ISBN 978-82-326-3164-3 (printed version)
ISBN 978-82-326-3165-0 (electronic version)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Doctoral Thesis
Muhammad Faisal Aftab

Doctoral theses at NTNU, 2018:187

Muhammad Faisal Aftab

Controller Performance Monitoring

Detection and Diagnosis of Oscillations in Control Loops

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Engineering Cybernetics

Doctoral theses at NTNU, 2018:187
NTNU

NTNU
Norwegian University of
Science and Technology

Muhammad Faisal Aftab

Controller Performance Monitoring

Detection and Diagnosis of Oscillations in Control Loops

Thesis for the degree of Philosophiae Doctor

Trondheim, May 2018

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Engineering Cybernetics

ISBN 978-82-326-3164-3 (printed version
ISBN 978-82-326-3165-0 (electronic version)
ISSN 1503-8181

ITK-report: 2018-5-W

Doctoral theses at NTNU, 2018:187

Printed by Skipnes Kommunikasjon as

Summary

The profitability of any industrial process is closely related to its ability to maintain near optimal operating conditions; therefore robust methods for root cause detection of any control performance degradation is crucial not only to maintain the desired operation of any industrial process but can result in marked improvement in productivity and over all economics.

In a large and complex industrial control system, disturbances/oscillations originating at one point tend to propagate both up- and downstream due to underlying interactions, process flows and recycle streams, thus leading to plant-wide effects. Effective controller performance monitoring therefore requires detection and diagnosis of such plant-wide disturbances so that targeted maintenance can be achieved within the shortest possible time.

Oscillations are considered to be the most important indicator of control performance degradation. Oscillations not only cause product variability and effect quality but also lead to adverse economical consequences owing to loss of precious resources like raw material and energy. Therefore, automated detection and diagnosis of oscillations needs to be reliable and effective as the large scale of the control system in industrial process plants makes manual observation and diagnosis practically impossible.

Detection and diagnosis of oscillatory control loops is not trivial due to the causes such as presence of multiple oscillations, unknown process dynamics, non stationary effects and noise corruption, to name a few. The work presented in this thesis is aimed at addressing these challenges and at the development of improved tools for both detection and diagnosis of oscillations and plant-wide disturbances. Moreover, data driven methods are preferred over model based diagnostics as they are more general and are independent of a specific plant model. Dynamic pro-

cess models can be costly to develop and maintain, and for many process plants dynamic models are not available.

This thesis has focussed on using the multivariate empirical mode decomposition (MEMD) and associated characteristics for the detection and diagnosis of oscillations in control loops. The methods proposed in thesis are aimed at addressing the shortcomings of the existing approaches and are fully data driven that require no *a priori* knowledge about the data or process itself. An improved approach for the oscillation detection has been presented that caters for non-stationary effects and reduces the mode mixing problems associated with the univariate empirical mode decomposition. Moreover, MEMD along with the proposed grouping algorithm provides a robust method to detect the plant-wide oscillations where different control variables, having common signatures of oscillations, are oscillating due to the same root cause. This helps in searching for the root cause only in the affected variables.

Another important contribution of the thesis is an automated detection of harmonics in an oscillating signal. Non-linearity induced oscillations give rise to limit cycles characterised by harmonics. The proposed harmonic detection algorithm is the only algorithm to date which fully automates the harmonics detection. Moreover, an index called degree of non-linearity (DNL) based on intra-wave frequency modulation has been used to quantify the extent of non-linearity to isolate the source of non-linearity induced oscillation. Moreover, the MEMD has been combined with the concept of delayed vector variance (DVV) to differentiate between linear and non-linear causes of oscillations. This has extended the applicability of DVV to the non-stationary signals as well. The delay vector variance method is also integrated with automatic calculation of embedding dimension to make it more robust and easy to automate.

Finally, plant-wide disturbances tend to propagate away from source owing to process flows and interactions among the control loops. Therefore, the cause and effect relationship can be exploited to isolate the source of disturbance. The causality analysis based on convergent cross mapping (CCM) approach is also explored for isolating the source of plant-wide oscillations. The CCM works well for the short span time series.

The research has been funded by the Siemens AS, Norway as a part of the Siemens-NTNU oil and gas offshore project.

Contents

List of Tables

List of Figures

Acknowledgements

First of all I am grateful to the the Creator and Sustainer of this universe, the God Almighty (Allah (SWT)), the Most Merciful and the Most Compassionate, for enabling me to accomplish this uphill task.

There are many people without whose help this thesis would not have been possible. First of all I am thankful to my supervisor Professor Morten Hovd for providing me opportunity to work on this project and also to Siemens AS, Norway for funding this research. Professor Morten has provided continuous assistance, paper reviews and fruitful suggestions during the course of PhD project. I would also like to mention my co-supervisor Dr. Selvanathan Sivalingam (Key expert, Siemens) for his cooperation and guidance during my duty work at Siemens. He also managed to get hold of the industrial data used in this research.

Moreover, I am grateful to the administrative staff at the Engineering Cybernetics department for their support and for maintaining a conducive environment. I would specially thank Leif Anderson, with whom I shared my office. He assisted me at numerous occasions and helped with lot of Matlab and Latex issues, during the past three years.

Finally I am much grateful to my parents for their prayers and my wife Halima, kids Talha, Abdullah and little Ismail for their support, patience and understanding of my busy schedule.

Muhammad Faisal Aftab
Trondheim, May 2018.

Chapter 1

Introduction

The increased level of automation in industry since the middle of the 20th century has resulted in increased standards of productivity, profitability, quality and safety. These objectives can be compromised to a significant extent due to any variation from the desired operating conditions. In a large industrial plant, the number of control loops can be in range of thousands and this level of automation necessitates a robust monitoring mechanism that can oversee the control system. Performance monitoring of control systems has attracted attention of the researchers in recent times for its importance in modern industrial applications.

Controller performance monitoring, if implemented, in principle can increase the reliability, efficiency and profitability of any industrial process in question regardless of its type, size or complexity. The main essence of controller performance monitoring, as the name suggests, is to diagnose and isolate all of control related problems which can significantly degrade the operation of an industrial process. The performance statistics of 26000 PID control loops operation spanning a period of two years are summarized in Figure 1.1 (Desborough and Miller 2002). It shows that only the third of the control loops in the process industry are considered to be performing well while rest suffer from one problem or the other. Moreover, around the same number, i..e., one third is reported to be oscillating (Srinivasan et al. 2007). The results from the recent survey by Starr et al. (2016) report that approximately 75% of control loops are actually not benefiting the plant operations. Thus, even with the passage of fifteen years the state of the affairs has not improved much.

Under these circumstances a robust controller performance monitoring mechanism is therefore strongly desirable in order to diagnose any cause of control related

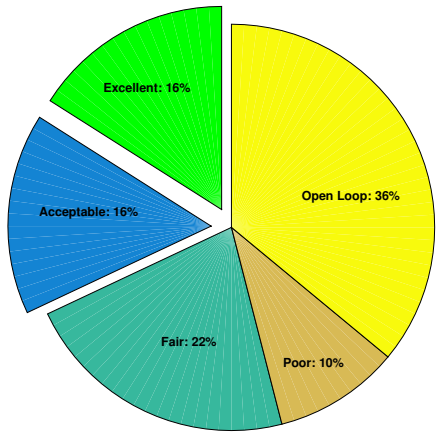

Figure 1.1: Industrial control loops performance statistics.

problem in the process. Efforts are being made to address this problem and different software packages such as Loop Performance Manager by ABB, ITUNE+ by ControlSoft, Loop Tuning Tune VP from Yokogawa, Plant Checkup by University of Pisa and Control Performance Monitor by HoneyWell have been made available commercially to aid in the controller performance monitoring (Bacci di Capaci and Scali 2018).

Most research from the controller performance monitoring perspective, of late, is being carried out on issues related to the chemical process industries probably due to the fact that the sheer size and complexity of a chemical plant calls for more automated monitoring measures as compared to the smaller systems. There can be many indicators of poor control loop performance such as *a*) poor reference tracking and disturbance rejection, *b*) oscillations, and *c*) excessive control action, to name a few. A common approach is to treat oscillations in control loops as an indicator of poor performance although the absence of oscillations does not necessarily mean that the control loop is performing well as it may perform poorly due to other reasons described above. Nevertheless, the presence of oscillations can be considered as first indication of the performance degradation that, if allowed to carry on unchecked, can significantly compromise performance and safety of any industrial process in general and chemical process plants in particular.

1.1 Root Causes of Poor Control Performance

The poor control performance can arise due to variety of issues, some of the issues are described briefly here (Choudhury et al. 2008).

Inappropriate Controller Tuning

The controllers are operating with default gains even when the operating conditions and process dynamics are changed considerably over time. The retuning is seldom done and the default settings can no longer give the acceptable performance.

Equipment Failure

Equipment failure, like that of sensor or an actuator can also be a major reason of poor control performance. The reduced manpower and increased level of automation results in lesser maintenance checks and hence increased probability of failure.

Poor Process Design

Control performance is greatly influenced by the appropriate process design. Sometimes the poor system design makes it impossible to achieve good performance.

Non-Linearities

Process non-linearities are major cause of poor control performance. The linear controller design methodology, widely employed in industry, assumes locally linear dynamics of underlying process. The linearity assumption may be violated due to variety of reasons like change in operating conditions, process degradation, inappropriate set points or non-linearity in actuators and sensors.

Inappropriate Control Structure

The unrealistic control structure like poor selection of control variables; inadequate input/output pairing and absence of derivative action can also contribute significantly to performance degradation. The increased interaction among different underlying loops may lead to plant-wide disturbance propagation and create performance issues.

1.2 Contributions

This thesis is aimed at exploring robust data driven methods to improve the existing performance monitoring paradigm. Considering that the oscillations are most common indicator of the performance degradation and product variability, this research has also focussed on the detection and diagnosis of oscillations in control

loops.

Karra et al. (2010) has outlined the presence of slow varying non-stationary trends and presence of multiple oscillations as a challenge in addressing the detection and diagnosis of oscillations. Likewise, though the presence of harmonics in non-linearity induced oscillations is well know, yet any automatic method to identify the harmonics, to detect non-linearity, has not been reported. The main contribution of the thesis is the development of methods that are fully data driven, automated and applicable to non-stationary time series data. The application of multivariate empirical mode decomposition (MEMD) and the associated properties, for the detection, diagnosis and characterization of the plant-wide oscillations has been studied in this work. Moreover, an automatic method to identify the presence of harmonics has also been developed. Though MEMD has been used in wide range of fields yet it has not been exploited for the control performance monitoring and diagnosis. Only the basic and univariate (time series of one variable) empirical mode decomposition (EMD) has been used by Srinivasan et al. (2007), Babji et al. (2009) and Srinivasan and Rengaswamy (2012) in the past.

No dependence on any type of *pre*-filtering or *pre*-processing of the data are the main advantages offered by the MEMD based methods. Moreover, it does not suffer from problems like mode mixing (will be explained later) that are associated with its predecessor univariate version. The key areas addressed in the thesis and the improvements offered therein are:

- An improved oscillation detection method is presented. The method is aimed at addressing the limitations of the existing EMD based method. The proposed method can detect multiple oscillations in the presence of noise and non-stationary effects.
- A new plant-wide oscillation detection technique based on the mode alignment property of MEMD is proposed. The proposed method does not require any *a-priori* assumption about the process, can handle non-stationary effects, helps in root cause diagnosis and is full automated.
- A fully automated algorithm to detect the non-linearity induced oscillations via harmonics analysis is also developed. Although detection of non-linearity via harmonics is an old concept; any automatic method has still not been reported. Therefore, the algorithm proposed in this thesis has provided the required solution.
- The domain of surrogate analysis has been extended to non-stationary data by combining the MEMD with the delay vector variance algorithm.

- Moreover, the convergent cross mapping (CCM) technique that is applicable to short term time series has also been exploited to identify source of the plant-wide disturbances.

1.3 Publications

1.3.1 Journal Publications

Paper A : Aftab et al. (2018) Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Improved Oscillation Detection via Noise Assisted Data Analysis*". In: Control Engineering Practice (Submitted)(2018).

Paper B : Aftab et al. (2017e) Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Plant-wide oscillation detection using multivariate empirical mode decomposition*". In: Computers and Chemical Engineering(Revised submitted) (2018).

Paper C : Aftab et al. (2017c) Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Detecting non-linearity induced oscillations via the dyadic filter bank property of multivariate empirical mode decomposition*". In: Journal of Process Control 60 (2017), pp. 68 - 81. ISSN: 0959-1524.

Paper D : Aftab et al. (2017d) Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Diagnosis of plant-wide oscillations by combining multivariate empirical mode decomposition and delay vector variance*". In: Journal of Process Control(Submitted) (2017).

1.3.2 Conference Publications

Paper E : Aftab et al. (2017a) M. F. Aftab, M. Hovd, and S. Sivalingam. "*Convergent cross mapping (CCM) based approach for isolating the source of plant-wide disturbances*". In: 2017 IEEE Conference on Control Technology and Applications (CCTA). Aug. 2017, pp. 1492 - 1498.

Paper F : Aftab et al. (2016) Muhammad Faisal Aftab, Morten Hovd, Norden E. Huang, and Selvanathan Sivalingam. "*An Adaptive Non-Linearity Detection Algorithm for Process Control Loops*". In: IFAC-PapersOnLine. Vol. 49. 7. 2016, pp. 1020 - 1025.

Paper G : Aftab et al. (2017b)Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*A Delay Vector Variance based Approach*

for Detecting and Isolating the Non-linearity Induced Oscillations in Control Loops". In: IFAC-PapersOnLine. Vol. 50. 1. July 2017, pp. 7975 - 7980.

1.3.3 Patents

PATENT A : Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Method for detection of harmonics of univariate signal*". European patent. Filing date 28-09-2016.

PATENT B : Muhammad Faisal Aftab, Morten Hovd, and Selvanathan Sivalingam. "*Detection of plant-wide oscillations in control variables for an industrial plant control system*". European patent. Filing date 14-09-2016.

1.4 Thesis Organization

The thesis is written as collection of papers with papers A-E forming chapters 3-7 respectively. The papers are reformatted to conform with the formatting of current thesis and also contain some additional/background information that cannot be included in original papers due to length constraints. The core work is divided into two parts namely *a*) Oscillation detection, and *b*) Oscillation diagnosis . Part I (Oscillation detection) comprises chapters 3 and 4, whereas Part II (Oscillation diagnosis) contains chapters 5- 7. The thesis is organised as follows.

Chapter 2 gives the background of the controller performance monitoring and the work done therein. Moreover, the preliminaries of the multivariate empirical mode decomposition (MEMD), are also discussed.

Chapter 3 (Paper A) presents an improved oscillation detection scheme via noise assisted data analysis to identify the multiple oscillations in individual control loops (Aftab et al. 2018).

Chapter 4 (Paper B) discusses the application of MEMD for the plant-wide oscillation detection. The approach groups different variables oscillating with the same frequency and hence common cause (Aftab et al. 2017e) .

Chapter 5 (Paper C) presents work pertaining to the detection of non-linearity induced oscillations in control loops using the harmonics analysis. The extent of non-linearity is quantified using an index called degree of non-linearity (DNL) (Aftab et al. 2017c; 2016).

Chapter 6 (Paper D) uses the time series analysis technique the delay vector variance (DVV) in combination with the MEMD to identify the non-linearity induced

oscillations. This integrated approach makes DVV analysis applicable to non-stationary time series as well as identifying the multiple sources of oscillations (Aftab et al. 2017d;b).

Chapter 7 (Paper E Aftab et al. (2017a)) explains the use of convergent cross mapping based causality analysis technique for isolating the source of plant-wide disturbance.

Chapter 8 outlines the conclusions and the future scope of work.

It is to be noted that each paper is written to be self-contained, and that the chapters 3-7 essentially are a collection of these papers. This means that there is some material (in particular related to the MEMD) that is presented several times. The reader may choose to skip these repeated presentations of background material in each paper. Note, however, that each paper makes a distinct contribution to some aspect of control performance monitoring.

Finally, the appendix contains the two conference papers (Papers F and G). These conference papers form the basis for the subsequent journal publications, the papers C and D respectively. The details given in these papers have been covered by the journal versions. Therefore, these papers are only included as appendices, in facsimile, for the sake of completeness.

Chapter 2

Background

Industrial control systems are aimed at reducing product variability and optimizing the utilization of resources. Thus, control loops need to perform well in order for the plant to fulfil both economic and environmental objectives. Unfortunately, owing to the reasons described in section 1.1, a large number of control loops actually degrade the system performance. The deviation from the desired performance may lead to increased power consumption, equipment wear and material loss. Therefore it is quite natural to look for the loops that are performing poorly and identify the source or the root cause of the poor performance.

Different techniques, both data driven and model based, have been investigated, in previous works, to monitor control performance and to diagnose the root cause of the degraded operation. Model based techniques, such as Srinivasan et al. (2005b), Lee et al. (2008), Jelali (2008), are difficult to use and need detailed process knowledge and mathematical models that are seldom available due to the cost and expertise required to develop and maintain such a model. Data driven techniques, on the other hand, being more flexible and requiring minimal process knowledge, are therefore more popular and sought-after. Moreover, the rapid advancement in computer hardware and software technologies, including cheap and high quality data storage, has made it possible to store large chunks of control loops data at higher sampling rates. These developments have provided further impetus to the data driven monitoring and diagnosis paradigm.

The two most important concepts in the data driven methods for control performance assessment are (Duan et al. 2014, Thornhill and Horch 2007) *a*) Detection of disturbances, *b*) Diagnosis of root cause/source of disturbance. Further classification for both these concepts, along with list of important research contributions,

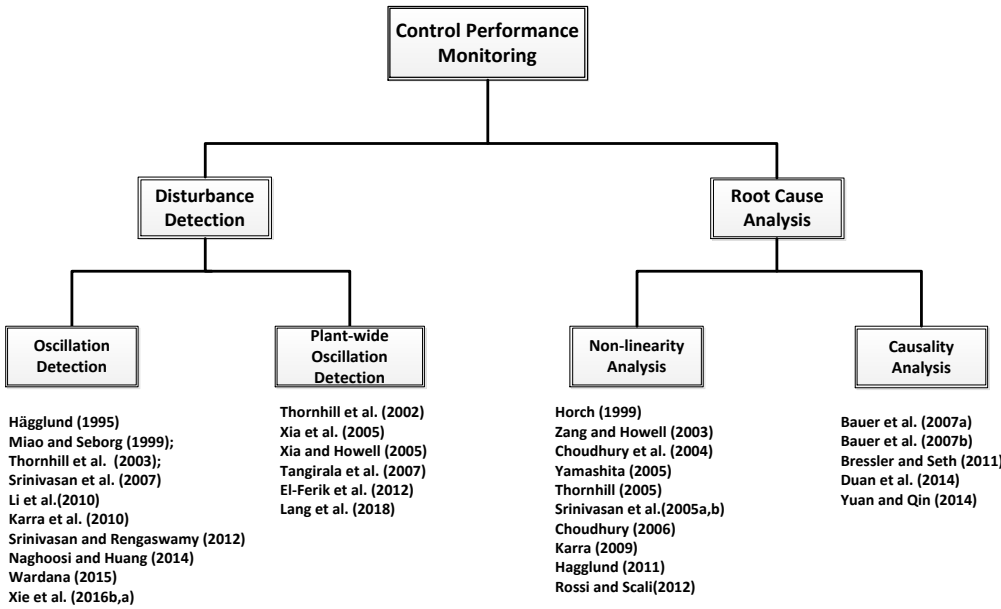

Figure 2.1: Control performance monitoring classification.

can be seen in Figure 2.1.

2.1 Disturbance/Oscillation Detection

Control performance degradation can have different manifestations, like sustained deviation from set-point and, more commonly, as oscillations. There are two types of oscillation detection concepts *a*) oscillation detection in individual loops, *b*) plant-wide oscillation detection. A brief overview of both these concepts and the associated research work has been provided in the subsequent sections.

2.1.1 Oscillation Detection in Individual Loops

This concept tries to identify in an automated manner about the state of an individual loop, i.e., whether it is oscillating or not. Reliable and accurate detection of oscillation is paramount in identifying control loops requiring further investigation. Such further investigation may involve looking for the cause of oscillation within the loop itself, or looking for causes elsewhere in the plant. The detection

process is not trivial, owing to the presence of noise, multiple oscillation frequencies and presence of non-stationary disturbances.

A simple on-line approach based on monitoring the control error has been proposed by Hägglund (1995). The method uses integral absolute error (IAE) between successive zero crossings to detect the presence of oscillations. A small value of IAE, between zero crossings, represents areas of good control, whereas a higher value indicates poor performance.

Srinivasan et al. (2007) proposed a modified Empirical Model Decomposition (EMD) method where a non-constant mean is removed using EMD. The zero crossings are identified from the cumulative area of this mean shifted data to detect oscillations. The Discrete Cosine Transform (DCT) has been used by Li et al. (2010) to detect multiple oscillations.

The majority of the researchers have used the Auto Covariance Function (ACF) for detection and characterization of oscillations. The ACF of the oscillating signal preserves the same oscillatory pattern as the original signal but is less affected by white noise. This idea is being used by Miao and Seborg (1999) to detect excessively oscillatory loops by measuring the decay ratio between consecutive extrema. Thornhill et al. (2003) used zero crossings of the ACF to identify the period and regularity index of the oscillations. Srinivasan and Rengaswamy (2012) proposed a method based on the combination of EMD, ACF and Fourier analysis for the oscillation detection in individual loops. Recently Naghoosi and Huang (2014) proposed another method where similar peaks of the ACF are clustered together and the time period between these peaks is used to detect multiple oscillations in the signal.

Although ACF based methods are robust against white noise, the presence of multiple oscillations make things complicated. Thornhill et al. (2003) proposed using a filter to separate out multiple oscillations, but the tuning of filter parameters is not trivial as the frequencies of oscillation are not known *a-priori*. The filtering requirement has been eliminated by the use of the EMD by Srinivasan and Rengaswamy (2012). Though the method can handle non-linear and non-stationary time series and gives better results than DCT (Li et al. 2010), it suffers from the mode mixing problem of the standard EMD process. The mode mixing problem can give erroneous results in case of multiple oscillations, large noise levels and coloured noise.

