasubramanian, Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets, *Engineering Applications of Artificial Intelligence* 17 (5) (2004) 501–518.
- [16] N. H. Ulerich, G. J. Powers, On-line hazard aversion and fault diagnosis in chemical processes: the digraph, fault-tree method, *IEEE Transactions on Reliability* 37 (2) (1988) 171–177.
- [17] F. Yang, S. L. Shah, D. Xiao, Signed directed graph based modeling and its validation from process knowledge and process data, *International Journal of Applied Mathematics and Computer Science* 22 (1) (2012) 41–53.
- [18] C. W. J. Granger, Investigating causal relationships by econometric models and cross-spectral methods, *Econometrica* 37 (3) (1969) 424–438.
- [19] B. Huang, S. X. Ding, N. Thornhill, Alternative solutions to multivariate control performance assessment problems, *Journal of Process Control* 16 (5) (2006) 457–471.
- [20] B. Huang, S. X. Ding, N. Thornhill, Practical solutions to multivariate feedback control performance assessment problem: reduced a priori knowledge of interactor matrices, *Journal of Process Control* 15 (5) (2005) 573–583.
- [21] M. Jelali, B. Huang, Detection and diagnosis of stiction in control loops: state of the art and advanced methods, Springer, London, 2010.
- [22] Hagglund, A control-loop performance monitor, *Control Engineering Practice* 3 (1995) 1543–1551.
- [23] N. Thornhill, B. Huang, H. Zhang, Detection of multiple oscillations in control loops, *Journal of Process Control* 13 (1) (2003) 91–100.
- [24] H. Jiang, M. S. Choudhury, S. L. Shah, Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method, *Journal of Process Control* 17 (2) (2007) 143–155.
- [25] R. Srinivasan, R. Rengaswamy, R. Miller, A modified empirical mode decomposition (EMD) process for oscillation characterization in control loops, *Control Engineering Practice* 15 (9) (2007) 1135–1148.
- [26] X. Li, J. Wang, B. Huang, S. Lu, The DCT-based oscillation detection method for a single time series, *Journal of Process Control* 20 (2010) 609–617.
- [27] C. Xia, J. Howell, N. F. Thornhill, Detecting and isolating multiple plant-wide oscillations via spectral independent component analysis, *Automatica* 41 (12) (2005) 2067–2075.
- [28] A. K. Tangirala, J. Kanodia, S. L. Shah, Non-negative matrix factorization for detection and diagnosis of plantwide oscillations, *Industrial and engineering chemistry research* 46 (3) (2007) 801–817.

- [29] M. R. Allen, A. W. Robertson, Distinguishing modulated oscillations from coloured noise in multivariate datasets, *Climate Dynamics* 12 (11) (1996) 775–784.
- [30] J. Li, P. Stoica, Efficient mixed-spectrum estimation with applications to target feature extraction, *IEEE Transactions on Signal Processing* 44 (2) (1996) 281–295.
- [31] M. A. A. Choudhury, S. L. Shah, N. F. Thornhill, Diagnosis of process nonlinearities and valve stiction: data driven approaches, Springer, 2008.
- [32] N. F. Thornhill, J. W. Cox, M. A. Paulonis, Diagnosis of plant-wide oscillation through data-driven analysis and process understanding, *Control Engineering Practice* 11 (12) (2003) 1481–1490.
- [33] S. Karra, M. N. Karim, Comprehensive methodology for detection and diagnosis of oscillatory control loops, *Control Engineering Practice* 17 (8) (2009) 939 – 956.
- [34] A. Horch, Condition monitoring of control loops, Ph.D. thesis, KTH, 2000.
- [35] S. Babji, U. Nallasivam, R. Rengaswamy, Root Cause Analysis of Linear Closed-Loop Oscillatory Chemical Process Systems, *Industrial and Engineering Chemistry Research* 51 (42) (2012) 13712–13731.
- [36] E. Radoi, A. Quinquis, A new method for estimating the number of harmonic components in noise with application in high resolution radar, *EURASIP Journal on Applied Signal Processing* (2004) 1177–1188.
- [37] K. Ogata, System dynamics, Pearson/Prentice Hall, 2004.
- [38] E. Naghooisi, B. Huang, Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations, *IECR* 53 (22) (2014) 9427–9438.
- [39] D. Seborg, T. Edgar, D. Mellichamp, Process dynamicss and control, John Wiley and Sons, 1989.
- [40] W. A. Fuller, Introduction to statistical time series, Wiley, 2009.
- [41] G. E. Box, G. M. Jenkins, G. C. Reinsel, Time series analysis: forecasting and control, John Wiley and Sons, 2013.
- [42] M. S. Bartlett, On the theoretical specification and sampling properties of autocorrelated time-series, *Supplement to the Journal of the Royal Statistical Society* 8 (1) (1946) 27–41.
- [43] C. Torrence, G. P. Compo, A practical guide to wavelet analysis, *Bulletin of the American Meteorological society* 79 (1) (1998) 61–78.
- [44] Z. K. Peng, F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, *Mechanical Systems and Signal Processing* 18 (2) (2004) 199–221.
- [45] Y. Yang, X. Li, X. Liu, X. Chen, Wavelet kernel entropy component analysis with application to industrial process monitoring, *Neurocomputing* 147 (5) (2015) 395–402.

- [46] S. Alabi, A. Morris, E. Martin, On-Line Dynamic Process Monitoring Using Wavelet-Based Generic Dissimilarity Measure, *Chemical Engineering Research and Design* 83 (6) (2005) 698–705.
- [47] K. Zhu, Y. S. Wong, G. S. Hong, Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results, *International Journal of Machine Tools and Manufacture* 49 (7) (2009) 537–553.
- [48] B. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *American Institute of Chemical Engineers Journal* 44 (7) (1998) 1596–1610.
- [49] J. Lin, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, *Journal of Sound and Vibration* 234 (1) (2000) 135–148.
- [50] T. Matsuo, H. Sasaoka, Y. Yamashita, Detection and Diagnosis of Oscillations in Process Plants, *Lecture Notes in Computer Science* (2003) 1258–1264.
- [51] Z. Guo, L. Xie, A. Horch, Y. Wang, H. Su, X. Wang, Automatic detection of nonstationary multiple oscillations by an improved wavelet packet transform, *IECR* 53 (40) (2014) 15686–15697.
- [52] M. I. Plett, Transient detection with cross wavelet transforms and wavelet coherence, *IEEE Transactions on Signal Processing* 55 (5) (2007) 1605–1611.
- [53] E. Naghooisi, B. Huang, Diagnosis between sinusoidal oscillations and oscillatory colored noise, in: *Proceedings of American Control conference*, 2014.
- [54] E. Naghooisi, B. Huang, Diagnosis of Oscillations Between Controller Tuning and Harmonic External Disturbances, *IEEE Transactions on Control Systems Technology* 23 (4) (2014) 1285–1293.
- [55] A. Grossmann, J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM J. of Math. Anal.* 15 (4) (1984) 723–736.
- [56] J. Q. Zhang, Y. Yan, A wavelet-based approach to abrupt fault detection and diagnosis of sensors, *Instrumentation and Measurement, IEEE Transactions on* 50 (5) (2001) 1389–1396.
- [57] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1999.
- [58] D. L. Donoho, I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *Journal of the american statistical association* 90 (432) (1995) 1200–1224.
- [59] D. L. Donoho, I. M. Johnstone, Threshold selection for wavelet shrinkage of noisy data, *Proceedings of the 16th Annual International Conference of the IEEE In Engineering in Medicine and Biology Society* (1994) A24–A25.
- [60] S. G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, *Image Processing, IEEE Transactions on* 9 (9) (2000) 1532–1546.