In order to address these issues, a noise-assisted approach (Rehman et al. 2013) is used in this work to provide a robust and improved oscillation detection method. The proposed method is shown to be robust against multiple oscillations, coloured

noise and varying noise strengths. Details are provided in Chapter 3.

2.1.2 Plant-Wide Oscillation Detection

Owing to control loop interactions as well as material and energy flows, a large number of loops can be affected by an oscillation emanating at a single point, leading to plant-wide oscillations. Plant-wide oscillation detection deals with detecting several control loops oscillating due to common cause and hence with a similar frequency content and grouping them together. Such a concept is important as different loops oscillating due to a common cause can be identified, and root cause analysis is then directed at only the loops affected by the particular oscillation, thereby reducing the problem size to a considerable extent.

The plant-wide oscillation detection problem has been mostly tackled via similarity in the power spectrum of the control loops data. Tangirala et al. (2005) used the correlation among the power spectra to construct the power spectral correlation map (PSCMAP). The map provides information regarding the similarity in the power spectra based on a power correlation index (PCI).

Dimensionality reduction techniques, like the principal component analysis (PCA) and spectral independent component analysis (ICA) have been proposed by Thornhill et al. (2002) and Xia et al. (2005), Xia and Howell (2005) respectively. Each principal or independent component represents a group of variables oscillating with common frequency.

Tangirala et al. (2007) highlighted the limitations of both PCA and ICA methods and instead proposed non-negative matrix factorization (NMF) for the dimensionality reduction. The pseudo singular values are used to reduce the dimensionality in an analogy to the use of singular value decomposition in PCA and ICA. Xia et al. (2007) has proposed NMF, with sparseness constraint, to detect the plant-wide oscillations.

The use of power spectra is the common building block of all these methods, thereby making these tools vulnerable in case of time varying disturbances (non-stationary effects). A fully data driven, adaptive and automated plant-wide oscillation detection method based on multivariate empirical mode decomposition (MEMD) is proposed in this thesis. The method, given in Chapter 4, is capable of handling non-stationary effects and requires minimum assumptions about the data or process.

2.2 Root Cause Analysis

Root cause analysis, also termed as oscillation diagnosis, is the most important step as it can facilitate targeted maintenance thereby reducing the shut down time. The

root cause analysis can be further subdivided into two sub categories (Thornhill and Horch 2007) (Figure 2.1), *a*) non-linearity analysis, *b*) causality analysis .

2.2.1 Non-Linearity Analysis

Non-linearity induced oscillations are characterised by the presence of limit cycles that exhibit harmonics. There can be many causes of non-linearity induced oscillations like *a*) valve stiction, *b*) sensor faults, and *c*) process non-linearities (Thornhill and Horch 2007) to name a few. Valve non-linearities are more common and account for around third of all non-linearity induced oscillations (Rossi and Scali 2012). Detection of non-linearity has attracted considerable attention and several different methods have been proposed.

Choudhury et al. (2008) proposed use of higher order spectra to detect non-linearities in control loops, using two indices called non-Gaussianity index (NGI) and non-linearity index (NLI). Non-linearity exhibited by symmetric waveforms (like square or triangular) exhibiting odd harmonics, cannot be captured (Thornhill 2005, Zang and Howell 2003) by this method. Horch (1999) proposed a correlation based method to detect valve stiction in non-integrating plants. Srinivasan et al. (2005a), Hägglund (2011) and Yamashita (2005) have proposed methods based on shape formalisms to detect stiction by comparing the waveform of the data with pre-defined trends. Non-linear time series methods have been investigated by Thornhill (2005) to detect presence of non-linearities in control loops.

The fact that the non-linearity induced oscillations are characterised by presence of harmonics (Thornhill et al. 2001) can be used to detect the presence of non-linearities. The dyadic filter bank property of multivariate empirical mode decomposition (MEMD) and a non-linearity measure degree called the non-linearity (DNL) has been investigated in this thesis (Chapter 5) (Aftab et al. 2016; 2017c). This method can work for non-stationary time trends and can isolate the source of the non-linearity as well. Similarly combination of time series analysis and MEMD are also presented (Aftab et al. 2017d;b) in Chapter 6.

2.2.2 Causality Analysis

Investigation of cause and effect relationships, also known as causality analysis, can be used to isolate the source of plant-wide disturbances. As already discussed, disturbance originating at one point in the plant may spread owing to process flows, therefore the cause and effect relationship can be exploited to ascertain the origin of disturbances.

Causality is defined as follows: *if prediction of time series y is improved by using the knowledge of another time series x, then x has a causal influence on y* (Wiener

1956). Granger (1969) proposed the concept of Granger causality (GC) that x "Granger causes (G-Causes)" y if removing x from the universe of all possible causative variables U decreases the predictability of y, i.e., $\sigma^2(y|U) < \sigma^2(y|(U - x))$ (σ^2 is variance of y). The GC has been applied for diagnosis of plant-wide oscillations by Yuan and Qin (2014).

The transfer entropy (TE) concept has been employed by Bauer et al. (2007a) to study the disturbance propagation path. The method estimates the joint probability density functions to ascertain the direction of information flow. The estimation of joint and conditional probability density functions (pdfs) is computationally intensive and needs sufficiently large data sets.

Another important concept is using the near neighbours, found from the time delayed embedding of the time series data. The time series $x(t)$ from a dynamical system is used for reconstruction of the attractor manifold, with embedding dimension m , using the time lagged vectors $\mathbf{x} = [x(t) \quad x(t-\tau) \quad x(t-2\tau) \ldots (t-(m-1)\tau)$; where τ is the time lag (Huffaker 2010, Takens 1981, Sauer et al. 1991, Kantz and Schreiber 2004). A number of studies have used this concept to find the relationship or interdependence among time series (Pereda et al. 2001, Chicharro and Andrzejak 2009, Arnhold et al. 1999). Bauer et al. (2007b) has used nearest neighbours method for for the diagnosis of plant-wide disturbances. This method is also sensitive to tuning parameters like embedding dimension, prediction horizon and the number of nearest neighbours used.

A convergent cross mapping (CCM) algorithm (Sugihara et al. 2012), applicable to short time series, has been used in this thesis to identify the source of the plant-wide oscillations (Aftab et al. 2017a). The method is also based on the nearest neighbour concept. Moreover, automatic determination of embedding dimensions has also been augmented with the CCM approach. The details are presented in chapter 7.

2.3 Empirical Mode Decomposition (EMD) and Variants

This thesis has made use of the multivariate version of empirical mode decomposition for detection and diagnosis of control loop oscillations. A summary of univariate and higher order variants of empirical mode decomposition is given in the subsequent sections whereas, the details and the associated properties of multivariate version is explained in the subsequent sections.

2.3.1 Empirical Mode Decomposition (EMD)

Empirical mode decomposition (EMD) is an adaptive signal tool that decomposes the signal into fast and slow orthogonal components via an iterative sifting process.

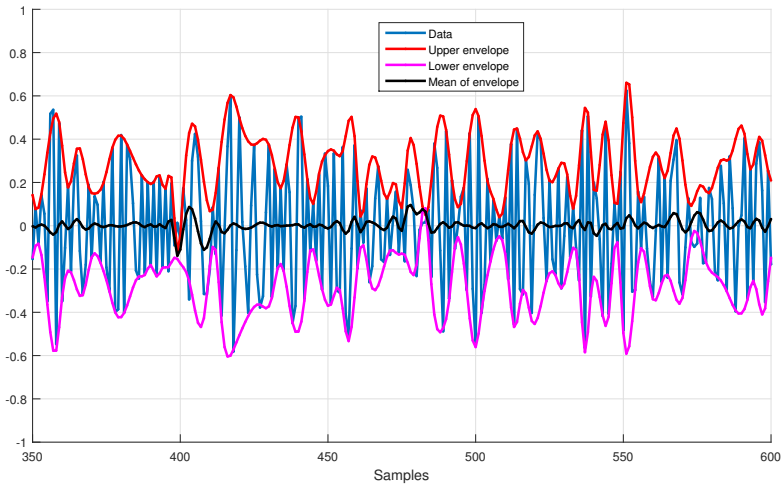

Figure 2.2: A snapshot of sifting envelope.

These orthogonal components are analogous to the basis functions in the conventional Fourier and wavelet analysis, but in contrast to Fourier and wavelet analysis the components resulting from EMD are not fixed. The components so obtained are called Intrinsic Mode Functions (IMFs). The IMFs are functions that have the following properties *a*) zero mean, *b*) the number of extrema and zero crossing differ at most by one. The sifting process starts with identifying the extrema of the signal $x(k)$ and fitting a spline envelope through these extrema (Figure 2.2). This envelope is averaged to generate mean envelope $m(k)$ that gives the slow component. The mean is then subtracted from the signal to find the fast component called detail $d(k)$, given by (Huang et al. 1998, Rilling et al. 2003):

$$d(k) = x(k) - m(k) \tag{2.1}$$

The procedure is iterated upon $d(k)$ until a valid IMF $c_1(k)$ is extracted. This IMF is subtracted from $x(k)$ and the sifting process is repeated on the remainder till there are no more IMFs to be extracted. The signal is thus decomposed into N IMFs, each denoted by $c_i(k)$, and a residue $b(k)$:

$$x(k) = \sum_{i=1}^{N} c_i(k) + b(k) \tag{2.2}$$

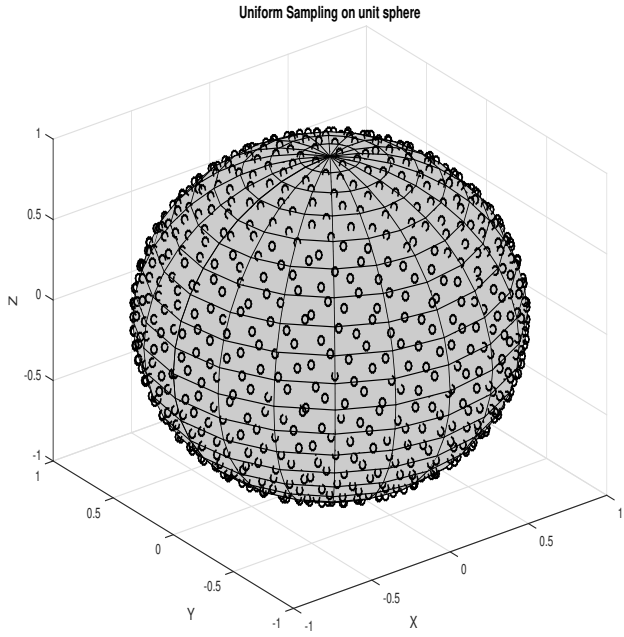

Figure 2.3: Random sampling with uniform distribution on the unit sphere (Aftab et al. 2017c).

2.3.2 Multivariate Empirical Mode Decomposition (MEMD)

Multivariate empirical mode decomposition (MEMD) is the extension of the univariate case (single variable signal) to the multi-dimensional signals (a signal consisting of more than one variable). The major challenge in this regard is the calculation of envelope and respective mean in the higher dimensions. Rehman and Mandic (2010a) proposed the use of signal projections in multiple directions to calculate the envelope. The multiple directions are represented by the direction vectors, formed by uniform sampling of the surface of the unit n-dimensional sphere (n is number of variables in multivariate signal). A snap shot for the three dimensional case, with X, Y, Z as coordinate axis, is given in Figure 2.3. We refer to subsequent chapters for further information on MEMD, as the procedure is described in more detail there.

2.4 Important Nomenclature

Some of the important terms used in the thesis are explained in this section.

1. The term *univariate signal* refers to a time series consisting of single variable. Similarly *bi-variate*, *tri-variate* and n-*variate* refers to the signal comprising of two, three and n variables respectively.

2. Multivariate signal means a signal consisting of more than one variable. For instance a multivariate signal (or n-dimensional signal) $\mathbf{X}$ consisting of n variables $x_1 \dots x_n$ with each variable having l samples is given by

$$\mathbf{X} = \begin{Bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(l) & x_2(l) & \dots & x_n(l) \end{Bmatrix} \tag{2.3}$$

3. The univariate signal and the channel are one and the same thing. Similarly n-channel, n-dimensional and n-variate signal are used interchangeably.

4. An *oscillatory mode* or a *tone* both represent a single oscillation with a particular frequency. Similarly two oscillatory modes or two tone signal refers to the presence of two oscillating frequencies in a signal.

2.5 Summary

This chapters gives an overview of the different concepts in the field of controller performance monitoring and presented a summary of the previously reported research contributions. Moreover, an introduction to EMD and its multivariate variant, the MEMD, is also provided along with the definition of certain terms used in the thesis.

Part I

Oscillation Detection

Chapter 3

An Improved Oscillation Detection Method

Paper A: Improved Oscillation Detection via Noise-Assisted Data Analysis

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

Control Engineering Practice (Submitted 2018)

Abstract

Oscillation detection is usually a precursor to more advanced performance monitoring steps such as plant-wide oscillation detection and root cause detection. Therefore any false or missed detection can have serious implications. Oscillation detection is a challenging problem due to the presence of noise and multiple oscillations in the plant data. This paper presents an improved and robust automatic oscillation detection algorithm based on noise assisted data analysis that can handle multiple oscillation frequencies in the presence of both coloured and white noise along with non-stationary effects. The dyadic filter bank property of multivariate empirical mode decomposition has been used to accurately detect the oscillations and to calculate the associated characteristics. This work improves upon the existing auto covariance function based methods. The robustness and reliability of the proposed scheme is demonstrated via simulation and industrial case studies.

3.1 Introduction

Oscillation detection is an important aspect of control loop performance assessment (CLPA) owing to the fact that about 30% of all industrial control loops are reported to be oscillating (Srinivasan et al. (2007)). Oscillations can be caused by a variety of issues like process degradation, poor controller tuning, presence of nonlinearities and external disturbances. The presence of oscillations is viewed as the foremost indicator of performance degradation and therefore an accurate and automatic oscillation detection tool is regarded as the backbone of any performance monitoring mechanism.

Oscillations can have significant impact on the profitability and economics of any industrial process, therefore their accurate and timely detection is a key ingredient of any performance monitoring tool. Reliable and accurate detection of oscillation is paramount in identifying control loops requiring further investigation. Such further investigation may involve searching for the cause of oscillation within the loop itself, or looking for causes elsewhere in the plant. Thus detection of any false mode (oscillation) or failure to identify any real oscillation can ruin the whole diagnosis process. Furthermore the accurate estimate of the oscillation characteristics like period of oscillation and amplitude are also helpful in plant-wide oscillation detection and fault localization.

Oscillation detection is an active research area and several methods have been developed over the years to detect the oscillating variables so that the proper remedial measures can be taken. Broadly speaking oscillation detection can be subdivided into two groups namely *a*) oscillation detection in individual loops/variables, and *b*) plant-wide oscillation detection where loops or variables oscillating with similar frequencies are grouped together to look for the common cause of performance degradation

The latter group is not an oscillation detection method in a stricter sense, as it only groups different variables without identifying the presence or extent of oscillations within each loop (Li et al. 2010). Therefore this work will be focused on the first type, i.e. oscillation detection, and a summary of some oscillation detection methods is presented here. More detailed description can be found in review papers by Thornhill and Horch (2007) and Bacci di Capaci and Scali (2018).

Hägglund (1995) proposed an on-line, simple and effective procedure based on monitoring of the control error and computing integral absolute error (IAE) between successive zero crossings, to detect the oscillations. The idea is that the IAE value will be small during the duration of good control and will be higher otherwise.

A modified Empirical Model Decomposition (EMD) method is proposed by Srinivasan

et al. (2007) where a non-constant mean from plant data is removed using modified EMD process. Zero crossings, identified from the cumulative area of this mean shifted data, are used for the detection of oscillations.

Li et al. (2010) have developed a method where the Discrete Cosine Transform (DCT) is used to isolate the different frequency components in the oscillatory signal followed by checking the regularity of zero crossings in these isolated components to identify the presence or absence of oscillations.

Several authors have investigated the use of the Auto Covariance Function (ACF) for detection and characterization of oscillations. The ACF of an oscillating signal oscillates with same frequency as the original signal and at the same time it is less sensitive to noise as white noise is confined to zero lag only. This idea is being used by Miao and Seborg (1999) to detect excessively oscillatory loops by measuring the decay ratio between consecutive extrema. Thornhill et al. (2003) proposed to use zero crossings of the ACF to identify the period and regularity index of the oscillations in the signal. Srinivasan and Rengaswamy (2012) proposed a method based on the combination of EMD, ACF and Fourier analysis to detect and characterize the oscillations in individual loops. Of late another method based on ACF is proposed by Naghoosi and Huang (2014) where similar peaks of the ACF are clustered together and the period of oscillation between these peaks is used to detect multiple oscillations in the signal.

Of all the methods listed above the ACF based methods are more robust and reliable in the presence of noise, that is always present in physical measurements and can complicate the analysis to a significant extent. But in spite of all these advantages the presence of multiple oscillations in the presence of white and coloured noise needs special treatment and care even in the ACF based methods. In the work by Thornhill et al. (2003), a filter is used to filter out the different oscillatory modes before the detection procedure. Tuning of filter parameters is an uphill task and needs good understanding of the underlying process dynamics. The EMD based approach given by Srinivasan and Rengaswamy (2012) gets rid of the filter requirement and is reported to be performing better than the DCT based approach given by Li et al. (2010), and can handle non-linear and non stationary time series, but it suffers from inherent limitations associated with EMD process itself, i.e. the EMD process is prone to mode mixing, especially in the presence of noise and multiple oscillations. Presence of coloured noise complicates the things further as the ACF of the coloured noise is no longer confined to the zero lag and significant correlation exists at other lags as well. This mode mixing can adversely affect the oscillation detection mechanism and has a tendency to cause erroneous results. The mode mixing problem will be discussed in detail in the subsequent sections. Moreover, the accuracy of existing schemes decrease significantly with

the increase in the noise variance and with presence of the coloured noise.

3.1.1 Contribution of Paper

In this work, the limitations of the EMD based oscillation detection process are highlighted using simulation studies as well as industrial data and an improved oscillation detection method based on Noise-Assisted Multivariate EMD (NA-MEMD) is presented. The dyadic filter bank property of MEMD is exploited to improve the accuracy and reliability of the oscillation detection mechanism. The proposed method can handle multiple oscillations in the presence of both white and coloured noise with equal robustness and reliability.

This paper is organized as follows. Section 3.2 gives the detailed description of the proposed method with a summary of EMD and the mode mixing issues related with it. It also highlights the basics of MEMD and NA-MEMD procedure. Simulations studies are presented in the section 3.3 and finally industrial case studies are given in section 3.4 followed by conclusions.

3.2 Noise-Assisted Oscillation Detection

In this work, the use of noise-assisted multivariate EMD (NA-MEMD) is proposed to detect the presence and frequency of oscillations in the signal. This method can handle both non-linear and non-stationary time series and is found to be better than existing EMD based method proposed by Srinivasan and Rengaswamy (2012). Here the mode alignment and dyadic filter bank property of MEMD is utilized to formulate a robust and reliable oscillation detection mechanism. The advantages of the proposed method over the standard EMD based method Srinivasan and Rengaswamy (2012) are highlighted using Monte-Carlo simulations with varying noise levels.

The proposed method is more accurate owing to the fact that it is less prone to the mode mixing problem, that is an inherent limitation of standard EMD. The input signal is decomposed into constituent IMFs using NA-MEMD. The IMFs so obtained are converted to the corresponding ACFs and the zero crossings are used to detect the presence or absence of oscillations. In this section the limitations of the standard EMD process, especially with regards to mode mixing problems, are highlighted. Next, a brief overview of the multivariate EMD is given, followed by the steps involved in proposed oscillation detection algorithm.

3.2.1 Empirical Mode Decomposition: Basics and Inherent Limitations

Empirical Mode Decomposition

Empirical mode decomposition is a recent addition to the family of adaptive signal processing techniques. This method adaptively decomposes the signal into sub-components called Intrinsic Mode Functions or IMFs in a purely adaptive and data driven way. Each IMF is a function that has zero mean and the number of extrema and zero crossing in the whole data set must either be equal or at most differ by one. The advantage lies in the fact that the procedure doesn't need any *a priori* assumption or knowledge about the underlying process dynamics (Huang et al. 1998). The EMD process basically extracts fast modes (oscillations) from the input data ($x(t)$) by iteratively removing slow modes. These slow modes are in fact local means $m(t)$ of the envelope defined by spline fitting of the extrema.

$$d(t) = x(t) - m(t) \tag{3.1}$$

where $d(t)$ represents the local fast oscillation (Rilling et al. 2003). The sifting process is iterated on d until it fulfils the IMF properties; which is then named $c_1(t)$. Once the IMF is extracted it is subtracted from the original signal and the sifting procedure is started again on the residue. This continues until there are no more IMFs to be extracted. If $c_i(t)$ is the i^{th} IMF and $b(t)$ is the residue, the sifting procedure gives

$$x(t) = \sum_{i=1}^{N} c_i(t) + b(t) \tag{3.2}$$

where N is the total number of IMFs. The details of the procedure can be seen in Huang et al. (1998) and Rilling et al. (2003).

Limitations of EMD

Although the EMD process is finding its way into a number of application areas yet it is not free from problems or shortcomings. The foremost of them all, that is well known and documented, is the mode mixing problem. Mode mixing is defined as when one oscillation or mode is present in more than one IMF or one IMF contains several contrasting modes or oscillations (Wu and Huang 2009, Rehman et al. 2013). The effect of mode mixing on the oscillation detection problem can be illustrated using examples given in Srinivasan and Rengaswamy (2012) and Li et al. (2010).

Illustrative Example

Consider a signal $x(k)$ formed as sum of two sinusoids with frequency $f_1 = 0.2Hz, f_2 = 1Hz$, a time varying bias and white noise $\nu(t)$ with variance σ_v^2

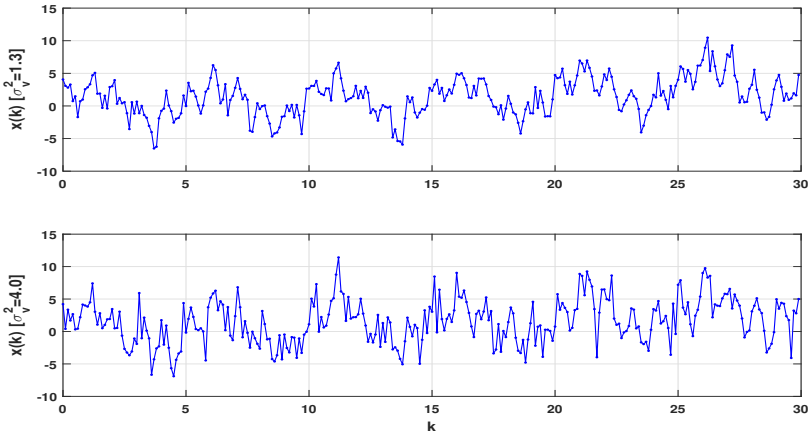

Figure 3.1: Signal x(k), as in equation (3.3), for noise variance 1.3 and 4.0.

as given in (3.3). The sampling frequency is $f_s = 10Hz$. The resulting signal for two values of noise variance $\sigma_v^2 = 1.3$ and $\sigma_v^2 = 4.0$ is shown in Figure 3.1.

$$x(k) = 0.05k^2 + \sin(2\pi f_1 k) + \sin(2\pi f_2 k) + \nu(k) \tag{3.3}$$

The two IMFs from the standard EMD method adopted in Srinivasan and Rengaswamy (2012) representing the oscillatory frequencies in $x(k)$, for two noise levels, are shown in the first two rows of Figures 3.2 and 3.3 respectively. The mode mixing problem in evident in IMF 2 as both modes with frequency f_1 and f_2 are present within one IMF. This effect is more pronounced in the IMFs corresponding to the signal with higher noise levels i.e $\sigma_v = 4.0$. Any use of these IMFs, with mode mixing, for oscillation detection is prone to produce erroneous results.

How to Remove Mode Mixing

The mode mixing problem as described in the preceding section needs to be eliminated in order to get accurate and reliable results. Several efforts have been made to eliminate this problem. Wu and Huang (2009) proposed a method called Ensemble EMD (EEMD), where independent noise realizations are added to the original signal, and the EMD is applied on each such realization. The ensemble of IMFs from M such iterations is taken to get the mean IMF. The procedure can be summarized as

Step I Generate M different signals $\tilde{x}$ by adding random noise such that $\tilde{x}_i(k) = x(k) + \omega_i(k)$; where ω_i is random noise and $i = 1..M$.

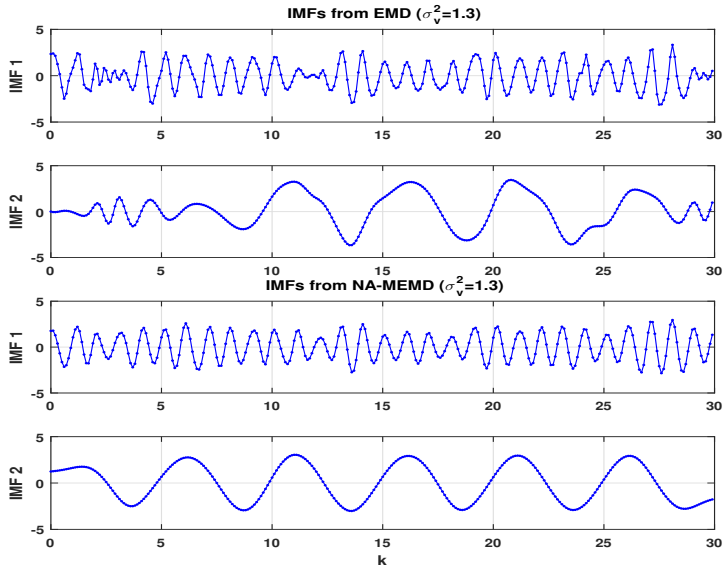

Figure 3.2: IMFs for signal in equation (3.3) for noise variance 1.3.

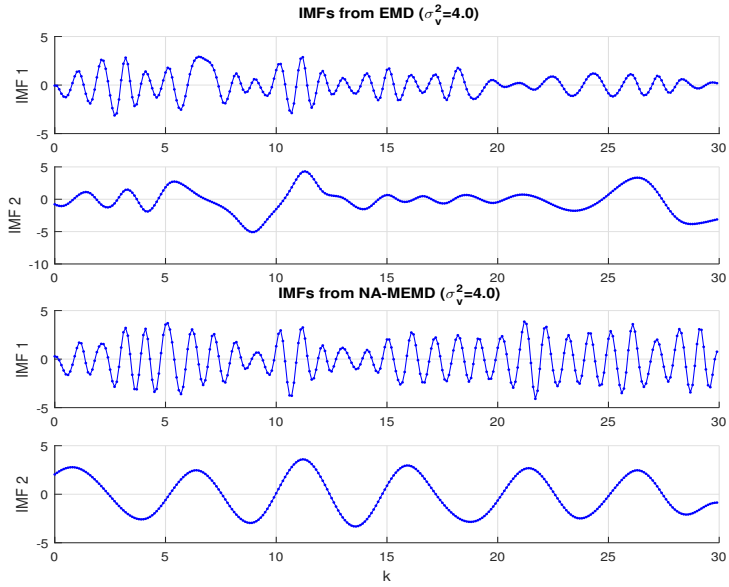

Figure 3.3: IMFs for signal in equation (3.3) for noise variance 4.0.

Step II For each i^{th} iteration perform EMD operation to get IMFs $C_j^i(k)$, where $j = 1...J$ with J total IMFs in each iteration.

Step III Get $\bar{C}_j$ the ensemble j^{th} IMF by averaging corresponding IMFs from M realizations.

$$\bar{C}_j = \frac{1}{M}\sum_{i=1}^{M} C_j^i(k) \tag{3.4}$$

Another similar approach is proposed by Torres et al. (2011) where only one IMF is extracted from an ensemble of M realizations and this IMF is subtracted from original signal to form a residue. The process is iterated upon the residue till all IMFs are extracted [1]. The Complementary Ensemble EMD (CEEMD) proposed by Yeh et al. (2010) tries to reduce the residual noise by pairs of complementary IMFs each generated from a pair of complimentary signals x_1, x_2. The complementary pair (x_1, x_2) corresponding to signal x are generated by adding and subtracting the white noise ν from the original signal, i.e.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} \tag{3.5}$$

These methods can remove mode mixing to some extent, but at the same time the direct addition of noise to the original signal has its own problems. The decomposed IMFs therefore will have noise artefacts and very large ensembles are required to effectively filter out the effects of the added noise, thereby significantly increasing computational overhead (Rehman et al. 2013). To overcome these issues a more robust and reliable method, called Noise-Assisted Multivariate Empirical Mode Decomposition (NA-MEMD) is proposed by Rehman et al. (2013).

3.2.2 Multivariate EMD (MEMD)

Multivariate EMD (MEMD) as the name suggests is an extension of the standard univariate EMD algorithm, to $n-$dimensional signals. Here $n-$dimensional signals mean signal consisting of n variables. The mathematical description of such a signal is given in equation 2.3.

The critical issue is how to get the envelope and its local mean in higher dimensional space[2]. Rehman and Mandic Rehman and Mandic (2010a) proposed to get the signal projection in multiple directions in n-dimensional space. Multiple directions are represented via direction vectors from the centre of a unit $n-$ dimensional sphere to the uniformly spaced points on its surface (Rehman and Mandic 2010a).

[1] details can be seen in Torres et al. (2011)

[2] higher here means three or more dimensions

Multivariate EMD (MEMD):: Algorithm

The MEMD of a multivariate signal $\mathbf{X}$ (given in equation 2.3) is carried out using the following steps:

Step I Set up K direction vectors $\mathbf{u^k}$ ($k = 1 \ldots K$) via uniform sampling on the n-dimensional sphere.

Step II Find the projections $p^k(t)$ of input signal along the direction vectors using relation

$$p^k(t) = \mathbf{X}u^k \quad \forall \quad k = 1 \ldots K \tag{3.6}$$

Step III Identify the extrema of projections $p^k(t)$ and the corresponding time instants t^k.

Step IV Generate n-dimensional envelope curve $\mathbf{e}^k(t)$ by interpolating $[t^k, \mathbf{X}(t^k)]$.

Step V The n-dimensional mean of envelope curve is then given by

$$\mathbf{m} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{e}^k(t) \tag{3.7}$$

Step VI Extract n-dimensional detail $\mathbf{d(t)}$ using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$.

Step VII Repeat the steps I-VI using $\mathbf{d(t)}$ as input till it fulfils the criteria for an IMF.

Step VIII Calculate the residue $\mathbf{b(t)} = \mathbf{X(t)} - \mathbf{d(t)}$ and iterate the procedure on $\mathbf{b(t)}$ till there are no more IMFs left to be extracted[3].

Dyadic Filter Bank Property of Multivariate EMD

An important aspect of EMD is its dyadic filter bank property in the presence of noise. By virtue of this property the EMD process essentially behaves like a sequence of band pass filters. Flandrin et al. (2004; 2014), Wu and Huang (2004) have empirically demonstrated the dyadic filter bank characteristics of the EMD by applying it to a white noise sequence. The same dyadic filter bank structure of standard EMD is preserved in higher order variants i.e. MEMD (Rehman and Mandic 2011).

The dyadic filter bank property is further elaborated here by processing a multivariate signal consisting of three white noise sequences using MEMD. The average power spectrum of resulting IMFs from 1000 noise realisations (Figure 3.4)

[3]The process is stopped when residue is an IMF or contains less than three extrema

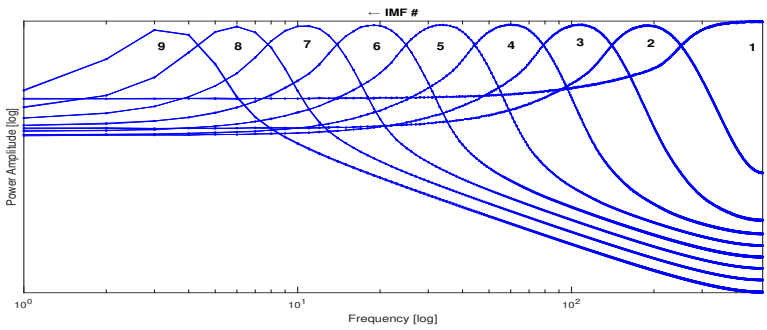

Figure 3.4: Dyadic filter bank property of multivariate EMD (Aftab et al. 2017c).

shows two important properties of multivariate EMD. *a*) The IMF power spectrum can be seen as the output of a series of band pass filters, with band frequencies decreasing with IMF index, and *b*) the mode alignment characteristics i.e. same indexed IMFs have similar frequency content.

3.2.3 Noise-Assisted MEMD (NA-MEMD)

The dyadic filter bank and mode alignment property of MEMD, as discussed in the previous section, is exploited to eliminate the mode mixing problem in the standard EMD procedure. The idea is to add two white noise sequences to the signal under study to make a multivariate signal consisting of three variables. This three variable signal is processed using MEMD and the IMFs corresponding to the original signal are retained whereas IMFs of the noise sequences are discarded Rehman et al. (2013). This procedure is called Noise-Assisted MEMD (NA-MEMD) and the detailed algorithm is given in Table 3.1.

As the noise signals span a broad frequency range, the MEMD arranges their IMFs according to the dyadic filter bank structure. The IMFs of the original signal also follow the same pattern, thereby reducing the mode mixing to a considerable extent. One important aspect is that the noise is not added to the actual signal and the issues highlighted in section 3.2.1 are completely avoided.

The power of this method is visible in the IMFs generated for signal in equation 3.3 with varying noise levels. The IMFs are shown in last two rows of Figures 3.2 and 3.3 respectively. It can be seen that by processing the same signal with NA-MEMD the mode mixing problem in standard EMD is significantly reduced and the IMFs can be effectively used for further analysis like oscillation detection.

Table 3.1: Noise-assisted multivariate EMD algorithm

Step I	Generate two uncorrelated white Gaussian noise sequences/signals with the same length as that of the original signal.
Step II	Add the two noise sequences (from step I) to the original signal to make a multivariate signal of three variables/channels [4].
Step III	Process the signal using MEMD. The resulting IMFs will have three channels.
Step IV	Retain the IMFs in the channels corresponding to the original signal, and discard the IMFs corresponding to the noise signals.

3.2.4 Discarding Spurious and Noisy IMFs

Discarding Spurious IMFs

The NA-MEMD method used in this work, like the standard EMD process, may also generate spurious IMFs, due to spline fitting issues, as highlighted in (Peng et al. (2005)), Aftab et al. (2016), Srinivasan and Rengaswamy (2012). These pseudo-components will be poorly correlated with the original signal as both EMD and MEMD yields near orthogonal IMFs. The correlation coefficient therefore can be used to identify significant IMFs as reported in Peng et al. (2005), Srinivasan and Rengaswamy (2012) and Aftab et al. (2016). The correlation coefficient ρ_i of the i^{th} IMF c_i with signal $x(t)$ is calculated from

$$\rho_i = \frac{Cov(c_i, x)}{\sigma_x \sigma_{c_i}}, \quad i = 1, 2, 3 \ldots N \tag{3.8}$$

where Cov denotes the covariance; σ_x and σ_{c_i} are the standard deviations of the signal and the IMF, respectively, and N is total number of IMFs. Next the correlation coefficient ρ_i is normalized using 3.9. IMFs with normalized coefficient λ greater than a certain threshold η are retained, while the others are eliminated and added to the residue. The threshold $\eta = 0.5$ is used in this work to remain consist-

[4] here channel is synonymous with variable

ent with the standard EMD based method Srinivasan and Rengaswamy (2012), so that results can be compared in an unbiased manner.

$$\lambda_i = \frac{\rho_i}{max(\rho_i)}, \quad i = 1, 2, 3 \ldots N \tag{3.9}$$

Discarding Noisy IMFs

NA-MEMD, like the standard EMD may produce noisy IMFs that are correlated with the original signal (especially when noise amplitude is higher). In order to eliminate these IMFs, the method proposed by Hoyer (2004) and used by Srinivasan and Rengaswamy (2012) is adopted. The sparseness index of signal X (X denotes the magnitude of power spectrum of the signal x), given by (3.10), will be approximately zero for a white noise signal and will be nearly one for periodically oscillatory signal.

$$Sparseness(x) = \frac{\sqrt{I} - \left(\sum_{i=1}^{I} |X_i| / \sqrt{\sum_{i=1}^{I} |X_i|^2} \right)}{\sqrt{I} - 1} \tag{3.10}$$

where I gives the number of frequency bins in power spectrum. IMFs with sparseness greater than certain threshold are retained. The threshold used in this work is 0.5 to be consistent with the standard EMD based procedure, so that the comparison of proposed method is not influenced by the choice of this threshold.

3.2.5 Auto Covariance Function (ACF) of IMFs

The Auto Covariance Function (ACF) of a signal is a popular choice to analyse oscillations owing to the fact that white noise is confined to zero lag. The zero crossings of ACF of the oscillatory IMFs, that have qualified the correlation and sparseness test, is used to compute the average period of oscillation $\bar{T}_p$ and regularity index r of oscillations. If Δt is the time interval between two successive zero crossings, then the average period of oscillation $\bar{T}_p$ for H such intervals will then be given by (Thornhill et al. 2003, Srinivasan and Rengaswamy 2012)

$$\bar{T}_p = \frac{2}{H} \sum_{i=1}^{H} (\Delta t_i) \tag{3.11}$$

and the corresponding r statistics, that shows the regularity of oscillations is then

$$r = \frac{1}{3} \frac{\bar{T}_p}{\sigma_{T_p}} \tag{3.12}$$

where σ_{T_p} is the standard deviation of the time intervals between zero crossings.

As discussed by Thornhill et al. (2003), oscillations are reported only if the regularity index r for that particular signal is greater than 1. Furthermore, as recommended by Thornhill et al. (2003), the first 11 zero crossings of the ACF are used in this work for calculating the average period of oscillation and regularity index .

3.2.6 Proposed Algorithm

The proposed oscillation detection algorithm based on Noise assisted MEMD is given below.

Step I Process the signal using NA-MEMD according to the steps in Table 3.1.

Step II Discard the IMFs that are noisy or uncorrelated with the original signal according to criteria given in section 3.2.4.

Step III Calculate the auto correlation function (ACF) of each retained IMF.

Step IV Calculate the regularity index of oscillation r, mean period of oscillation $\bar{T}p$ and standard deviation σ_{T_p}.

Step V Report the presence of oscillation if regularity index $r > 1$ along with the estimates of $\bar{T}p$ and σ_{T_p}.

3.2.7 Illustrative Example Continued

The efficacy of the proposed scheme is now highlighted by analysing the signals in the example of section 3.2.1 by the existing and proposed methods. The results for this particular case are given in Table 3.2.

It can be seen that due to mode mixing, the existing EMD based method fails to detect the second frequency for $\sigma_v^2 = 1.3$; both frequencies are detected when $\sigma_v^2 = 4.0$, but accuracy reduced considerably. In comparison the scheme proposed

Table 3.2: Oscillation detection illustrative example

Case	Method	IMF	r value	T_p	σ_{T_p}	Oscillation
$\sigma_v^2 = 1.3$	EMD based Method	1	∞	1.0	0.0	yes
		2	0.8	4.1	1.72	no
	Proposed Method	1	∞	1.0	0.0	yes
		2	17.45	5.0	0.09	yes
$\sigma_v^2 = 4.0$	EMD based Method	1	5.16	0.98	0.063	yes
		2	1.16	4.86	1.4	yes
	Proposed Method	1	∞	1.0	0.0	yes
		2	12.10	5.0	0.13	yes

in this paper is robust and gives accurate estimates for both cases. It is emphasized again that this result is for a particular realization of added noise and is reported just to highlight the advantages of the proposed scheme, whereas the detailed analysis for 10000 such noise realizations with varying strength is given in the next section.

3.3 Simulation Studies

The simulation studies are aimed at highlighting the advantages of the proposed scheme in presence of both white and coloured noise. Unlike the white noise the ACF of the coloured noise is not confined to zero lag but instead significant correlation exists at lags other than zero. Thus, presence of coloured noise will make the detection of oscillations more challenging. Therefore, the coloured noise case is a good example to assess the efficacy of the proposed method and its performance in comparison to the existing methods.

The simulations studies, for white and coloured noise, are carried out using examples from Srinivasan and Rengaswamy (2012) and Li et al. (2010). The results for 10000 such simulation runs for each case are compared with the existing standard EMD based approach to ascertain the robustness and reliability of the proposed scheme. The standard EMD based scheme, as already discussed, is prone to report inaccurate results due to the mode mixing problem and this inaccuracy increases with increase in the noise level.

Moreover, a case of non-sinusoidal oscillations such as the ones generated by stiction non-linearity is also analysed to see how the proposed and existing methods work in this scenario.

3.3.1 Example 1: Oscillation Detection with Varying White Noise Levels

Consider a signal containing two modes or frequencies and a non-stationarity corrupted with white noise (same as Example 3 from Srinivasan and Rengaswamy (2012) and Li et al. (2010)).

$$x(k) = 0.05k^2 + \sin(2\pi f_1 k) + \sin(2\pi f_2 k) + \nu(k) \tag{3.13}$$

with $f_1 = 0.2Hz$,$f_2 = 1.0Hz$, sampling rate 0.1 second and white Gaussian noise ν with variance σ_v^2. Note that the parabolic term introduces non-stationarity in the data. Simulations with 10000 different realizations of white noise are carried out to compare the performance of proposed and existing methods for different noise variances; $\sigma_v^2 = 1.3$, $\sigma_v^2 = 3$ and $\sigma_v^2 = 4$ that gives the signal to noise ratio (SNR)[5] of 2.5, 1.7 and 1.4 respectively. The results in terms of success rate $\%$ are summarized in Table 3.3.

[5]SNR as defined in Li et al. (2010)

The oscillation detection is defined as successful if two oscillatory modes (frequencies) are reported for a simulation. Any other number of oscillatory modes mean that the detection has failed.

It can be seen that the proposed methods performed much better than the EMD based method whose detection capability is greatly compromised with increasing noise levels. For barely 39% of the simulations it could detect the two frequencies successfully when noise variance was 4 (SNR=1.4) as compared to 94% for the proposed method. The histograms of estimated mean periods of oscillation (with bin size 0.1) for the successful iterations are plotted in Figures 3.5-3.7. The vertical axis of the figures show the number of iterations for a particular estimate of period, where as the horizontal axis shows the spread of estimated period of oscillations. The proposed method gives quite accurate estimate, with almost all estimates confined to one bin. In comparison the period estimates from the EMD based method show quite significant spread specially for higher noise levels.

3.3.2 Example 2 :: Effect of Coloured Noise

The presence of coloured noise in the oscillating signal makes the detection more difficult as discussed in Karra et al. (2010), Li et al. (2010). The coloured noise ACF is not confined to zero lag, and oscillation detection in the presence of colored noise therefore needs extra care. Therefore, this section studies the efficacy of the proposed method in identifying the presence of multiple oscillations in presence of coloured noise. The signal composition remains the same as in example 1 (section 3.3.1) with coloured noise replacing the white noise. The coloured noise is generated by passing a white noise of unit variance through a low pass filter $1/(1-0.7z^{-1})$. The simulations for 10000 different realizations of coloured noise are carried out and results are summarized in last row of Table 3.3. It can be seen that in this case again the proposed method performs better with success rate of 80% as compared to 66% of EMD based method in determining the oscillatory modes correctly. The proposed method also gives better estimates of the period of oscillation (Histogram Figure 3.8).