- [61] F. Luisier, T. Blu, M. Unser, A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding, *Image Processing, IEEE Transactions on* 16 (3) (2007) 593–606.
- [62] Y. Xu, J. B. Weaver, D. M. Healy, J. Lu, Wavelet transform domain filters: a spatially selective noise filtration technique, *Image Processing, IEEE Transactions on* 3 (6) (1994) 747–758.
- [63] T. Alexandrov, S. Bianconcini, E. B. Dagum, P. Maass, T. S. McElroy, A review of some modern approaches to the problem of trend extraction, *Econometric Reviews* 31 (6) (2012) 593–624.
- [64] D. M. T. Karlsruhe, A Novel Wavelet Based Approach for Time Series Data Analysis, thesis: Institut für Technologie, K. I. T. .
- [65] Z. Zhang, J. C. Moore, Comment on” Significance tests for the wavelet power and the wavelet power spectrum by Ge (2007), *Annales Geophysicae-Atmospheres Hydrospheres and Space Sciences* 30 (12).
- [66] G. E. Box, J. S. Hunter, W. G. Hunter, *Statistics for experimenters: design, innovation, and discovery*, Wiley-Interscience, New York, 2 edn., 2005.
- [67] W. Gardner, A unifying view of coherence in signal processing, *Signal Processing* 29 (2) (1992) 113–140.
- [68] M. Rosenblatt, J. W. V. Ness, Estimation of the bispectrum, *Annals of Mathematical Statistics* (420-436).
- [69] C. L. Nikias, A. Petropulu, *Higher-order spectra: A nonlinear signal processing frame work*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [70] M. J. Hinich, Testing for Gaussianity and linearity of a stationary time series, *Journal of Time Series Analysis* 3 (1982) 169–176.
- [71] M. S. Choudhury, S. L. Shah, N. F. Thornhill, Diagnosis of poor control-loop performance using higher-order statistics, *Automatica* 40 (10) (2004) 1719–1728.
- [72] B. P. V. Milligen, C. Hidalgo, E. Sanchez, Nonlinear phenomena and intermittency in plasma turbulence, *Physical review letters* 74 (3).
- [73] B. P. V. Milligen, E. Sanchez, T. Estrada, C. Hidalgo, B. Branas, B. Carreras, L. Garcia, Wavelet bicoherence: A new turbulence analysis tool, *Physics of Plasmas* 2 (8) (1995) 3017–3032.
- [74] Z. Ge, Significance testing for wavelet bicoherence and its application in analyzing nonlinearity in turbulent shear flows, *Physical Review* 81 (5).
- [75] E. Bristol, On a new measure of interaction for multivariable process control, *IEEE Transactions on Automatic Control* 11 (1) (1966) 133–134.
- [76] M. van de Wal, B. de Jager, A review of methods for input/output selection, *Automatica* 37 (2001) 487–510.

- [77] B. Halvarsson, Comparison of some gramian based interaction measures, in: Proceedings of IEEE International Conference on Computer-Aided Control Systems, 2008.
- [78] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, H. Witte, Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems, *Signal Processing* 85 (11) (2005) 2137–2160.
- [79] B. Gourevitch, R. L. Bouquin-Jeannes, G. Faucon, Linear and nonlinear causality between signals: methods, examples and neurophysiological applications, *Biological Cybernetics* 95 (4) (2006) 349–369.
- [80] F. Yang, P. Duan, S. Shah, T. Chen, Capturing connectivity and causality in complex industrial processes, Springer Science and Business Media, 2014.
- [81] J. Geweke, Measurement of linear dependence and feedback between multiple time series, *Journal of the American Statistical Association* 77 (378) (1982) 304–313.
- [82] K. J. Blinowska, R. Ku, M. Kamiski, Granger causality and information flow in multivariate processes, *Physical Review E* 70 (5) (2004) 050902.
- [83] M. Kaminski, K. Blinowska, A new method of the description of the information flow in the brain structures, *Biological Cybernetics* 65 (1991) 203–210.
- [84] L. Baccala, K. Sameshima, Overcoming the limitations of correlation analysis for many simultaneously processed neural structures, *Prog Brain Res* 130 (2001a) 34–47.
- [85] H. Lutkepohl, New introduction to multiple time series analysis, Springer, New York, 2005.
- [86] J. M. Dufour, E. Renault, Short run and long run causality in time series: theory, *Econometrica* (1998) 1099–1125.
- [87] S. Gigi, A. K. Tangirala, Quantitative analysis of directional strengths in jointly stationary linear multivariate processes, *Biological cybernetics* 103 (2) (2010) 119–133.
- [88] C. D. Mol, D. Giannone, L. Reichlin, Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?, *Econometrics* 146 (2008) 318–328.
- [89] M. Banbura, D. Giannone, L. Reichlin, Large Bayesian vector auto regressions, *Applied Econometrics* 25 (1072) (2010) 71–92.
- [90] R. Giacomini, H. White, Tests of conditional predictive ability, *Econometrica* 74 (6) (2006) 1545–1578.
- [91] A. Zellner, An introduction to Bayesian inference in econometrics, Willey, Toronto, 1996.
- [92] C. A. Sims, T. Zha, Bayesian methods for dynamic multivariate models, *International Economic Review* (1998) 949–968.