Table 3.3: No of successful iterations as %age of 10000 noise realizations

	Noise Type	Success rate(%)	
		EMD based Method	Proposed Method
Example 1	White $\sigma_v^2 = 1.30$	76%	97 %
	White $\sigma_v^2 = 3.0$	50%	96%
	White $\sigma_v^2 = 4.0$	39%	94%
Example 2	Coloured	66%	80%

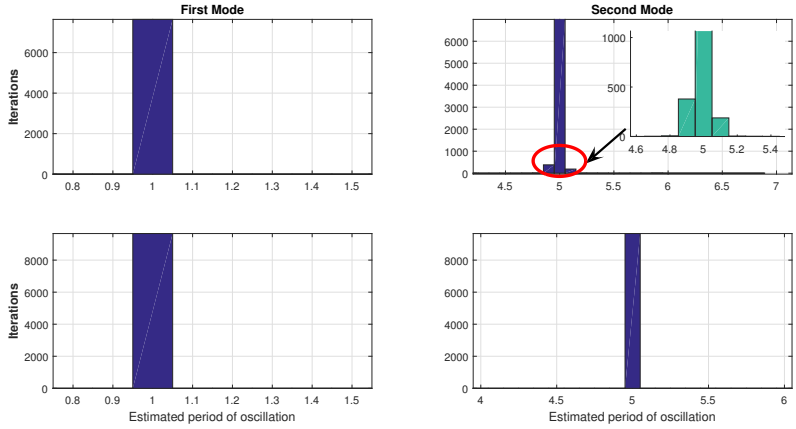

Figure 3.5: Histograms for estimates of period of oscillation, example 1 (white noise $\sigma^2 = 1.3$). First row (EMD based method), second row (proposed method).

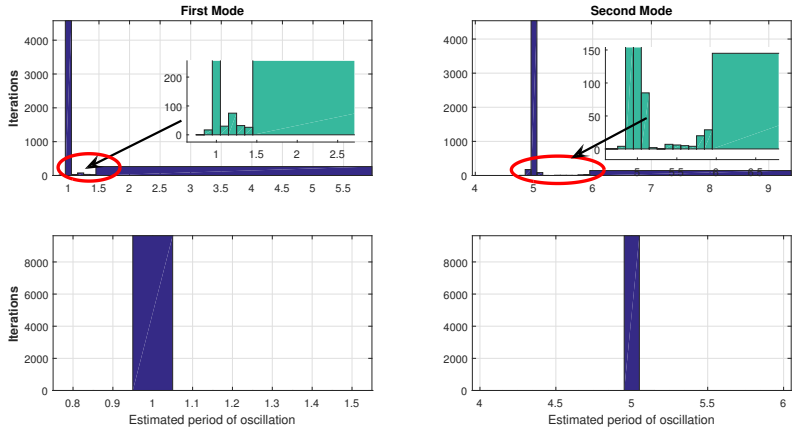

Figure 3.6: Histograms for estimates of period of oscillation, example 1 (white noise $\sigma^2 = 3$). First row (EMD based method), second row (proposed method).

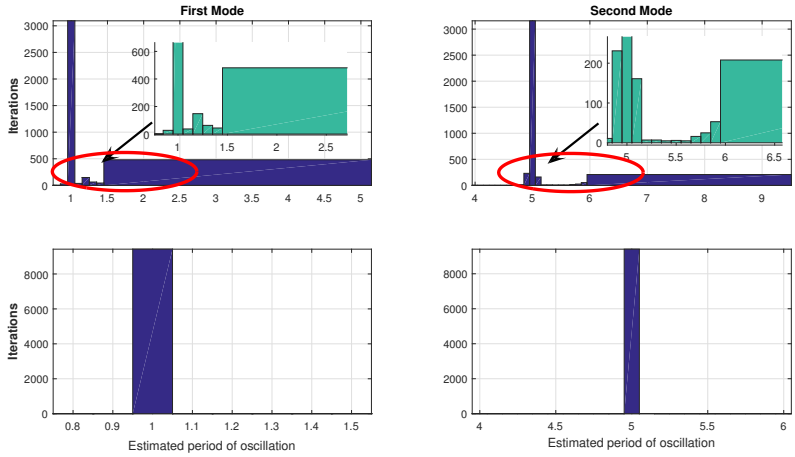

Figure 3.7: Histograms for the estimates of period of oscillation, example 1 (white noise $\sigma^2 = 4$). First row (EMD based method), second row (proposed method).

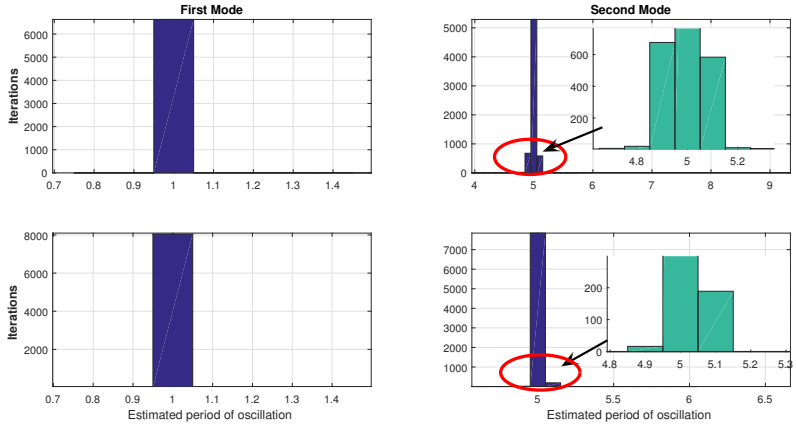

Figure 3.8: Histograms for the estimates of period of oscillation, example 2 (coloured noise). First row (EMD based method), second row (proposed method).

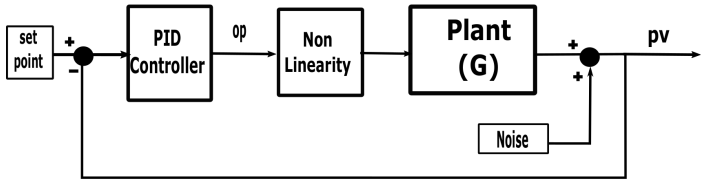

Figure 3.9: Closed loop system for stiction induced oscillations.

3.3.3 Example 3: Stiction Induced Oscillations

It is well known fact that the oscillations due to non-linearity are non-sinusoidal in nature and exhibit harmonics. This example is added to observe the performance of the proposed and existing oscillation detection method for such oscillations. The example is taken from our recent work Aftab et al. (2016) where stiction induced oscillations are generated in a SISO closed loop system (Figure 3.9) using the LuGre friction model (Olsson 1996). The system, controller and LuGre model parameters are the same as given in Aftab et al. (2016) and the noise variance is 0.2.

The oscillations are symmetric and therefore exhibit odd harmonics. The system response along with the IMFs from standard EMD and proposed methods are shown in Figure 3.10. It can be seen that the IMF given by the standard EMD shows significant leakage effects whereby the third harmonic leaked into the fundamental harmonic clearly indicating the case of mode mixing. In contrast, by virtue of the dyadic filter bank property of the NA-MEMD there is no such leakage effect for the IMF from NA-MEMD. The estimated period of oscillation for the identified fundamental harmonics is provided in Table 3.4 for both the methods. It can be seen that the proposed method gives more accurate estimate as the IMF obtained is free from mode mixing.

The reason for the absence of third and higher harmonics is the choice of the correlation coefficient selected to retain the significant IMFs (see section 3.8). In this case the correlation coefficient of third harmonic is less than the chosen threshold of 0.5 and is therefore excluded from the analysis. The main objective of presenting this example here is to highlight the advantages offered by the proposed method in case of non-sinusoidal oscillation. Moreover, the industrial case study sections contain examples where the proposed method highlights the presence of harmon-

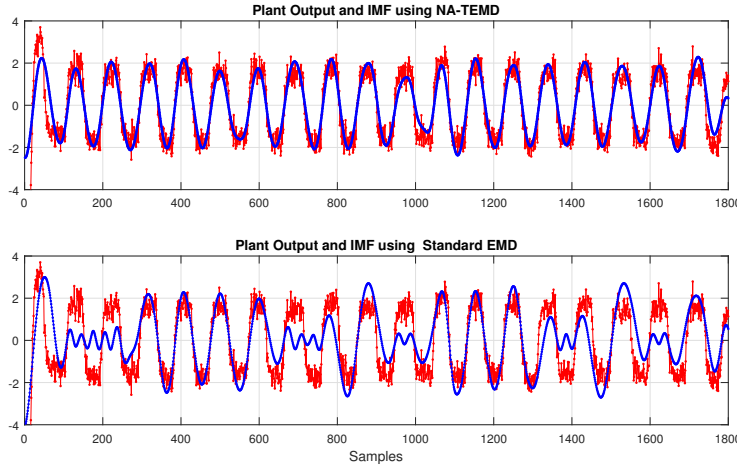

Figure 3.10: Stiction induced oscillations (Red:: Plant Output, Blue:: IMFs).

ics in case of non-linearity induced oscillations. The detailed discussion about the detection of harmonics using the NA-MEMD has been provided in chapter 5 of the thesis and Aftab et al. (2017c).

3.4 Industrial Case Studies

In this section the robustness and advantages of the proposed scheme are highlighted using the the three industrial case studies. The measurements have multiple oscillatory modes or frequencies and also contain non-linearity induced oscillations. Therefore, the case studies can effectively test the ability of the proposed scheme to detect multiple oscillations in a real plant data.

3.4.1 Case Study-I

The data from hydrogen reformer, also analysed in Thornhill et al. (2002), Tangirala et al. (2005), Aftab et al. (2017c), Karra et al. (2010), is used in this case study. Five different tags 6, 11, 14, 19 and 33 (Figure 3.11) show presence of multiple

Table 3.4: Oscillation characteristics for example-3 (stiction induced oscillations)

Method	IMF	r value	T_p	σ_{T_p}
EMD based Method	1	7.9	92.54	3.9
Proposed Method	1	30.8	93.27	1.0

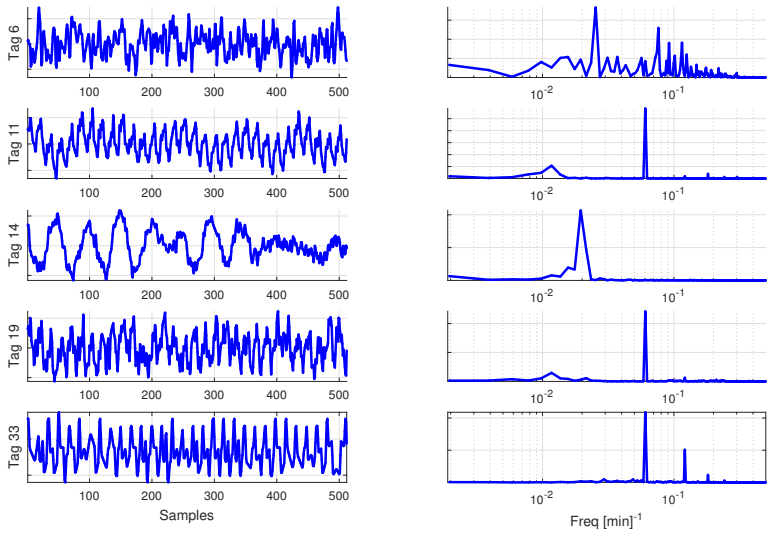

Figure 3.11: Time trends and spectra for case study-I.

oscillations that are hard to identify manually. The trends are processed by the existing and proposed oscillation detection methods. Though there are 37 tags in this case study and the proposed method works well with all of them, but all the tags are not included due to two reasons, *a*) only those tags are reported where the proposed method gives significant improvement over the EMD based methods. Most of the tags have only one oscillation frequency and is handled equally well by the EMD method, and *b*) the aim is to provide the proof of concept and therefore the results for all 37 tags are not reported to keep the discussion concise and to the point.

The results given in Table 3.5 show that the proposed method is more effective and accurate in detecting the presence of multiple oscillations and related characteristics. Moreover, the proposed method also highlights the presence of harmonics in case of oscillations generated due to non-linearity.

The proposed method is able to identify multiple oscillations in tag 6, whereas the existing EMD based method could only identify one oscillatory frequency. Similar results are reported for tag 11 where standard EMD based method fail to identify the slow frequency oscillation. In case of tags 14 and 19, the proposed method gives more reliable and accurate estimates of oscillations periods. As far as tag 33 is concerned the proposed method has correctly highlighted the presence of harmonics a characteristic of oscillations due to non-linearity. It has been reported

Table 3.5: Oscillation characteristics case study-I

Tag	Method	IMF	r value	$\bar{T}_p$ [min]	σ_{T_p} [min]	Oscillation
Tag 6	EMD	1	1.20	21.6	6.0	yes
	Proposed	1	2.51	17.6	2.3	yes
		2	5.1	39.8	2.6	yes
Tag 11	EMD	1	5.5	16.7	1.0	yes
	Proposed	1	5.9	16.5	0.93	yes
		2	2.13	83.2	13.00	yes
Tag 14	EMD	1	3.2	48.2	4.9	yes
	Proposed Method	1	7.9	49.8	2.1	yes
Tag 19	EMD	1	3.46	16.4	1.577	yes
		2	2.73	70.6	8.59	yes
	Proposed Method	1	6.48	16.4	0.84	yes
		2	5.74	82.20	4.70	yes
Tag 33	EMD	1	5.90	16.54	0.93	yes
	Proposed Method	1	1.95	5.45	0.934	yes (3^{rd}harmonic)
		2	4.52	8.18	0.60	yes (2^{nd}harmonic)
		3	5.90	16.54	0.93	yes

in number of studies (Aftab et al. 2017c, Thornhill 2005, Choudhury 2006) that the said oscillations are generated due to valve non-linearity, thus the proposed oscillation detection method has also confirmed this fact.

3.4.2 Comparison with Other Methods

This case study has been mostly used for the plant-wide oscillation detection and root cause analysis. Karra et al. (2010) has used the same data set for the oscillation detection problem. The results reported in Karra et al. (2010) did not identify the multiple oscillations in different tags. Moreover, the periods of the single frequencies reported in Karra and Karim (2009) are also less accurate as compared to the method proposed in this paper. The work by Karra et al. (2010) requires additional filtering to isolate multiple oscillations and noise effects and that requires the frequency known to be in advance.

3.4.3 Case Study-II

The data from four control loops, represented by tags 8, 9, 11 and 12 (time trends and spectra given in Figure 3.12), from the challenge problem (Thornhill et al.

Table 3.6: Oscillation characteristics case study-II

Tag	Method	IMF	r value	$\bar{T}_p$ [min]	σ_{T_p} [min]	Oscillation
Tag-8	EMD	1	r<1			no
	Proposed	1	13.3	517.3	12.94	yes
Tag-9	EMD	1	5.29	514.33	32.4	yes
	Proposed	1	76.98	253	1.09	yes (harmonic)
		2	45.71	510.6	3.72	yes
Tag-11	EMD	1	3.0	23.27	2.5	yes
		2	8.2	51.0	2.07	yes
	Proposed	1	4.08	22.9	1.86	yes
		2	16.9	51.27	1.0	yes
Tag-12	EMD	1	4.0	572.3	46.7	yes
	Proposed	1	23.2	501.2	7.24	yes

(2002)) are also analyzed. The results for both the schemes, the proposed and existing one, are given in Table 3.6.

The time trend and spectrum of tag 8 shows a slow oscillation with period around 500 mins. The existing EMD based method cannot capture this oscillation ($r < 1$). In contrast the proposed method is able to detect this oscillation in the signal. Moreover, the proposed method has correctly pointed out the presence of harmonics in tag 9, thereby confirming the non-linearity as the source of oscillation. EMD based method only identified the fundamental harmonic. The oscillation statistics reported by proposed method are also more accurate than the standard EMD based method.

Similarly for the multiple oscillations in tag 11, the 'r' statistic and σ_{T_p} shows that the proposed method gives more accurate estimates of related characteristics as compared to the existing method. Same is the case with the slow oscillation in tag 12.

3.4.4 Case Study-III

The data for three control loops, namely tag 1,2 and 3 from an industrial paper plant has been analysed for the presence of oscillations. The time trends and the spectra are shown in Figure 3.13. The results presented in Table 3.7 clearly indicates the advantages of the proposed method.

The proposed method identifies three oscillation frequencies in tag 1 but the stand-

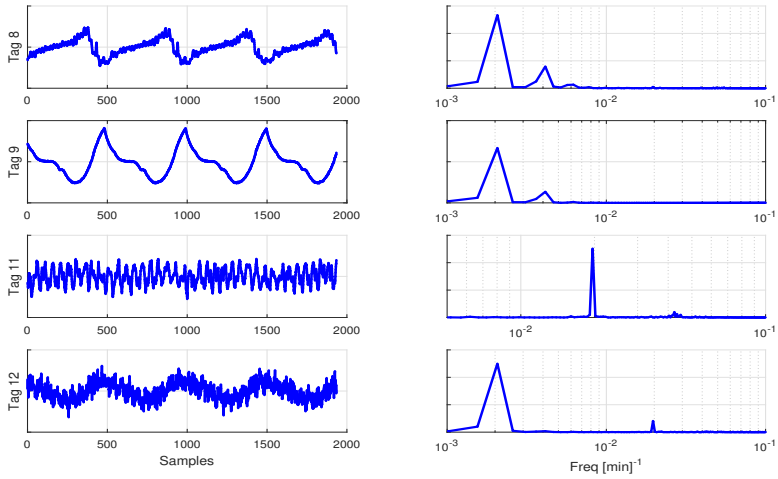

Figure 3.12: Time trends and spectra for case study-II.

ard EMD approach can only identify two of them. Similarly for tag 2 the standard EMD can identify only one oscillation whereas the proposed method indicates the presence of multiple (three) oscillations. Moreover, the noise assisted technique has correctly identified the presence of single oscillation in tag 3 that the standard EMD method could not detect.

Table 3.7: Oscillation characteristics case study-III

Tag	Method	IMF	r value	$\bar{T}_p$ [min]	σ_{T_p} [min]	Oscillation
Tag 1	EMD	1	3.98	5.15	0.43	yes
		2	1.6	19.7	4.0	yes
	Proposed	1	4.8	5.03	0.34	yes
		2	1.87	8.24	1.47	yes
		3	3.37	18.85	1.86	yes
Tag 2	EMD	1	1.0	6.72	2.24	yes
	Proposed	1	1.38	3.51	0.84	yes
		2	2.0	6.67	1.07	yes
		3	3.28	8.85	0.90	yes
Tag 3	EMD	< 1	–	–	–	no
	Proposed	1	4.27	11.78	0.91	yes

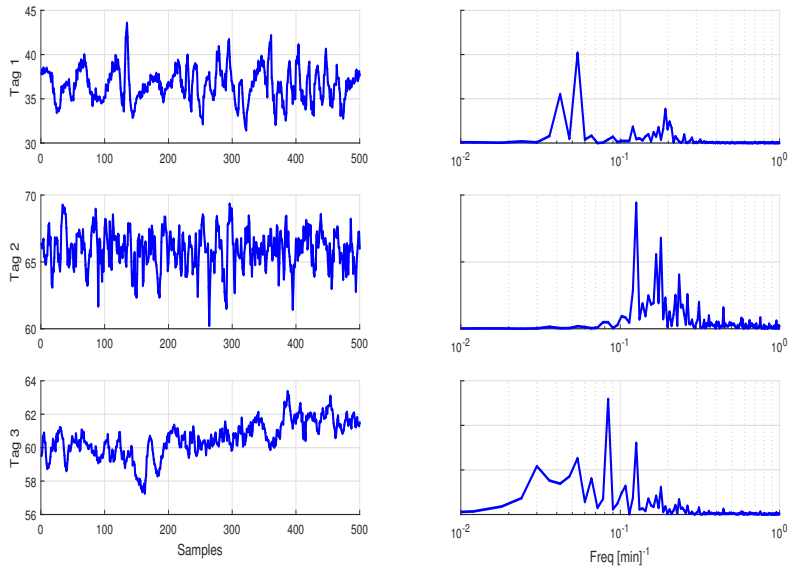

Figure 3.13: Time trends and spectra for case study-III.

3.5 Conclusions

An improved and robust oscillation detection method based on Noise-Assisted Multivariate EMD (NA-MEMD) is presented. It is shown with the help of both simulations and industrial case studies that the proposed method is less prone to mode mixing problem, and is more reliable and accurate in determining the presence of multiple oscillations and related oscillation characteristics as compared to the standard EMD based approach. The proposed method can handle the non-stationary effects and can highlight the presence of harmonics in non-linearity induced oscillations as well.

Chapter 4

Plant-Wide Oscillation Detection

Paper B: Plantwide Oscillation Detection using Multivariate Empirical Mode Decomposition

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

Computers and Chemical Engineering (Revised Submitted 2018)

Abstract

Plant-wide oscillation detection is an important task in the maintenance of large-scale industrial control systems, owing to the fact that in an interactive multi-loop environment oscillation generated in one loop may propagate to the different parts of the plant. In such a scenario, it is required that different loops oscillating due to a common cause and hence similar frequency may be grouped together. In this paper an adaptive method for plant-wide oscillation detection based on multivariate empirical mode decomposition (MEMD) along with a grouping algorithm is proposed. The method can identify multiple oscillation groups among different variables as well as variables with random noise only. The proposed method is also applicable to both non-linear and non-stationary time series where the techniques based on the conventional Fourier analysis are prone to errors. Within each group that oscillate due to a common cause, the method can also indicate the location of the probable root cause of oscillations. The efficacy of the proposed method is established with the help of both simulation and industrial case studies.

4.1 Introduction

The observation that only quarter of all industrial controllers give acceptable performance (Starr et al. (2016), Desborough and Miller (2002)) led to increased re-

search in the field of oscillation detection and root cause diagnosis. The topic is important because oscillations give rise to product variability and greatly impact the profitability of plant operations. Sustained oscillations are also harmful for plant equipment, leading to increased wear and possibly premature failure. There can be thousands of loops running in a plant, and oscillations in any of them can propagate to different parts of the plant owing to underlying interactions and coupling effects. It is therefore important to find and group the variables that oscillate due to a common cause. This will be very helpful in finding a root cause of a specific plant-wide oscillation, as there is little reason for searching for a root cause among variables not affected by the oscillation in question.

Oscillations can be caused due to a variety of reasons like poor controller tuning, external disturbances and process non-linearities, to name a few (Choudhury et al. 2008). Oscillation detection can be subdivided into two categories, namely oscillation detection in individual loops and plant-wide oscillation detection (Li et al. 2010). The present work deals with the latter subcategory only. The presence of multiple oscillations, noisy measurements along with non-stationary and non-linear effects make plant-wide oscillation detection a challenging problem. The salient features of a reliable and robust plant-wide oscillation detection method, as described in literature (Thornhill and Horch 2007), can be outlined as:

- The ability to group different variables oscillating with similar frequency.
- The capacity to detect multiple oscillations (oscillations due to multiple sources) in a variable and to assign each oscillation to corresponding group. This also includes the ability to separate out variables void of any oscillation, thus containing random noise only.
- Adaptability, i.e., the minimum possible dependence on any *a priori* assumptions, pre-processing or underlying process dynamics.
- Ability to identify the set of variables/nodes that are closest to the source of the oscillations.