- [93] D. F. Waggoner, T. Zha, A Gibbs sampler for structural vector autoregressions, *Journal of Economic Dynamics and Control* 28 (2) (2003) 349–366.
- [94] L. Tierney, Markov chains for exploring posterior distributions, *The Annals of Statistics* (1994) 1701–1728.
- [95] D. F. Waggoner, T. A. Zha, Normalization, probability distribution, and impulse responses, *Federal Reserve Bank of Atlanta Working Paper* 97 (11).
- [96] A. Inoue, L. Kilian, Inference on impulse response functions in structural VAR models, *Journal of Econometrics* 177 (1) (2013) 1–13.
- [97] G. Koop, Aggregate shocks and macroeconomic fluctuations: A Bayesian approach, *Applied Econometrics* 7 (4) (1992) 395–411.
- [98] C. A. Sims, T. Zha, Error bands for impulse responses, *Econometrica* 67 (5) (1999) 1113–1155.
- [99] H. Lutkepohl, A. Staszewska-Bystrova, P. Winker, Comparison of methods for constructing joint confidence bands for impulse response functions, *International Journal of Forecasting* 31 (3) (2015) 782–798.
- [100] S. Gigi, A. K. Tangirala, Quantification of interaction in multiloop control systems using directed spectral decomposition, *Automatica* 49 (5) (2013) 1174–1183.
- [101] H. Jiang, R. Patwardhan, S. L. Shah, Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix, *Journal of Process Control* 19 (2009) 1347–1354.
- [102] M. R. Maurya, R. Rengaswamy, V. Venkatasubramanian, A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis, *Ind. Eng. Chem. Res.* 42 (2003) 4789–4810.
- [103] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, N. F. Thornhill, Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy, *IEEE Transactions on Control Systems Technology* 15 (1) (2007) 12–21.
- [104] N. Thornhill, S. Shah, B. Huang, Detection of Distributed Oscillations and Root-Cause Diagnosis, *Proceedings of Chemfas* 4 (2001) 167–172.
- [105] M. Stockmanna, R. Habera, U. Schmitzb, Source identification of plant-wide faults based on k nearest neighbor time delay estimation, *Journal of Process Control* 22 (33) (2012) 583–598.
- [106] T. Schreiber, Measuring Information Transfer, *Physical Review Letters* 85 (2) (2000) 461–464.
- [107] S. Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.
- [108] C. E. Shannon, W. Weaver, R. Blahut, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949.
- [109] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

- [110] A. Kaiser, T. Schreiber, Information transfer in continuous processes, *Physica* 166 (1) (2002) 43–62.
- [111] M. P. Wand, Fast Computation of Multivariate Kernel Estimators, *Journal of Computational and Graphical Statistics* 3 (4) (1994) 433–445.
- [112] M. C. Jones, J. S. Marron, S. J. Sheather, A Brief Survey of Bandwidth Selection for Density Estimation, *Journal of the American Statistical Association* 91 (433) (1996) 401–407.
- [113] R. Marschinski, H. Kantz, Analysing the information flow between financial time series, *Eur. Phys. J.* 30 (2) (2002) 275–281.
- [114] M. Palus, M. Vejmelka, Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections, *Physical Review E* 75 (5).
- [115] P. Duan, F. Yang, T. Chen, S. L. Shah, Direct Causality Detection via the Transfer Entropy Approach, *IEEE Transactions on Control Systems Technology* 21 (6) (2013) 2052–2066.

## .1 Appendix

### .1.1 Examining correlation between residuals of a multivariate regression model estimated based on Bayesian framework

The method to detect instantaneous relations between variables is based on fitting a multivariate regression model to the variables and examining the covariance matrix of residuals. A multivariate regression model is of the following form:

$$Y = XB + U \quad (1)$$

where  $Y = [y_1, y_2, \dots, y_m]$  is an  $n \times m$  matrix of observations on  $m$  variables,  $X$  is  $n \times k$  matrix of independent variables with rank  $k$ ,  $B = [\beta_1, \beta_2, \dots, \beta_m]$  is  $k \times m$  matrix of regression parameters and  $U = [u_1, \dots, u_m]$  is  $n \times m$  matrix of disturbances. The rows of  $U$  matrix are independently distributed with positive definite  $m \times m$  covariance matrix. The pdf of  $Y$  can be written in the following form.

$$p(Y|X, B, \Sigma) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}(Y - XB)'(Y - XB)\Sigma^{-1}\right] \quad (2)$$