The above mentioned points do provide solid set of guidelines for detection and diagnosis of plant-wide oscillations. Nearly all existing methods have tried to address these issues one way or the other, but an important piece still missing in the puzzle is handling of non-stationary effects. Therefore, in addition to the above mentioned features the effective plant-wide oscillation method should also be able to address the non-stationary effects such as time varying disturbances and changing set points. Therefore, it is desirable that the plant-wide oscillation detection

method also contain the following property in addition to the ones described above, i.e.

- Applicability to signals with both non-stationary and non-linear effects.

The method proposed in this work does provide this added feature and before going into the details a brief overview of the existing plant-wide oscillation detection methods has been provided. Different attempts have been made in recent years by researchers to address the plant-wide oscillation detection problem. The primary focus has been on the frequency domain analysis as it is insensitive to phase lags.

The use of principal component analysis (PCA) for the detection of plant-wide oscillations is proposed by Thornhill et al. (2002). In this method, normalized power spectra of different time trends are subjected to dimensionality reduction using PCA and each principal component represents a group oscillating with similar characteristics. Apart from the power spectrum, PCA of the Auto Covariance Function (ACF), which is independent of phase lags, is also shown to give similar performance (Thornhill et al. 2002).

Another method based on similarity among the power spectra of different variables is put forward by Tangirala et al. (2005). In this method a correlation index among the power spectra is used to construct the power spectral correlation map (PSCMAP). Xia et al. (2005) and Xia and Howell (2005) proposed a spectral independent component analysis (ICA) based method where dimensionality reduction and grouping is accomplished by employing ICA, but with a condition that sources of oscillation are statistically independent.

The limitations of the PCA and ICA based method are highlighted in Tangirala et al. (2007) and another method based on Non-Negative Matrix factorization (NMF) is proposed instead that is aimed at overcoming these limitations. In an analogy to the singular value decomposition performed in standard PCA, the proposed method uses the pseudo singular value decomposition (PSVD) to determine the size of basis space. The main issues in NMF as highlighted in Tangirala et al. (2007) are initialization of basis functions and the absence of any solid method for basis size determination, though singular value decomposition (SVD) of the spectral matrix is used as a tool for the basis initialization in the work by Tangirala et al. (2007). The NMF based method is powerful tool yet the emphasis is more on determining the basis shape of the linearly independent oscillations in a plant-wide setting and at times it fails to group the variables oscillating with similar frequencies. This aspect will be illustrated by the simulation example later in the paper. Moreover, a GA based matrix factorisation technique has been adopted by El-Ferik et al. (2012). A method based on intrinsic time-scale decomposition has

been proposed recently by Lang et al. (2018).

Almost all the methods used for the plant-wide oscillation detection, except Lang et al. (2018), impose the condition of stationarity on the data. Lang et al. (2018) has proposed the method that can handle non-stationary effects but it still has not been fully automated as all the individual signals need to be checked separately for the presence of oscillations. In order to overcome these problems a fully adaptive and automated plant-wide oscillation detection method based on Multivariate Empirical Mode Decomposition (MEMD) has been proposed in this work.

The MEMD is the extension of the standard univariate Empirical Mode Decomposition (EMD) to higher dimensions. The MEMD like its predecessor EMD is applicable to both non-linear and non-stationary time series and can adaptively decompose the n-dimensional signal, consisting of n recorded variables, into functions called Intrinsic Mode Functions (IMFs), without any *a priori* assumptions or knowledge about the underlying process. MEMD is finding its ways into areas like data fusion, remote sensing, bio-medical engineering and communication engineering, owing to its particular properties and adaptive nature.

Though MEMD has been around for a while yet it has not been explored for the controller performance monitoring and detection and diagnosis of plant-wide oscillations. The novelty of the current work stems from the fact that the MEMD, along-with fully automated grouping algorithm, has been used first time for the detection of plant-wide oscillations. An important consequence of MEMD is its mode alignment property where common oscillations in data are extracted in the same indexed IMFs. This mode alignment property along with a grouping algorithm is exploited to give a robust plant-wide oscillation detection method in this work. Another advantage is the ease of calculating the period of oscillation for each group of oscillations that is not available in existing plant-wide oscillation detection methods (details in section 4.6). Moreover, the proposed method does not require any pre-processing or assumptions regarding the signal generating process and is both applicable to non-linear and non-stationary time series. The method can handle multiple oscillations and group the variables oscillating due to the same cause. Furthermore, the technique is also helpful in identifying the variables associated with the root cause of oscillations.

The paper is organized as follows. Section 4.2 gives an overview of the EMD and its extension to higher dimensions. The mode alignment property and proposed grouping algorithm are discussed in sections 4.3 and 4.4 respectively. Section 4.5 and 4.6 outlines the default parameter settings and the proposed algorithm. Simulation examples are discussed in section 4.7 followed by industrial case studies (section 4.8) and conclusions.

4.2 Overview: EMD and its Extensions

4.2.1 Empirical Mode Decomposition (EMD)

Empirical mode decomposition, a recently developed adaptive signal processing technique, aims at decomposing the input signal into intrinsic mode functions or IMFs $c_i(t)$ and a bias term or residue $b(t)$.

$$x(t) = \sum_{i=1}^{N} c_i(t) + b(t) \tag{4.1}$$

where $x(t)$ is the input time series and N is the total number of IMFs. An IMF is defined as a signal having zero mean where the number of extrema and zero crossing must either be equal or at most differ by one (Huang et al. 1998). The EMD is an iterative process that extracts fast or high frequency modes or oscillations from the slower ones. The sifting process is initiated by identifying the extrema and calculating the local mean $m(t)$ from the envelope defined by spline fitting of the extrema. This local mean is then subtracted from the signal to get a local fast mode or detail $d(t)$ (Rilling et al. 2003), that can be written mathematically as

$$d(t) = x(t) - m(t) \tag{4.2}$$

The same procedure is repeated on d till the time it is an IMF $c_i(t)$. The first IMF gives the highest frequency and to get other frequencies this IMF is subtracted from original signal and the sifting process starts again on the residue $h_i(t)$, given by

$$h_i(t) = x(t) - c_i(t) \tag{4.3}$$

This iteration goes on until there are no more IMFs left to be extracted. The residue left behind after extraction of last the IMF is the final residue $b(t)$. The detailed description of the procedure can be found in Huang et al. (1998) and Rilling et al. (2003).

4.2.2 Multivariate EMD (MEMD)

The standard EMD caters for univariate signals (signals consisting of one variable) only. Therefore, in order to process signals of higher dimensions (signals with more than one variables) extended versions of the standard EMD termed as Bivariate EMD (Rilling et al. (2007)), Trivariate EMD (Rehman and Mandic (2010b)) and Multivariate EMD (Rehman and Mandic (2010a)) have been developed. Bivariate and Trivariate EMD deals with two and three dimensional signals respectively, whereas MEMD is generalization of the same to n dimensions. Here the two, three and n-dimensional signals means a signal comprising of

two, three and n variables respectively. A multivariate signal is a signal consisting of more than one variables. For instance a multivariate signal $\mathbf{X}$ consisting of n variables $x_1 \dots x_n$ with each variable having l samples is given by

$$\mathbf{X} = \begin{Bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(l) & x_2(l) & \dots & x_n(l) \end{Bmatrix} \tag{4.4}$$

The standard EMD forms a spline envelope through the extrema of the signal and calculate the mean to extract different IMFs (sees section 4.2.1). The primary challenge in extending the univariate EMD to the multivariate case is the calculation of such an envelope and the local mean in higher dimensions.

Rilling et al. (2007) proposed that, for the bivariate case (two variables), the envelope can be constructed by projecting the signal in multiple directions in the complex plane and by interpolating the extrema of projected signals via spline fitting. The multiple directions are formed by generating unit vectors spanning the complex plane. The same concept is extended to three and general n dimensions by considering uniformly sampled points on the unit three and n-dimensional spheres respectively. The multiple directions are then represented by the direction vectors from the centre of the n-dimensional sphere to these points on the surface. The uniformly spaced points are generated via Halton sequence as discussed in Rehman and Mandic (2010a) and Aftab et al. (2017c). A short mathematical description is also provided here.

The k^{th} sample of a one dimensional Halton sequence $\Phi_p(k)$ is written as

$$\Phi_p(k) = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} \dots \frac{a_r}{p^{r+1}} \tag{4.5}$$

where prime base-p representation of k can be expressed as (Wong et al. 1997)

$$k = a_0 + a_1 p + a_2 p^2 \dots a_r p^r \tag{4.6}$$

where each a_i is an integer in $[0, p-1]$. Starting with $p_1, p_2 \dots p_q$ as q prime numbers, the kth sample of the q-dimensional, Halton sequence then becomes

$$\left(\Phi_{p_1}, \Phi_{p_2}, \dots \Phi_{p_q}\right) \tag{4.7}$$

Thus the points formed using the equation 4.7 are uniformly distributed and can used for the generating direction vectors.

The uniform distribution of the direction vectors generated by Halton sequence deteriorate rapidly with the increase in the number of dimensions (Niederreiter 1992). In order to cater for this problem the Halton sequences generated by the reverse radix based algorithm, proposed by Kocis and Whiten (1997), are used in this work that gives fairly uniform distribution for higher dimensions

The MEMD method to decompose multivariate input signal $\mathbf{X}$ (equation 4.4) into IMFs is summarized below (Rehman and Mandic 2010a).

Step I Set up K direction vectors u^k with $k = 1 \dots K$ by choosing uniformly distributed points on the n dimensional sphere.

Step II Find the projections $p^k(t)$ of the input signal $\mathbf{X}$ along the direction vectors u^k using

$$p^k(t) = \mathbf{X}u^k \quad \forall \quad k = 1 \dots K$$

Step III Identify the maxima of projections $p^k(t)$ and corresponding time instants t^k.

Step IV Generate n-dimensional envelope curve $\mathbf{e}^k(t)$ by interpolating $[t^k, \mathbf{X}(t^k)]$.

Step V The n-dimensional mean of envelope curve is then given by

$$\mathbf{m} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{e}^k(t) \tag{4.8}$$

Step VI Extract n-dimensional detail $\mathbf{d(t)}$ using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$.

Step VII Repeat the steps I-VI using $\mathbf{d(t)}$ as input till it fulfils the criteria for an IMF.

Step VIII Calculate the residue $\mathbf{b(t)} = \mathbf{X(t)} - \mathbf{d(t)}$ and iterate the procedure on $\mathbf{b(t)}$ till there are no more IMFs left to be extracted.

4.3 Mode Alignment Property of MEMD

An important consequence of MEMD is its mode alignment property. By virtue of this property the different common oscillating frequencies/modes in the multivariate signal are aligned in the same indexed IMFs (Rehman and Mandic 2010a). This can be explained by the use of following illustrative example.

Table 4.1: Signal composition (mode alignment example)

Variables ↓	f_1	f_2	f_3	f_4	Phase angle	Bias	Noise variance
A	✓	–	✓	✓	$\angle f_4 = \pi/2$	+ve	0.2
B	–	✓	✓	–	$\angle f_3 = -\pi/2$	-ve	0.2
C	✓		✓	–	$\angle f_1 = \pi/4$	–	0.2
D	✓	✓	–	✓	–	–	1.0

4.3.1 Mode Alignment :: Illustrative Example

In order to accentuate the mode alignment property, we consider a multivariate signal consisting of four variables A, B, C and D containing four oscillatory frequencies/modes (f_1, f_2, f_3, f_4). The signals are generated in such a way that they have some oscillations common among them. In order to illustrate the robustness of the MEMD, different initial phase angles are also incorporated. A time varying bias ($b = 0.1t$) is also added in channels A and B to make the signals non-stationary. The signals composition, variance of added noise and the associated phase angles are given in Table 4.1 (only non-zero phase angles are mentioned). The D variable also contains significantly higher noise level in addition to the frequencies f_1, f_2 and f_3.

The objective is to analyse this signal comprising four variables using MEMD and to observe whether it can align common frequencies in the same indexed IMFs. The signals and the resultant IMFs from MEMD process are shown in Figure 4.1.

In total nine (9) IMFs and a residual are extracted. The noise content has been extracted in first three IMFs corresponding to D variable. It can be seen in Figure 4.1 that oscillations common between A, B and C with frequency f_3 are present in the fifth IMF of A, B and C, whereas the f_2 frequency, common between B and D, is extracted in the sixth IMF. Similarly, f_1 is present in the eighth IMFs of A, C and D. The frequency f_4, present in A and D, being the highest frequency oscillation is extracted in the fourth IMF of A and D variables, i.e., the first IMF after the noise content. The non-stationary terms arising due to the time varying bias are extracted in residue of A and B. Thus, this example illustrates the ability of the MEMD to align the common oscillations in different signals that can be utilized for the plant-wide oscillation detection.

4.4 Grouping Algorithm

It has been shown that MEMD is capable of aligning common oscillations in multivariate signal in same indexed IMFs, but in order to have automatic plant-wide detection and grouping of different oscillations based on MEMD, some post pro-

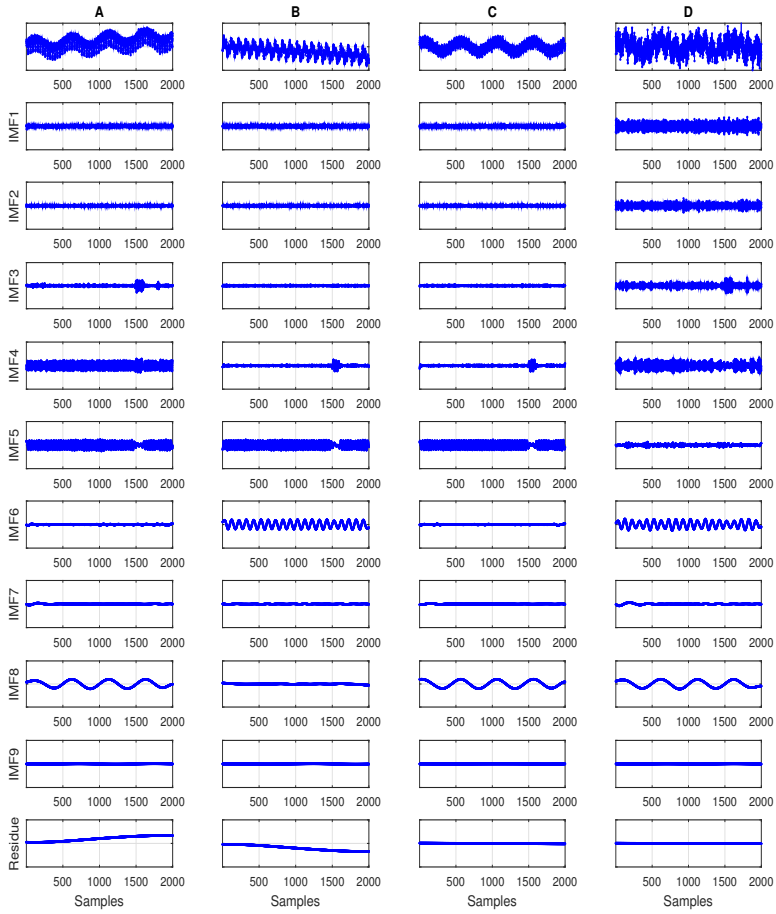

Figure 4.1: Mode alignment illustrative example.

cessing is required. In this section, we propose a grouping algorithm that is used for automatic plant-wide oscillation detection.The major steps involved in grouping algorithm are explained next.

4.4.1 Normalized Correlation Coefficient Matrix

The first step in the algorithm is the calculation of a correlation coefficient matrix Λ. This is required because the MEMD, like the standard EMD process, may also generate spurious IMFs, due to spline fitting issues as highlighted in Peng

et al. (2005), Aftab et al. (2016), Aftab et al. (2017c), Srinivasan and Rengaswamy (2012) and leakage effects where one principal mode may leak into other IMFs. As the IMFs so generated are nearly orthogonal, these pseudo-IMFs will be poorly correlated with the original signal.

Therefore the correlation coefficient can be used to assess the significance of IMFs as reported in Peng et al. (2005), Srinivasan and Rengaswamy (2012) and Aftab et al. (2016). For the j^{th} variable of the multivariate signal $x_j(t)$, the correlation coefficient ρ_{ij} with corresponding i^{th} IMF is calculated from

$$\rho_{ij} = \frac{Cov(c_{i_j}, x_j)}{\sigma_{x_j}\sigma_{c_{i_j}}}, \quad i = 1, 2, 3 \ldots N \tag{4.9}$$

where Cov is the covariance, $c_{ij}(t)$ is the i^{th} IMF corresponding to j^{th} variable and $x_j(t)$ is the j^{th} variable in input signal; σ_{x_j} and $\sigma_{c_{i_j}}$ are the standard deviations of the j^{th} variable and corresponding IMF i, respectively. N is the total number of IMFs and J is the total number of variables in the multivariate signal.

The normalized correlation coefficient λ_{ij}, for each variable j, is calculated for all N IMFs using

$$\lambda_{ij} = \frac{\rho_{ij}}{max_i(\rho_{ij})}, \quad i = 1, 2, 3 \ldots N \tag{4.10}$$

The matrix Λ is then constructed from the individual λ_{ij}, in such a way that each row contains the correlation coefficient of same indexed IMF with j^{th} variable, thus giving N by J matrix of correlation coefficients.

$$\Lambda = \begin{Bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1(J-1)} & \lambda_{1J} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2(J-1)} & \lambda_{2J} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{(N-1)1} & \lambda_{(N-1)2} & \cdots & \lambda_{(N-1)(J-1)} & \lambda_{(N-1)J} \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{N(J-1)} & \lambda_{NJ} \end{Bmatrix} \tag{4.11}$$

4.4.2 Grouping Algorithm

Although owing to the mode alignment property of MEMD all the same indexed IMFs (each row in matrix Λ) represent similar oscillations, yet due to presence of pseudo IMFs (as highlighted in previous section), a grouping algorithm is required, so that only the dominant oscillations in significant IMFs shall be considered for plant-wide oscillation detection.

Thus the proposed grouping technique searches every row (same indexed IMFs for all J variables) for all the correlation coefficients greater than certain threshold, say

η. The variables in a row, that are greater than the threshold, are grouped together representing same oscillation. In this way all rows are searched one after another to group common oscillations across all variables. The grouping is accomplished using the procedure outlined in Table 4.2 and is applied on the simulation example in section 4.7.

In case of multiple oscillations, one j^{th} variables may reside in more than one group, thereby giving the insight into the different oscillation frequencies present in each variable.

4.4.3 Finding Noisy IMFs and Regrouping

Noisy IMFs may have significant correlation with the input signal, in particular for signals comprising of random noise only or with higher noise content. In order to find the IMFs (identified in the previous step), void of any oscillatory mode, the method proposed by Hoyer (2004) and used by Srinivasan and Rengaswamy (2012) is considered here.

The process called sparseness index (SI) calculation uses the fact that the noise signal will have power in a broad frequency range, whereas the oscillatory signal will have dominant magnitude at only one frequency. So the sparseness index of the signal x given by (4.12) will be approximately zero for the noise signal and will be nearly one for purely oscillatory signals.

$$SI(x) = \frac{\sqrt{I} - \left(\sum_{i=1}^{I} |X_i| / \sqrt{\sum_{i=1}^{I} |X_i|^2} \right)}{\sqrt{I} - 1} \tag{4.12}$$

where X is the frequency spectrum of times series $x(k)$ and I is the number of frequency bins. Spectrum is computed without using any window.

The SI of the IMFs grouped in each Group G_k is calculated from their frequency

Table 4.2: Grouping algorithm

1. For each row i of matrix Λ, find variables j with $\lambda_{ij} \geq \eta$.
2. Assign the variables with $\lambda_{ij} \geq \eta$ to group G_k.
3. The j variables within the same G_k represent oscillation group with similar frequency.

spectrum and only those IMFs are retained that have sparseness near to 1 or greater than some threshold γ, while rest are shifted to group G_0 representing noise.

4.5 Default Parameter Settings

The grouping algorithm described in the previous section relies on tuning of parameters like number of direction vectors for MEMD and thresholds for the correlation coefficient and sparseness index. This section briefly outlines the motivation for choosing the default settings of these parameters. The settings are also summarized in Table 4.3.

4.5.1 Number of Direction Vectors

The multidimensional envelope for MEMD process is generated using the signal projection in different directions. The number of direction vectors shall be sufficient enough so that accurate envelopes can be generated and multivariate sifting process yields consistent results. Too high number gives rise to unnecessary computational overload without improving the results much.

It has been observed that results are quite consistent if the number of direction vectors is about four times the number of measurements/variables in question. Increasing the number of direction vectors further does not effect the results much. Therefore, the default setting for number of direction vectors used in this work is four times the number of measurements used in the analysis.

4.5.2 Correlation Coefficient (η)

The grouping algorithm requires the definition of the threshold η for the normalized correlation coefficient. In the work by Srinivasan and Rengaswamy (2012) the threshold used to distinguish the dominant IMFs from the pseudo-IMFs is taken as 0.5. The same threshold is used in this work as it ensures that only dominant oscillations are captured. The same threshold is also used in our recent work Andersson et al. (2017). This threshold can be lowered but this will increase the risk of getting spurious groupings.

4.5.3 Sparseness Index Threshold (γ)

Sparseness index is used to distinguish between IMFs with significant noise content and ones with pure oscillations. In order to define the threshold the spectra of the four signals from the industrial data along with SI are shown in Figure 4.2. It can be seen that the signals with SI up-to 0.54 can have a energy spread across broad range of frequency and may not exhibit distinct frequency oscillation. The signal with SI$= 0.6$ contains distinct harmonics due to non-linearity induced oscillations. As the objective is to group variables having distinct oscillation the

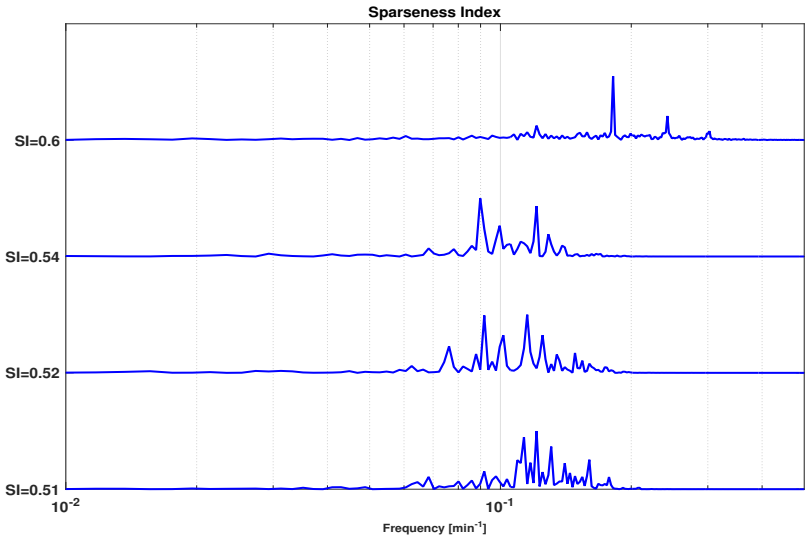

Figure 4.2: Frequency spectra of noise-dominated signals with sparseness index 0.6 top curve, 0.54 second curve, 0.52 third curve and 0.51 bottom curve.

threshold used in this work is taken to be 0.58 (same threshold used in previous work Aftab et al. (2017c)). The lower value is not desirable as only the dominant and distinct oscillations need to be considered. The threshold can be increased but it may risk some boundary line cases to be excluded.

Table 4.3: Default parameters setting

Parameter	Default value
Direction Vectors	4 x number of measurements
Correlation Threshold (η)	0.5
Sparseness Threshold (γ)	0.58

4.6 Proposed Algorithm

The proposed algorithm, for plant-wide oscillation detection can be divided into following steps.

Step I Process the multivariate input signal using MEMD.

Step II Calculate the correlation coefficient ρ_{ij} of each i^{th} IMF with corresponding input variable, i.e., j^{th} variable of multi-variable input as given

in equation (4.9).

Step III Calculate corresponding normalized correlation coefficient λ_{ij} using relation in equation (4.10).

Step IV Construct the Correlation Coefficient Matrix Λ according to equation (4.11).

Step V Search each row of Λ to group similar oscillations using the grouping algorithm given in section 4.4.2.

Step VI Calculate the sparseness index of all selected IMFs per criteria given in section 4.4.3.

Step VII IMFs with sparseness less than the given threshold are transferred to 0^{th} group G_0 .

Step VIII If transferring of IMFs as per step VII makes any group empty then re-number the group index k (will be explained in simulation example).

Step IX Find the characteristic power spectrum of each group using the member with maximum sparseness.

Step X Calculate the frequency/period of oscillations for each group using IMF with maximum sparseness via ACF method (Thornhill et al. 2003) .

Step XI In case of two or more groups show same frequency, the variables are merged into one group.

4.7 Simulation Example

In this section the proposed algorithm is applied on the simulation example consisting of multivariate signal comprising four variables (same signal as discussed in Section (4.3.1)). The MEMD process generated 9 IMFs and a residue. The normalized correlation coefficient matrix (Λ) is given in Table 4.4.

The elements of Λ that fulfil the grouping criteria in each row are highlighted. The index of the row is same as the index of IMFs.

4.7.1 Initial Grouping

The initial grouping (shown in Table 4.5) is carried out (using correlation coefficients in Table 4.4) as follows:

Table 4.4: Normalized correlation coefficient matrix (simulation example)

↓ IMF / Variable →	A	B	C	D
1	0.18	0.20	0.16	0.93
2	0.13	0.14	0.14	0.69
3	0.36	0.09	0.12	0.70
4	1.00	0.32	0.34	1.00
5	0.95	0.96	0.97	0.32
6	0.12	1.00	0.10	0.95
7	0.15	0.08	0.03	0.12
8	0.86	0.47	1.00	0.96
9	0.19	0.19	0.11	0.23
Residue	0.63	0.80	0.16	0.24

1. Correlation coefficient of first IMFs (first row of Table 4.4) corresponding to all four variables are analysed. It can be seen that only IMF that have correlation greater than threshold belongs to the variable D. Thus variable D is assigned to first Group, i.e., G_1 (see Table 4.5).

2. Similarly the same procedure as discussed in step 1 is repeated for second and third IMFs respectively. Here again the IMFs belonging to variable D fulfil the criterion and hence variable D is assigned to the groups G_2 and G_3.

3. In case of fourth row, IMFs of both variables A and D have correlation greater than the threshold. This shows that an oscillation is common between these two variables. Therefore, both the variables are assigned to fourth group, i.e., G_4.

4. Similar, analysis for fifth row shows that variables A, B, and C share this oscillation. Thus A, B and C are placed in group G_5.

5. Variables B and D have common oscillation extracted in sixth IMFs and these variables are assigned to group G_6.

6. Seventh row has no IMF that fulfils the required criterion, so it is skipped. No group is populated in the Table 4.5 and group number is not incremented.

7. Common oscillations among variables A, C and D (highlighted IMFs) are extracted in eighth IMF and seventh group G_7 is populated with these variables.

Table 4.5: Initial grouping for the simulation example

↓ Initial Groups /Variables →	A	B	C	D
G_1				✓
G_2				✓
G_3				✓
G_4	✓			✓
G_5	✓	✓	✓	
G_6		✓		✓
G_7	✓	✓		✓
G_8	✓	✓		

8. No element is highlighted in ninth row and it is skipped just like the seventh row.

9. The tenth row has two IMFs belonging to variables A and B that are highlighted. These IMFs represent the non-stationary term and are assigned to the group G_8.

Therefore initially eight different types of groups are formed. The IMFs included in each groups are analysed to form the final grouping.

4.7.2 Final Grouping

The final grouping is carried out in the following steps.

1. The Sparseness indices (SI) of all the IMFs highlighted in Table 4.4 are calculated. The IMFs with SI less than threshold γ represent noise and are therefore shifted to NULL group G_0. The variables included in G_0 contain significant noise level.

2. In case of present example the first three IMFs from variable D, forming the groups $G_1 - G_3$ have SI less than the threshold so they are transferred to NULL group. Now three groups $G_1 - G_3$ have become vacant.

3. Lastly the remaining five groups are re-numbered as $G_1 - G_5$.

4. The oscillation frequency of each group $G_1 - G_4$ can be calculated using steps IX-X outlined in section 4.6.

Table 4.6: Final grouping for the simulation example

↓ Final Groups / Variables →	A	B	C	D
G_0				✓
G_1	✓			✓
G_2	✓	✓	✓	
G_3		✓		✓
G_4	✓	✓		✓
G_5	✓	✓		

4.7.3 Analysing the Groups

Analysis of final grouping in Table 4.6 shows that D variable is member of G_0, G_1, G_2 and G_4 thereby confirming that it contains multiple oscillations as well as significant noise content.

Groups $G_1, G_2 \ldots G_4$ represent oscillations with frequencies $f_4, f_3 \ldots f_1$ respectively with highest frequency residing in the first group (as the frequency of IMF decreases with IMF index). The results show that oscillation frequency f_4 (group G_1) is present in variables A and D whereas f_3 is common among A, B and C and is represented by G_2. Frequency f_2 is common between B and D (group G_3) and the slowest f_1 is present in variables A, B and D (G_4). G_5 represents the non stationary term that is part of A and B variables only.

Thus the proposed grouping algorithm has correctly deciphered the simulated multivariate signal and highlighted the common oscillations among different variables in a fully automated way.

4.7.4 Simulation Example with NMF Method

The same signals are analysed with NMF method for comparison purposes (with non-stationary bias terms removed from the data for this analysis). The spectral decomposition methods, like principal component analysis (PCA), independent component analysis (ICA) and NMF, work on the matrix of power spectra $P(f)$, where rows of $P(f)$ are formed by power spectrum of m signals. The spectra are then decomposed as a sum of p basis functions $w_1^{'} \ldots w_p^{'}$ (Thornhill and Horch 2007) given by

$$P = \begin{pmatrix} t_{1,1} \\ \vdots \\ t_{m,1} \end{pmatrix} w_1^{'} + \begin{pmatrix} t_{1,2} \\ \vdots \\ t_{m,21} \end{pmatrix} w_2^{'} + \cdots + \begin{pmatrix} t_{1,p} \\ \vdots \\ t_{m,p} \end{pmatrix} w_p^{'} \tag{4.13}$$

Table 4.7: Spectral shape associated with variables (NMF method)

↓ Basis Functions	Variables
w_1	A,C
w_2	B
w_3	D
w_4	A,D

where each i^{th} spectrum in P is mapped to the p dimensional space with coordinates $\left(t_{i,1} \dots t_{i,p}\right)^T$. The similar power spectra have similar t coordinates in the new p dimensional space. Thus the spectra with similar t coordinates are clustered together representing the common oscillation. The basis functions determine spectrum of each cluster. In NMF these basis function are either positive or zero and therefore can be used for the analysis of positive quantity like power spectrum.

The results depicting the frequency spectrum of each basis function (from NMF) are shown in Figure 4.3. Variables associated with different spectral shapes (basis functions) are given in Table 4.7. It can be seen that basis functions contain multiple frequency peaks, that are not too different from the individual power spectra of the variables. Moreover grouping based on common oscillations is not so obvious and needs further scrutiny of the basis spectra.

The comparison shows that the proposed method, based on MEMD, points out different common oscillations in all loops quite effectively and an in an automated manner while catering for the the non stationary trends in the data as well.

4.8 Case Studies

4.8.1 Case Study-I

Time series data from 12 tags/measurements of the challenge problem (1934 samples with sampling rate 1 minute) (Tangirala et al. 2005, Thornhill et al. 2002) is analyzed using the proposed scheme. The data is from a pulp manufacturing process simulation where a stream of desired composition, from the combination of soft and hard wood pulp, is formed (process schematic shown in Figure 4.5). It is being reported that the trends exhibit two oscillations, one at frequency 0.002min^{-1} (≈ 500 mins) and other with higher frequency 0.02min^{-1} (≈ 50 mins) (Thornhill et al. 2002). The objective is to process 12 tags (shown in Figure 4.4) using the proposed method, and to analyze how well the method can group the time trends with similar oscillation frequencies.

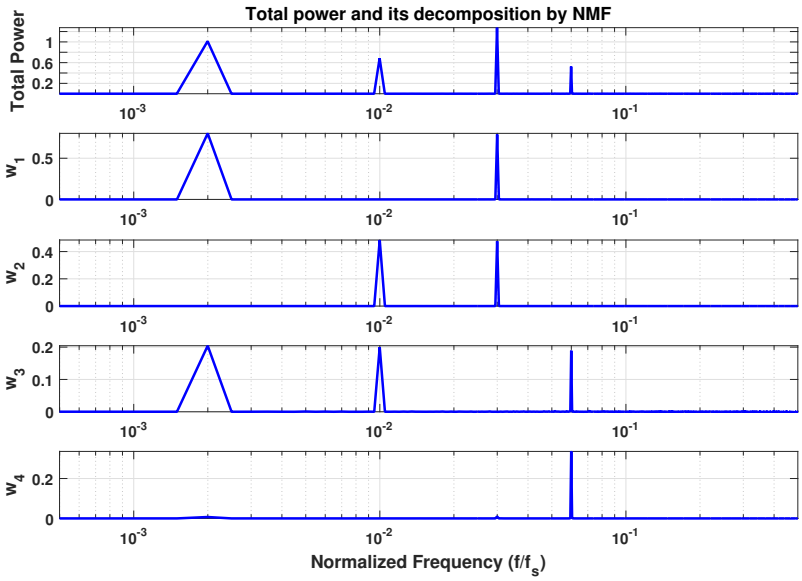

Figure 4.3: Basis functions from NMF method (simulation example).

The correlation coefficient matrix (equation 4.11) formed by the proposed method is given in Table 4.8. The entries highlighted in green are the ones that fulfilled the grouping criteria given in section 4.4.2.

Analysis of Oscillation Groups

The final oscillation groups formed using the information provided by correlation coefficient values (Table 4.8) are shown in Table 4.9. The grouping is performed using the same procedure as adopted for the simulation example.

It can be seen that both tags 2 and 5 have common oscillations in first three IMFs. The sparseness of these IMFs is found to be less than the threshold, showing significant noise content in both these tags. Therefore, tags 2 and 5 are placed together in group G_0.

In the fourth row, only tag 11 IMF fulfils the correlation criteria. This IMF has oscillation frequency of 0.0242 min^{-1}. Group G_1, with tag 11 as member, represents this oscillation . The common oscillations between Tags 1 and 11, given by IMFs in fifth and sixth row exhibits frequency of 0.02 min^{-1}. This oscillation is grouped in G_2.

Oscillations captured by seventh IMF is found to be present in tag 4 only (as per

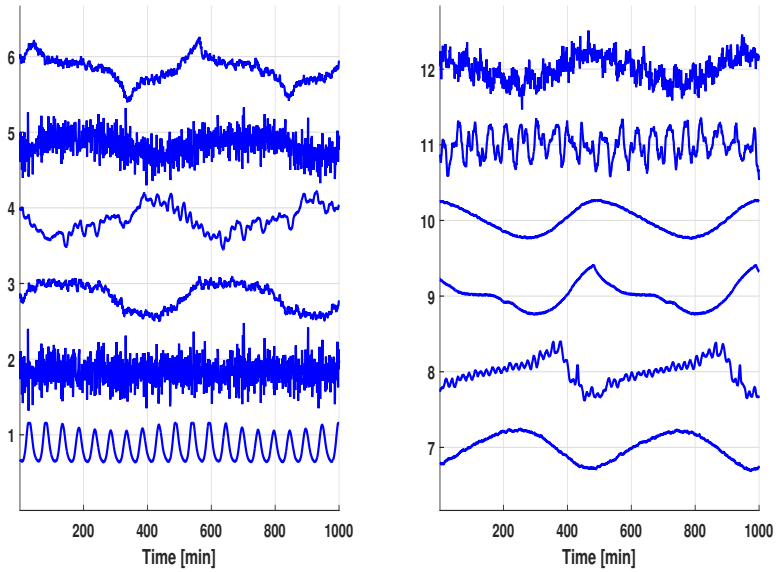

Figure 4.4: Time trends (normalized) case study-I.

criterion in section 4.4). This oscillation (frequency 0.004 min^{-1}) is grouped in G_3.

The eighth IMF gives the largest cluster and represents primary oscillation with frequency 0.002min^{-1}. Tags 3-10 and 12 are shown to exhibit this oscillations (highlighted as green in Table 4.8) and are placed in group G_4. The frequencies and the members of each group are summarized in Table 4.10. The corresponding power spectra of the each group are shown in Figure 4.6.

It is to be noted that the group G_1 represents presence of a oscillation with frequency 0.0424 min^{-1}. The presence of this oscillation has not been reported by the PCA (Thornhill et al. 2002) and GA (El-Ferik et al. 2012) based analysis, whereas both the proposed and NMF based method successfully reported its presence in Tag 11. The proposed method has also highlighted the presence of multiple oscillations in tag 11 (groups G_1 and G_2). Moreover, tag 5 is member of both G_0 and G_4 confirming that it contains significant noise as well as an oscillation with frequency 0.002min^{-1}. The IMFs with maximum SI in each group are used to calculate the period of oscillation using the ACF method (Thornhill et al. 2003). This is an added feature that is not available in existing plant-wide oscillation detection methods.

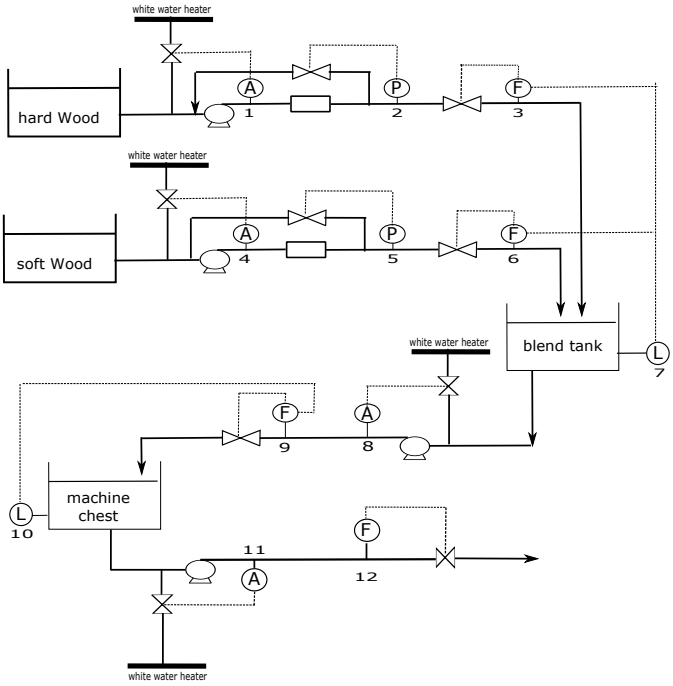

Figure 4.5: Process schematic case study-I.

Root Cause Analysis

The results (Table 4.9) show that 9 out of 12 tags (tags 3-10 and 12) are oscillating with frequency 0.002min^{-1} and group G_4 represents this primary oscillation. It is therefore, quite intuitive to search for the root cause of this oscillation.

In case this oscillation arises due to non-linearity some of the tags in G_4 (frequency 0.002min^{-1}) must exhibit higher order harmonics (Thornhill et al. 2001). A close inspection reveals that the tags included in G_3 do exhibit higher order harmonics of 0.002min^{-1}) and at the same time G_3 is subset of G_4. Thus presence of higher order harmonics confirms the presence of non-linearity as the source of oscillation, as non-linearity induced oscillations exhibit strong harmonic content (Thornhill 2005). In order to locate the source of non-linearity the tags grouped in G_4 can be analyzed using measures like surrogate analysis (Thornhill 2005), harmonics analysis (Aftab et al. 2017c).

4.8.2 Case Study-II

We now consider a more complex case study, where time trends from a refinery process are analysed. The process is a hydrogen reformer and the same case study

Table 4.8: Correlation coefficient matrix Λ for case study-I

IMF	Tags (j)											
(i)↓	1	2	3	4	5	6	7	8	9	10	11	12
1	0.01	1.00	0.12	0.03	1.00	0.13	0.03	0.05	0.02	0.02	0.11	0.47
2	-0.01	0.75	0.07	0.05	0.80	0.10	0.02	0.03	0.02	0.02	0.14	0.42
3	0.06	0.52	0.11	0.09	0.57	0.11	0.03	0.10	0.01	0.01	0.21	0.36
4	0.25	0.33	0.12	0.22	0.35	0.11	0.01	0.24	0.01	0.03	0.69	0.34
5	0.69	0.19	0.13	0.23	0.17	0.08	0.07	0.22	0.01	0.00	0.88	0.31
6	1.00	0.13	0.10	0.26	0.15	0.19	0.01	0.17	0.06	0.03	1.00	0.42
7	0.17	0.11	-0.02	0.61	0.29	-0.05	0.16	0.36	0.31	0.12	0.20	0.23
8	0.11	0.07	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.28	1.00
9	-0.01	0.07	0.12	-0.01	0.01	0.07	-0.02	-0.13	-0.07	-0.11	-0.03	0.06
10	-0.00	0.08	-0.09	-0.10	-0.04	-0.15	0.01	-0.15	-0.09	-0.03	-0.04	0.06

has been analyzed by Thornhill et al. (2002), Tangirala et al. (2005) and Tangirala et al. (2007) using PCA, PSCMAP and NMF based methods respectively. The schematic of refinery process is shown in Figure 4.7. The data consists of 37 Tags (Figure 4.8) each with 512 data points recorded at 1 min sample rate.

The data trends from 37 tags are processed using the proposed method and the final grouping is given in Table 4.11. In total 9 groups (G_0-G_8) are identified with G_0 representing random noise. Frequencies associated with each group can also be seen in spectra plotted in figure 4.9.

Oscillations with low frequency (< 0.0004 min^{-1}) are captured in G_8 and contains Tags 5, 21, 22, 31 and 36. Tag 5 also finds its place in slightly higher frequency group i.e. G_7 thereby confirming the fact that Tag 5 also contains somewhat higher frequency as compared to other members of G_8, thus confirming the result reported by Thornhill et al. (2002). The next group G_6 gives the trends with oscillations having frequency around 0.0008 min^{-1} and contains Tags 8, 12 and 35-37. Oscillations with frequency around 0.001min^{-1} are represented by G_5 and Tags 7-8, 12, 14-15, 31 and 35-37 are part of this group. Here again Tags 35-37 are part of both G_5 and G_6 correctly indicating presence of more than one frequencies in both tags. The frequencies around 0.002min^{-1} are represented by G_4 with Tags 12, 14-18 and 26 as the members.

The biggest cluster is the one oscillating with frequency 0.06min^{-1} and the oscillations are attributed to valve non-linearity (Thornhill 2005). The results show that

Table 4.9: Final grouping case study-I

Tag # ↓	G_0	G_1	G_2	G_3	G_4
1			✓		
2	✓				
3					✓
4				✓	✓
5	✓				✓
6					✓
7					✓
8					✓
9					✓
10					✓
11		✓	✓		
12					✓

Table 4.10: Groups and associated frequencies (case study-I)

Group # ↓	Frequency	Tags
G_0	Noise	$2, 5$
G_1	0.0424 min^{-1}	11
G_2	0.02 min^{-1}	$1, 11$
G_3	0.004 min^{-1}+ harmonics	4
G_4	0.002 min^{-1}	$3, 4, 5, 6, 7, 8, 9, 10, 12$

tags 1-4, 6-8, 10-11, 13, 15-20, 24-26, 28, 30, 33-34 and 37 exhibit this oscillation. Tags 2, 11 and 33-34 are shown to have oscillation of frequency 0.12min^{-1} and grouped in G_2. Tag 37 with frequency $0.3min^{-1}$ is the only member of group G_1. Group G_0 contains the tags that represent variables with only noise or significant noise contribution. Tag 23, 27 , and 29 are members of G_0 only thereby contain only noise, whereas other members 6-9, 13, 15-16, 24-26 and 30 have significant noise content in addition to the presence of oscillations. A strange inclusion in G_0 is tag 32, that contain high frequency oscillation at 0.3min^{-1} as reported in Thornhill et al. (2002). This is so because significant noise content in tag 32 gives the sparseness index of 0.46 and is therefore not included in G_1 along side tag 37.