Considering that

$$\begin{aligned} (Y - XB)'(Y - XB) &= (Y - X\hat{B})'(Y - X\hat{B}) + (B - \hat{B})'X'X(B - \hat{B}) = \\ &S + (B - \hat{B})'X'X(B - \hat{B}) \end{aligned} \quad (3)$$

where  $\hat{B} = (X'X)^{-1}X'Y$  and  $S = (Y - X\hat{B})'(Y - X\hat{B})$ . The likelihood function can be written as:

$$l(B, \Sigma|Y, X) \propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{tr}S\Sigma^{-1} - \frac{1}{2} \text{tr}(B - \hat{B})'X'X(B - \hat{B})\Sigma^{-1}\right) \quad (4)$$

The prior distribution is considered to be non-informative as

$$p(B, \Sigma) = p(B)p(\Sigma) \propto |\Sigma|^{-(m+1)/2} \quad (5)$$

The joint posterior of the parameters can be obtained as:

$$p(B, \Sigma|Y, X) \propto |\Sigma|^{-1/2(n+m+1)} \exp\left(-\frac{1}{2} \text{tr}(S + (B - \hat{B})'X'X(B - \hat{B}))\Sigma^{-1}\right) \quad (6)$$

which can be written as  $p(B, \Sigma|Y, X) = p(B|\Sigma, Y, X)p(\Sigma|Y, X)$ .

$$p(B|\Sigma, Y, X) \propto |\Sigma|^{-k/2} \exp\left(-\frac{1}{2} \text{tr}((B - \hat{B})' X' X (B - \hat{B}) \Sigma^{-1})\right) \quad (7)$$

The conditional posterior of  $B$  given  $\Sigma$  is multivariate normal with mean  $\hat{\beta}$  and covariance matrix  $\Sigma \otimes (X' X)^{-1} = (\Sigma^{-1} \otimes X' X)^{-1}$  ( $|\Sigma^{-1} \otimes X' X|^{1/2} \propto |\Sigma|^{-k/2}$ ).

The conditional posterior for one specific equation, for example,  $\beta_1$  can also be obtained as

$$p(\beta_1|\Sigma, Y, X) \propto \frac{1}{\sigma_{11}^{k/2}} \exp\left(-\frac{1}{2\sigma_{11}} (\beta_1 - \hat{\beta}_1)' X' X (\beta_1 - \hat{\beta}_1)\right) \quad (8)$$

where  $\beta' = [\beta'_1, \beta'_2, \dots, \beta'_m]$  and has covariance matrix equal to  $(X' X)^{-1} \sigma_{11}$ .

$$p(\Sigma|Y, X) \propto |\Sigma|^{-\nu/2} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} S\right) \quad (9)$$

where  $\nu = n - k + m + 1$ . The form of the pdf is called inverted Wishart. It is also shown that partitioning the  $\Sigma$  matrix to two parts  $\Sigma_{11}$  and  $\Sigma_{22}$  where  $\Sigma_{11}$  is the  $p \times p$  upper left-hand principal minor matrix of  $\Sigma$  with  $p < m$ , leads to the following posterior distribution.

$$p(\Sigma_{11}|Y, X) \propto |\Sigma_{11}|^{-(\nu-2(m-p))/2} \exp\left(-\frac{1}{2} \text{tr} \Sigma_{11}^{-1} S_{11}\right) \quad (10)$$

where  $S_{11}$  is the upper left-hand principal minor of  $S$ .

If  $p = 2$  the posterior for  $\rho_{12} = \frac{\sigma_{12}}{(\sigma_{11}\sigma_{22})^{1/2}}$  can be obtained as

$$p(\rho_{12}|Y, X) \propto \frac{(1 - \rho_{12}^2)^{(n'-3)/2}}{(1 - \rho_{12}r_{12})^{n'-1/2}} S_{n'}(\rho_{12}r_{12}) \quad (11)$$

where  $n' = n - k - (m - 2)$ ,  $r_{12} = s_{12}/(s_{11}s_{22})^{1/2}$  and

$$S_{n'}(\rho_{12}r_{12}) = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \frac{1^2 \cdot 3^2 \dots (2l-1)^2}{(n' + \frac{1}{2}) \dots (n' + l - \frac{1}{2})} \left(\frac{1 + \rho_{12}r_{12}}{8}\right)^l \quad (12)$$

This posterior can be used in order to check if there is correlation between the residuals. If 0 is not within the posterior of  $\rho_{12}$  it implies there is instantaneous relations between the two respective variables.