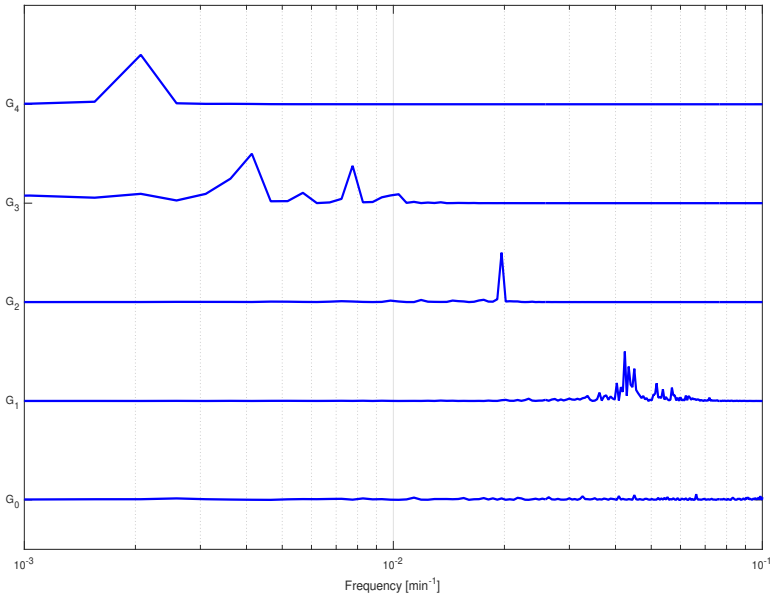

Figure 4.6: Power spectrum of each group (case study-I)

Oscillation Diagnosis

The main oscillation in this case study is the one with frequency $0.06min^{-1}$ and represented by group G_3. The majority of the control loops are affected by this oscillation and hence can be considered as the primary oscillation in the process. It is therefore interesting to see if the proposed method can give some hint about the source of this oscillation. Table 4.11 shows that group G_2 is subset of G_3 and oscillation (frequency $0.12min^{-1}$) is twice as that of G_3 (Figure 4.9). Therefore G_2 can be considered as containing the second harmonic of oscillations captured by G_3 and the non-linearity is the source of this oscillation in the plant.

The tags included in G_2 can help in locating the source of this non-linearity induced oscillation. The analysis reported by Thornhill (2005), Choudhury (2006) and Zang and Howell (2007) also show that tags included in G_2 exhibit non-linearity signatures. Though tags 33 and 34 are declared most probable source of this oscillation by Thornhill (2005), and Zang and Howell (2007) respectively.

Therefore the proposed method is able to detect different clusters in the plant data on the basis of dominant oscillations and at the same time identify the set of variables that are most likely to be the source of the primary oscillation. Clearly, it is

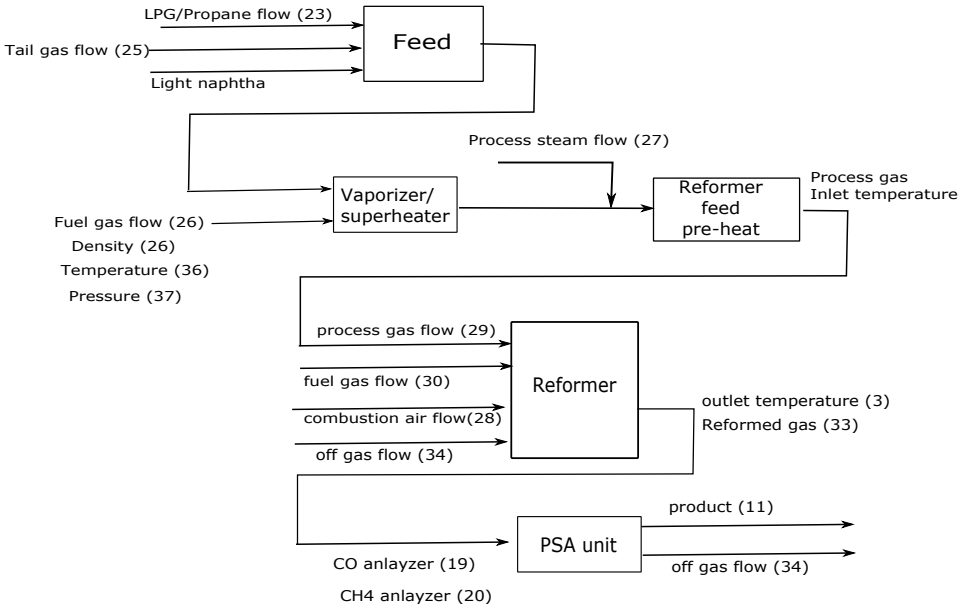

Figure 4.7: Process schematic case study-II.

recommended that the this analysis is complemented by additional investigations to accurately pinpoint the source of the primary oscillation. However, the preceding analysis has significantly reduced the number of tags that need to be analyzed in more detail.

4.9 Conclusions

In this paper a method for plant-wide oscillation detection based on adaptive multivariate data analysis is presented. The method is shown to be robust in identifying dominant oscillations and grouping them together, using both simulation and industrial case studies. The proposed method is capable of handling non-linear and non-stationary time series, where the existing methods are liable to be erroneous. The proposed method is also shown to be capable of indicating the variables that are most likely to be the root cause of oscillation. Furthermore, the method is adaptive in nature and doesn't require *a priori* assumption or pre-processing or any kind of filtering.

4.10 Acknowledgements

The financial support from Siemens AS, Trondheim is gratefully acknowledged. The authors would like to thank Prof A.K. Tangirala for providing the industrial data.

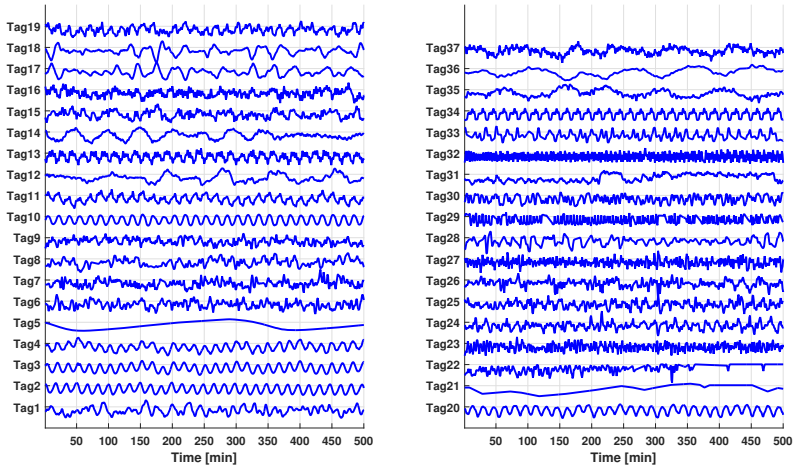

Figure 4.8: Time trends (normalized) case study-II.

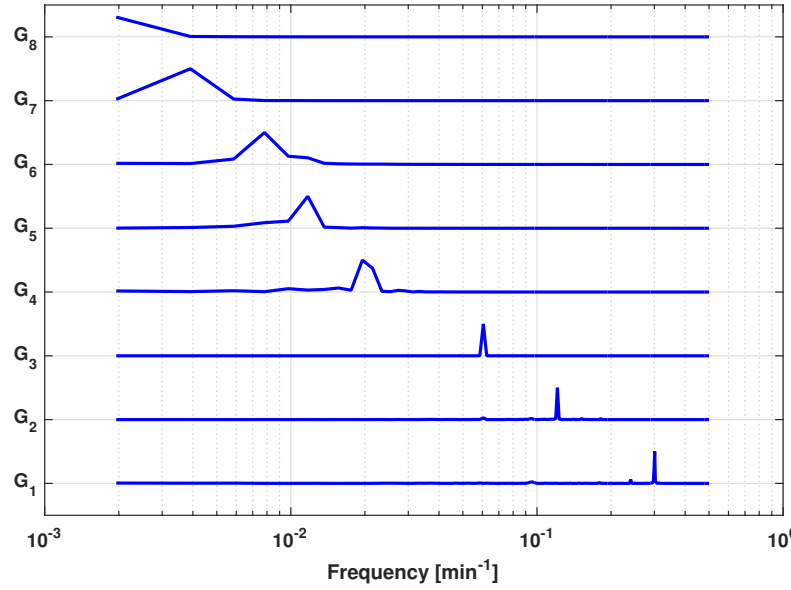

Figure 4.9: Frequency content of each group (case study-II).

Table 4.11: Final grouping case study-II

Tag # ↓	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
1				✓					
2			✓	✓					
3				✓					
4				✓					
5								✓	✓
6	✓			✓					
7	✓			✓		✓			
8	✓			✓		✓	✓		
9	✓								
10				✓					
11			✓	✓					
12					✓	✓	✓		
13	✓			✓					
14					✓	✓			
15	✓			✓	✓	✓			
16	✓			✓	✓				
17				✓	✓				
18				✓	✓				
19				✓					
20				✓					
21									✓
22	✓								✓
23	✓								
24	✓			✓					
25	✓			✓					
26	✓			✓	✓				
27	✓								
28	✓			✓					
29	✓								
30	✓			✓					
31									✓
32	✓								
33			✓	✓					
34			✓	✓					
35						✓	✓		
36						✓	✓		✓
37		✓		✓		✓	✓		

Part II

Oscillation Diagnosis

Chapter 5

Detecting Non-Linearity Induced Oscillations in Control Loops via Harmonics

Paper C: Detecting Non-Linearity Induced Oscillations via the Dyadic Filter Bank Property of Multivariate Empirical Mode Decomposition

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

Journal of Process Control 60 (2017) 68-81

Abstract

Non-linearity induced oscillations in control loops are characterized by the presence of higher order harmonics. In this paper the dyadic filter bank property of the Multivariate Empirical Mode Decomposition (MEMD) is exploited to reveal the harmonic content of the oscillatory signal to indicate the presence of non-linearity. Once the harmonics are identified the extent of non-linearity is evaluated automatically using Degree of non-linearity measure (DNL) introduced in our previous work (Aftab et al. 2016). Although detection of non-linearity via harmonics is an old concept; any automatic method has still not been reported. Moreover, the existing methods suffer from the restrictive assumption of signal stationarity. The proposed method is more robust in identifying the non-linearity induced oscillations and is adaptive and data driven in nature and thus requires no a priori assumption about the underlying process dynamics. The proposed method can also differentiate among the different sources of multiple oscillations, for example combinations

of non-linearity and linear sources or two non-linear sources. Apart from detecting the non-linearities the proposed method can also contribute in locating the source of non-linearity using the degree of non-linearity index, thereby reducing the maintenance time to a considerable extent. The robustness and effectiveness of the proposed method is established using industrial case studies and results are compared with existing methods based on higher order statistics (Choudhury et al. 2008) and surrogate based methods (Thornhill 2005).

5.1 Introduction

Oscillations are one of the major causes of degraded control performance in industrial control systems. Product variability, equipment wear, and reduced profitability are the major after-effects associated with oscillatory control loops. Oscillations can be due to multiple causes like poor controller tuning, disturbances and non-linearities. Non-linearities, ranging from inherent non-linearities in the process itself to ones associated with sensor and actuator faults, are among the major causes of oscillation in industrial control systems. The high complexity and large size of modern industrial processes necessitate the early and correct diagnosis of the oscillation source for timely maintenance and reduced shut down time.

Detection of non-linearity induced oscillations has attracted considerable attention from the research community for more than a decade. Detailed literature reviews can be found in Thornhill and Horch (2007) and Bacci di Capaci and Scali (2018), while only a brief overview of some procedures is provided here.

Horch (1999) proposed that an odd correlation function between manipulated and process variable indicates the presence of valve non-linearities in non-integrating plants. Other methods based on shape analysis formalisms for detecting valve non-linearities are put forth by Srinivasan et al. (2005a), Hägglund (2011) and Yamashita (2005).

Methods pertaining to non-linear time series analysis are also being used by some authors to detect oscillations caused by non-linearities in control loops. One such method, based on higher order statistics (HOS), is put forward by Choudhury et al. (2008), Chaudhry et al. (2004). The bi-spectrum of the non-linearity induced oscillations shows peaks at the corresponding bi-frequencies. Two indices, termed the Non-Gaussianity Index (NGI) and the Non-Linearity Index (NLI), are used to detect the presence of non-linearity. The signal is classified as the output of a linear Gaussian process if NGI ≤ 0.0001; whereas NLI≥ 0.01 classifies the signal to be outcome of a non-linear process. Both the NGI and the NLI need to be above their respective thresholds for the signal to be identified as non-linear. In addition, a total non-linearity index (TNLI) is defined to calculate the total non-linearity in

the signal. An important limitation for this method is that symmetric waveforms (like square or triangular) exhibiting odd harmonics, cannot be captured (Thornhill 2005, Zang and Howell 2003).

Another important method is that of surrogate testing proposed by Thornhill (2005). Surrogates are time series which have exactly the same power spectrum as original time series, but phase is randomized to remove any kind of phase coupling. The algorithm makes use of the fact that the time series from the non-linear source depicts phase coupling and hence is more predictable than its surrogate counterpart. A non-linearity measure called N-measure is defined to accept or reject the null hypothesis that the signal is outcome of a linear Gaussian process. Different parameters like embedding dimensions, the number of nearest neighbors and number of oscillation cycles [1] need to be tuned in order to obtain reliable results.

Non-linearity detection methods based on the Hilbert Huang Transform (HHT) and intra-wave frequency modulation, that are applicable to non-stationary time series, are proposed by Babji et al. (2009) and Aftab et al. (2016). The former gives only the qualitative picture while the latter provides a measure to quantify the severity of non-linearity. The method by Aftab et al. (2016) provides an automatic way to detect non-linearities but it suffers from the inherent mode mixing limitation of empirical mode decomposition (EMD), the first step in obtaining the HHT. The mode mixing problem may result in false reporting of the non-linearity in presence of noise and multiple oscillations.

The idea that the oscillations caused by non-linearities contain higher order harmonics is explored by Thornhill et al. (2001). A measure called the distortion factor D is introduced that measures the energy spread among the fundamental frequency and harmonics to detect the extent of non-linearity. This method has certain limitations, the foremost being the manual detection of harmonics and the requirement of stationary data. It is also prone to report inaccurate results in the presence of noise and multiple oscillations due to other causes. This paper is aimed at addressing the limitation of the previous work (Aftab et al. 2016), where the mode mixing problem can result in false detection of non-linearity. In the method proposed here, instead of relying on the intra-wave frequency modulation to classify the time series data, the harmonic content is analyzed using the noise assisted multivariate EMD (MEMD). The presence of harmonics is taken as an indication of non-linearity and then intra-wave frequency modulation is used to measure the extent of non-linearity in the signal.

The advantages offered by the proposed method are threefold. First it removes the mode mixing problem associated with non-linearity detection method given

[1] See the original reference for details

in(Aftab et al. 2016) while retaining the data driven ability of the EMD process. Second, the proposed method is automatic, can sift out the harmonic content adaptively, and can work in the presence of noise and oscillations caused by multiple sources. Third, the extent of non-linearity using intra-wave frequency modulation can help in isolating the source loop of non-linearity.

The paper is organized as follows: Section 5.2 gives an overview of the EMD, Multivariate EMD and the associated dyadic filter bank property. Section 5.3 outlines the steps involved in harmonic extraction. The degree of non-linearity and criterion to isolate the non-linearity are discussed in section 5.4 and section 5.5 respectively. Section 5.6 gives the detailed algorithm, followed by simulation and industrial case studies in section 5.7 and 5.8 respectively, followed by conclusions.

5.2 Empirical mode Decomposition (EMD) and Variants

5.2.1 Standard EMD

Standard EMD is also referred to as univariate EMD, as it caters only for one dimensional signals. EMD is a data driven procedure that adaptively sifts out different components, called Intrinsic Mode Functions (IMFs), from the signal. An IMF is defined as a function that has zero mean and the number of extrema and zero crossings at most differ by one. The method sifts out fast components from the slower ones through an iterative procedure that involves identification of local extrema and fitting an envelope through them using cubic splines. The low frequency components are local means $m(t)$ of the envelope and the local fast component $d(t)$ is then given by (Rilling et al. 2003)

$$d(t) = x(t) - m(t) \tag{5.1}$$

The sifting process is repeated until $d(t)$ fulfils the criteria to be an IMF. The IMF is then subtracted from the original signal and the sifting procedure is started again on the residue and continues till no more IMFs are left to be extracted. The signal can then be expressed as the sum of IMFs and residue as

$$x(t) = \sum_{i=1}^{N} c_i(t) + b(t) \tag{5.2}$$

where $x(t)$ is the input time series, $c_i(t)$ is the i^{th} IMF, $b(t)$ is the residue, and N is the total number of IMFs. The details of the procedure can be found in Huang et al. (1998) and Rilling et al. (2003).

5.2.2 Multivariate EMD (MEMD)

Given the data driven nature and distinct properties of EMD, efforts were made to extend its applicability to multivariate signals (signals with more than one variables), with the major obstacle to such an extension being the generation of envelope and its mean in the higher dimensions. Rilling et al. (2007) proposed that the envelopes in the 2-dimensional space can be generated by extrema sampling of multiple signal projections in a complex plane. The mean is then calculated by averaging the envelopes from these projections.

Rehman and Mandic (2010a) generalized the same concept to n-dimensional signals (signals with n variables), the so called Multivariate EMD (MEMD), by generating envelopes in n-dimensional space. The multiple directions are represented by vectors from the center of the unit n-dimensional sphere to the uniformly distributed points on its surface. The points are generated using uniform sampling by means of Hammersley and Halton sequences.

Generating Uniform Points

Halton sequence is used to generate points for uniform sampling. The kth sample of a one dimensional Halton sequence $\Phi_p(k)$ is written as

$$\Phi_p(k) = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} \dots \frac{a_r}{p^{r+1}} \tag{5.3}$$

where prime base-p representation of k can be expressed as (Wong et al. 1997)

$$k = a_0 + a_1 p + a_2 p^2 \dots a_r p^r \tag{5.4}$$

where each a_i is an integer in $[0, p-1]$. Starting from $k = 0$ and with $p_1, p_2 \dots p_q$ as q prime numbers, the kth sample of the q-dimensional, Halton sequence then becomes

$$\left(\Phi_{p_1}, \Phi_{p_2}, \dots \Phi_{p_q}\right) \tag{5.5}$$

The uniform points created using Equation (5.5) can be used to generate direction vectors representing multiple directions in $q-$dimensional space. The dimensions (q in this case) are equal to the number of variables in the multivariate signal. The uniformly sampled points on a three dimensional unit sphere, using the equation 5.5, are given in Figure 5.1.

5.2.3 Envelope and Mean in *n*-dimensions

The next step is to generate the envelope of all the projection curves, using cubic splines which are in turn averaged to calculate the mean envelope $m(t)$. Fast modes are extracted, in an analogy to the univariate case, by $d(t) = x(t) - m(t)$

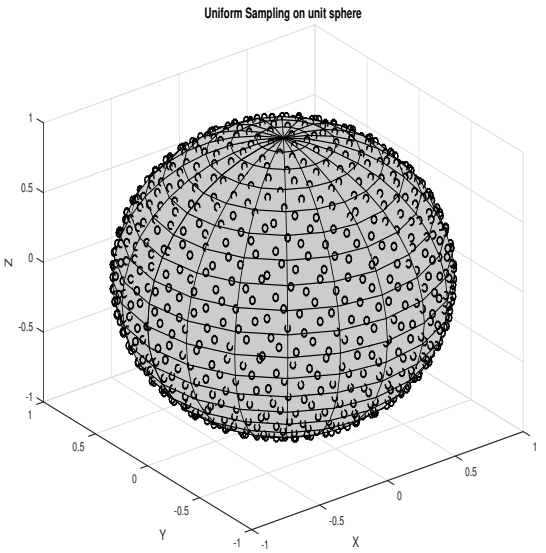

Figure 5.1: Random sampling with uniform distribution on the unit sphere.

and the procedure is iterated in a similar fashion till there are no more IMFs to extract.

The IMFs so generated are also n-dimensional signals, with each dimension relating to the corresponding variable in the input signal. The detailed algorithm to process a multivariate signal $\mathbf{X}$ (given in equation 2.3) with n variables is as follows.

Step I Set up K direction vectors ($K = 64$ [2] is used here) u^k with $k = 1 \dots K$ by choosing uniformly spaced points on the n dimensional sphere.

Step II Find the projections $p^k(t)$ of the input signal $\mathbf{X}$ along the direction vectors u^k using

$$p^k(t) = \mathbf{X}u^k \quad \forall \quad k = 1 \dots K$$

Step III Identify the maxima of projections $p^k(t)$ and corresponding time instants t^k.

Step IV Generate n-dimensional envelope curve $\mathbf{e}^k(t)$ by interpolating $[t^k, \mathbf{X}(t^k)]$.

[2] The number of direction vector should be sufficiently large than number of variables in the signal

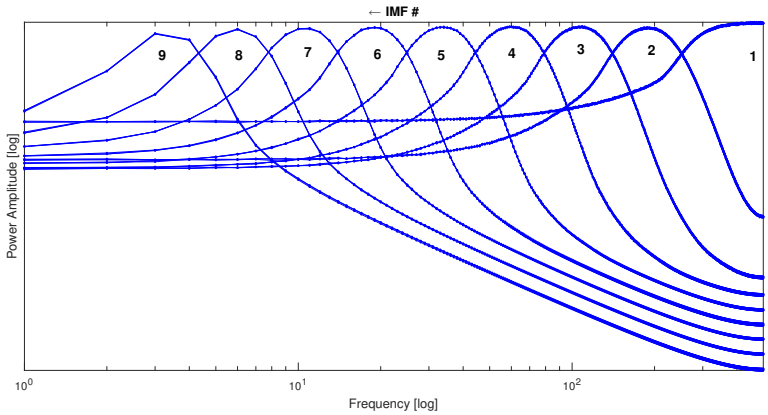

Figure 5.2: Dyadic filter bank property of multivariate EMD (MEMD).

Step V The n-dimensional mean of envelope curve is then given by

$$\mathbf{m} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{e}^k(t) \tag{5.6}$$

Step VI Extract n-dimensional detail $\mathbf{d(t)}$ using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$.

Step VII Repeat the steps I-VI using $\mathbf{d(t)}$ as input till it fulfils the criteria for an IMF.

Step VIII Calculate the residue $\mathbf{b(t)} = \mathbf{X(t)} - \mathbf{d(t)}$ and iterate the procedure on $\mathbf{b(t)}$ till there are no more IMFs left to be extracted.

5.2.4 Dyadic Filter Bank Property of the MEMD

An important consequence of the MEMD is that it acts like a series of band pass filters in the presence of white noise, analogous to the wavelet decomposition (Rehman and Mandic 2011). This property is referred to as the dyadic filter bank property of MEMD. In order to elaborate this concept an average power spectrum of IMFs from 1000 different realizations of a multivariate signal consisting of three white noise sequences is shown in Figure 5.2. The IMFs can be seen as the output of a series of band pass filters, with the frequency of each band decreasing with the IMF index.

Table 5.1: Enforcing dyadic filter bank property

Step 1 Generate two uncorrelated white Gaussian noise sequences with length same as that of the original signal.
Step 2 Add the two noise sequences to the original signal to make three channel/variable signal (here channel and variable are used interchangeably).
Step 3 Process the signal using Multivariate EMD. The resulting IMFs will have three channels.
Step 4 Retain the IMFs in the channel corresponding to the original signal, and discard the IMF components corresponding to noise channels

5.3 Extraction of Harmonics in Univariate Signals

The dyadic filter bank property of MEMD can be enforced on the univariate signals using the so called noise-assisted MEMD (NA-MEMD). To achieve this the univariate signal is added with two or more noise sequences to make a multivariate signal which is then processed using the MEMD algorithm. The details of the procedure are discussed next.

5.3.1 Enforcing the Dyadic Filter Bank Property

The harmonics exhibited by the non-linearity induced oscillations can be extracted by virtue of the dyadic filter bank property of Multivariate EMD. In order to enforce the dyadic filter bank property of MEMD, input signal of one variable is added with two or more noise sequences to generate a multivariate signal. This multivariate signal is processed using MEMD algorithm (the detailed steps are described in Table 5.1) (Rehman et al. 2013). The IMFs corresponding to the input signal are retained (noise IMFs are discarded) and are aligned according to the filter bank structure. These IMFs are ready to be tested for the presence of harmonics, but first non-significant and noisy IMFs have to be discarded.

5.3.2 Discarding Spurious and Noisy IMFs

Discarding Pseudo IMFs

The EMD process is prone to produce pseudo IMFs that are poorly correlated with the input signal due to spline fitting issues (Peng et al. 2005, Srinivasan and

Rengaswamy 2012, Aftab et al. 2016). The MEMD is also no exception. Therefore in oder to get the significant IMFs, the correlation index of each IMF with original signal is calculated using relation

$$\rho_i = \frac{Cov(c_i, x)}{\sigma_x \sigma_{c_i}}, \quad i = 1, 2, 3 \ldots N \tag{5.7}$$

where Cov denotes the covariance; σ_x and σ_{c_i} are the standard deviations of the signal and the IMF, respectively, and N is the total number of IMFs. The normalized correlation coefficient λ_i is calculated for each IMF

$$\lambda_i = \frac{\rho_i}{max(\rho_i)}, \quad i = 1, 2, 3 \ldots N \tag{5.8}$$

Only the IMFs with ($\lambda > \eta$) are retained, where the value of threshold η is discussed in Section 5.3.4.

Discarding Noisy IMFs

It may happen that, in case of signals with large noise levels, IMFs with large noise content and little information are retained in the previous step. Therefore these noisy IMFs need to be discarded as the objective is to identify the harmonics in the IMFs and these harmonics will have distinct peaks in the frequency spectrum.

To get rid of IMFs consisting mainly of noise, a method based on the sparseness index(Hoyer 2004, Srinivasan and Rengaswamy 2012) is used. The frequency spectrum of noise dominated IMFs will be spread across a broad frequency range whereas for oscillatory IMFs it will exhibit distinct peaks. The sparseness index SI of frequency spectrum $X(f)$ of signal $x(t)$, given by (5.9) will be almost zero for a white noise signal whereas it will attain a value near one for the oscillatory signals.

$$SI(x) = \frac{\sqrt{I} - \left(\sum_{i=1}^{I} |X_i| / \sqrt{\sum_{i=1}^{I} |X_i|^2} \right)}{\sqrt{I} - 1} \tag{5.9}$$

Here X is the frequency spectrum and I is total number of frequency bins. The IMFs with $SI > S_{Thresh}$, containing distinct peaks, are retained for further analysis. The default value of S_{Thresh} is discussed in Section 5.3.4.

5.3.3 Extracting Harmonics

Once the significant and oscillatory IMFs are obtained using the procedure laid down in Section 5.3.2, the next step is to assess the IMFs for presence of harmonics. The presence of oscillations in retained IMFs and the corresponding time period is determined using the Auto Covariance Function (ACF) method proposed

by Thornhill et al. (2003). This is due to the fact that the ACF of an oscillatory signal oscillates with the same frequency with noise confined to zero lag only. The IMFs are converted to the corresponding ACF and zero crossings are evaluated.

If Δt is the time interval between two successive zero crossings and for H such intervals, the average period of oscillation $\bar{T}_p$ will be given by [Thornhill et al. (2003) and Srinivasan and Rengaswamy (2012)]

$$\bar{T}_p = \frac{2}{H} \sum_{i=1}^{H} (\Delta t_i) \tag{5.10}$$

The regularity of the oscillation is determined from the r statistics calculated using the relation

$$r = \frac{1}{3} \frac{\bar{T}_p}{\sigma_{T_p}} \tag{5.11}$$

where σ_{T_p} is the standard deviation of the the time intervals between zero crossings. The oscillation is detected if $r > 1$ and the mean period of oscillation $\bar{T}_p$ is reported.

5.3.4 Default Parameter Settings

The default settings of different parameters used in this work are outlined in Table 5.3. The discussion about these default settings, for each parameter, are discussed next.

Number of Zero Crossings

The presence or absence of harmonics is ascertained from the zero crossings of ACF using (5.10) and (5.11). Thornhill et al. (2003) recommended to use the first eleven zero crossings for the detection of oscillation. This is so because in the absence of persistent oscillation the auto covariance function decays as a function of lags and may contain spurious zero crossings at large lags. A smaller number of zero crossings tend to make the estimates of $\bar{T}_p$ and σ_{T_p} unreliable. Using eleven zero crossings balances these concerns.

Correlation Threshold η

In order to discard the spurious IMFs generated in MEMD process, the normalized correlation coefficient, λ, given in (5.8), is used. Only the IMFs with ($\lambda > \eta$) are retained with the threshold η set to 0.25 in this work. This value is chosen such that the IMFs representing significant harmonics can be captured while discarding most pseudo components. The threshold is lower than the value of 0.5 used by Srinivasan and Rengaswamy (2012) because the objective in this work is to capture

Table 5.2: Correlation coefficient of different IMFs for square wave

IMF	$\bar{\lambda}$	Harmonic
1	0.09	Thirteenth
2	0.14	Eleventh
3	0.16	Ninth
4	0.22	Seventh
5	0.28	Fifth
6	0.33	Third
7	1.00	Fundamental

both the fundamental oscillation as well as the associated harmonics, if they are present.

To elaborate this further consider the example of a square wave with added white noise (shown in the bottom row of Figure 5.4) which is characterized by the presence of odd harmonics. The average correlation coefficients $\bar{\lambda}$ of extracted IMFs, for 1000 different noise realizations are summarized in Table 5.2. It is clear that selecting the threshold of 0.25 will enable the proposed method to identify third and fifth harmonics which is sufficient for the purpose of detecting the non-linearity. The chosen threshold is a compromise between the number of higher order harmonics that are discarded and avoiding spurious IMFs that may corrupt the analysis.

Sparseness Threshold (S_{Thresh})

In order to extract harmonics, IMFs with sparseness index near to one are retained. In order to get an idea about the value of sparseness index, the spectra of four signals (from industrial data) with different sparseness are shown in Figure 5.3. The corresponding SI values are given on y-axis. The first row shows the spectrum with harmonic content and has sparseness of 0.67. It can be seen (bottom three spectra) that sparseness index up-to 0.54 may contain spectrum spread across broad frequency range. As we are looking for distinct peaks corresponding to harmonics, the threshold (S_{Thresh}) used in this paper to select the oscillatory IMF can be set as 0.58. The threshold is based on the authors' experience and can be lowered but it will increase the risk of spurious harmonic detection. The effects of changing this threshold on the industrial case study are given in Section 5.8.

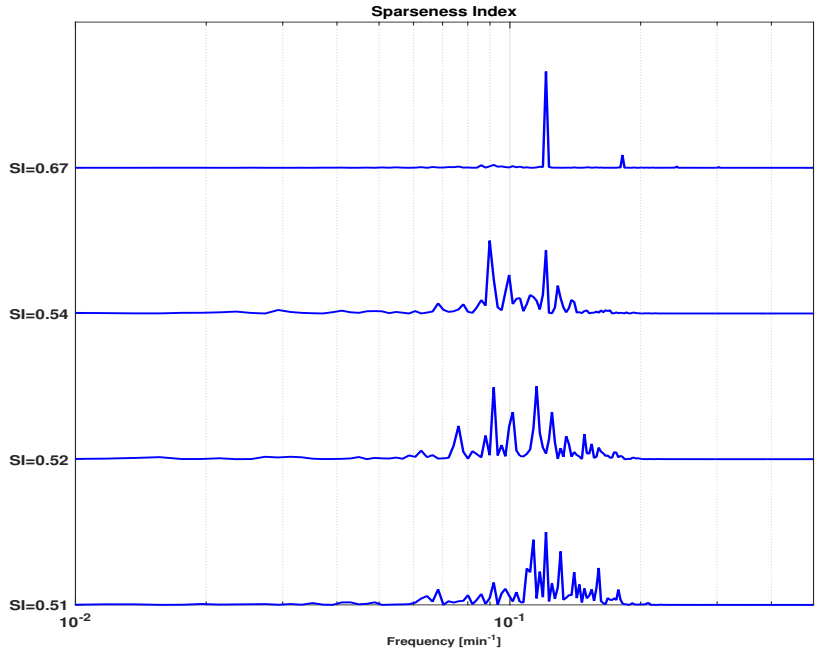

Figure 5.3: Frequency spectra with sparseness index 0.67 (first row), 0.54 (second row), 0.52 (third row) and 0.51 (bottom row).

5.3.5 Automatic Detection of Harmonics

The presence of harmonics is established if there exists mean oscillation frequencies of IMFs (with $r > 1$) that are integral multiples. The automatic detection of harmonics, the core element of this work, is carried out via following steps

1. Calculate the mean frequency $\bar{\Omega}$ and maximum(minimum) frequency Ω_{max} (Ω_{min}) of oscillation, for each IMF with $r > 1$, using

$$\begin{aligned} \bar{\Omega} &= \frac{1}{\bar{T}_p} \\ \Omega_{max} &= \frac{1}{\bar{T}_p - \sigma_{T_p}} \\ \Omega_{min} &= \frac{1}{\bar{T}_p + \sigma_{T_p}} \end{aligned} \tag{5.12}$$

2. Identify the most correlated IMF i.e with $\lambda = 1$ and corresponding frequency $\bar{\Omega}_{\lambda_{max}}$.

Table 5.3: Default parameters setting

Parameter	Default value
Zero Crossings	11
Correlation Threshold (η)	0.25
Sparseness Threshold (S_{Thresh})	0.58

3. The presence of harmonics is confirmed if oscillations are found at integer multiples of the base frequency; i.e. if for IMF $i, i \in \mathbf{Z}^+$[3] (different from the fundamental component)

$$\Omega_{max_i} \leq k\bar{\Omega}_{\lambda_{max}} \leq \Omega_{min_i} \quad k > 1 \tag{5.13}$$

4. In case no harmonics are detected for the most correlated IMF; the steps $2-3$ can be repeated for the next most correlated IMF and so on.

5. The condition that the IMF containing the fundamental frequency of oscillation has higher correlation with the actual signal than its harmonics must be fulfilled, to declare two IMFs as a harmonic fundamental pair, i.e.

$$\lambda_f > \lambda_{h_i} \quad i \in \mathbf{Z}^+ \tag{5.14}$$

   where λ_f and λ_{h_i} are the normalized correlation coefficient of fundamental and i^{th}harmonic IMF respectively.

6. Similarly the relation in (5.14) shall also hold for different harmonics if they arise from the same non-linearity induced oscillation; that is for k^{th} harmonic.

$$\lambda_{h_k} > \lambda_{h_{k+1}} \quad k \in \mathbf{Z}^+ \tag{5.15}$$

   If two IMFs fulfil (5.13) ; but not (5.15), they are considered two different oscillations and not harmonics. This is important to avoid spurious harmonic detection when a signal is affected by multiple sources of oscillation.

The signals whose IMFs exhibit harmonics are then classified as the ones oscillating due to non-linearity. The extent of non-linearity can then be evaluated using the Degree of Non-Linearity (DNL) measure as given in our previous work Aftab et al. (2016). The DNL is based on the concept of intra-wave frequency modulation and instantaneous frequency (IF); a brief overview of these concepts is provided later in section 5.4.

[3] Z^+ is the set of positive integers

5.3.6 Illustrative Example

The rules for harmonic detection outlined in the preceding section can be explained by considering the following illustrative example. Two cases are considered here.

Multiple Oscillations Case

Consider time series data from a closed loop system suffering oscillations due to a combination of a stiction non-linearity and an external sinusoidal disturbance. The fundamental frequency of the stiction non-linearity is 0.0134sec^{-1} whereas the sinusoidal disturbance has frequency of 0.0067sec^{-1}. The simulation model is explained in detail in Section 5.7 . The oscillatory response of the system is shown in the first row of Figure 5.4. In total three oscillatory modes are identified. The details of these modes are given in Table 5.4.

It can be seen that first two IMFs form a fundamental harmonic pair with second IMF designated as fundamental and first IMF as 3^{rd} harmonic according to the rules given in (5.13-5.15). It is to be highlighted that the stiction oscillation, in this case, is characterized by the presence of odd harmonics that can not captured by the bi-spectrum (higher order statistics) based method as discussed by Thornhill (2005) and Zang and Howell (2003).

It is to be noted that the 3^{rd} IMF represents the external sinusoidal disturbance, which is regarded as an oscillation separate from the stiction induced oscillation by virtue of the rule given in (5.15). Thus the proposed algorithm can identify multiple oscillation caused by different sources. Oscillations caused by two different non-linear sources can be detected in a similar fashion.

Symmetric Waves with Odd Harmonics

In order to further establish the fact that the proposed scheme can correctly identify the harmonics in the symmetric waveforms exhibiting odd harmonics the square wave corrupted by white noise of variance 0.1 is considered. The square wave can be seen in the second row of Figure 5.4. The results summarized in Table 5.4 show that the proposed method can identify the presence of odd harmonics (3^{rd} and 5^{th}) in the first and second IMFs with the fundamental frequency residing in the third IMF. The same signal is declared linear by the bi-spectrum based method (NGI=-0.0087).

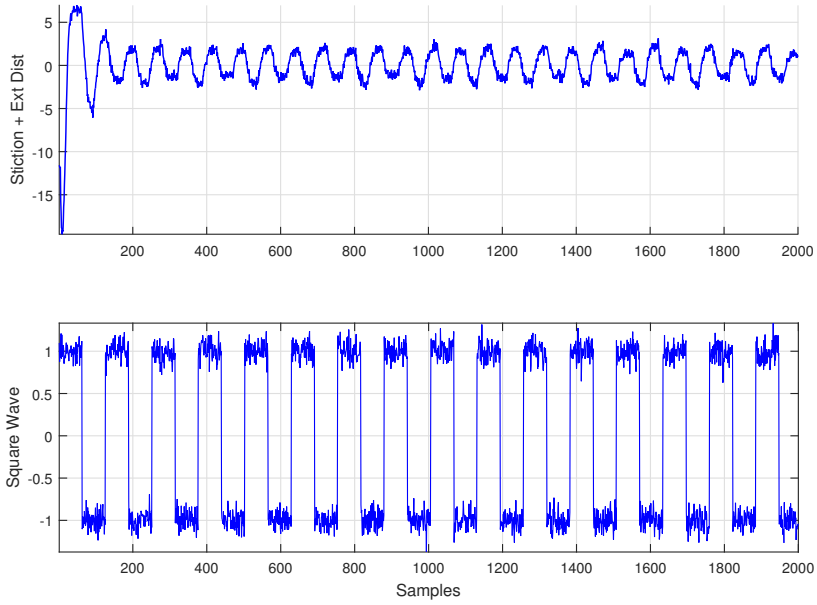

Figure 5.4: Harmonic detection illustrative example: stiction plus sinusoidal disturbance (first row); square wave (second row).

5.4 Intra-Wave Frequency Modulation and Degree of Non-linearity

5.4.1 Instantaneous Frequency (IF)

The first step in the computation of the Instantaneous Frequency (IF) of the signal is the creation of an analytic signal using the Hilbert transform. The Hilbert transform $Y(t)$ of a signal $W(t)$, also regarded as convolution of $W(t)$ and $1/\pi t$, is given by

$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{W(\tau)}{t-\tau} d\tau = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{W(t-\tau)}{\tau} d\tau \tag{5.16}$$

where P indicates Cauchy's principal value of the integral. The analytic signal $Z(t)$ is then given by

$$\begin{aligned} Z(t) &= W(t) + jY(t) = a(t)e^{j\theta t} \\ a(t) &= \sqrt{W^2(t) + Y^2(t)}, \quad \theta(t) = \arctan \tfrac{Y(t)}{W(t)} \end{aligned} \tag{5.17}$$

Here the amplitude and phase are functions of time and the instantaneous frequency (IF) is defined as the time derivative of the phase function $\theta(t)$; given by

Table 5.4: Harmonic detection illustrative example

Dist	IMF	λ	$\bar{\Omega}$	Ω_{min}	Ω_{max}	Harmonics	Type
Ext Dist Plus Stiction	1	0.53	0.0348	0.0278	0.0466	Yes (3^{rd})	Non-Linear
	2	1.0	0.0134	0.0132	0.0137		
	3	0.30	0.0067	0.0062	0.0072	No	Linear
Square Wave	1	0.29	0.0399	0.378	0.0422	Yes (3^{rd}) (5^{th})	Non-Linear
	2	0.32	0.0239	0.0236	0.0243		
	3	1.0	0.0079	0.0079	0.0080		

(Huang et al. (2009))

$$\omega(t) = \frac{d\theta(t)}{dt} = \frac{1}{a^2}[W\dot{Y} - Y\dot{W}] \tag{5.18}$$

The analytical signal formed in (5.17) can give the correct IF only if the original signal $W(t)$ fulfils the properties of an IMF. Therefore the IMF generation via EMD is the first step in calculating the IF and this combined procedure of applying EMD and Hilbert transform to arrive at the analytical signal is termed the Hilbert Huang Transform (HHT); further details can be seen in Huang et al. (1998).

5.4.2 Intra-Wave Frequency Modulation

The term intra-wave frequency modulation has been coined by Huang et al. (2009; 2014) and unlike the traditional meaning of modulation it refers to change of instantaneous frequency within one cycle of oscillation. In order to explain this fact consider two waveforms (shown in first row of Figure 5.5), a triangular wave and a sine wave both having 4 Hz frequency. The triangular wave contains odd harmonics, a characteristic of non-linearity induced oscillations.

The instantaneous frequency for both the waveforms is presented in second row of Figure 5.5. It can be seen that in case of triangular wave the instantaneous frequency changes within an oscillation cycle and varies between fundamental harmonic (4 Hz) and third harmonic(12 Hz). This fluctuation of IF within one cycle is termed as intra-wave frequency modulation. The IF for the sine wave is flat at the value of 4Hz showing no fluctuation or intra-wave frequency modulation.

The non-linearity induced oscillations thus give rise to intra-wave frequency modulation i.e variation of the IF within one period of oscillation (Huang et al. 1998, Babji et al. 2009, Wang et al. 2012, Aftab et al. 2016). The intra-wave frequency modulation increases with increase in non-linearity of the signal (the details can be

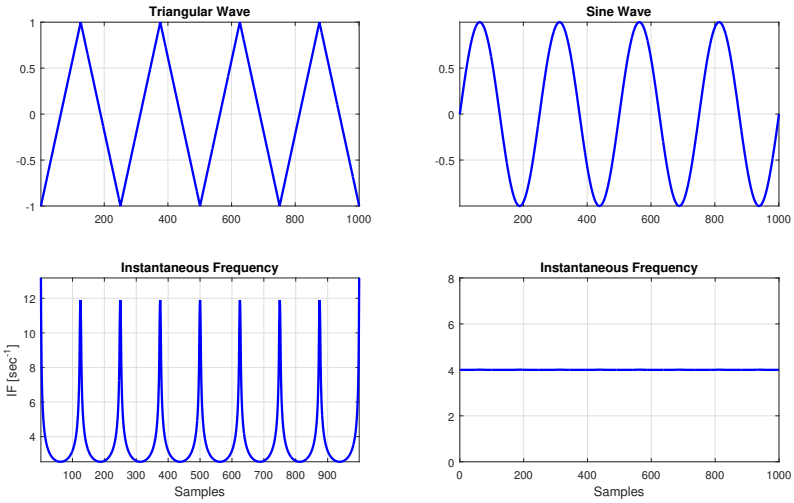

Figure 5.5: IF comparison for triangular and sine wave.

seen in (Aftab et al. 2016) given in appendix A) and the extent of this modulation can be used to quantify the extent of non-linearity.

A closer look at the filter bank characteristics (Figure 5.2) reveals an overlapping region among the spectra of adjacent IMFs. This overlapping allows the presence of more than one harmonic within an IMF and hence gives rise to intra-wave frequency modulation.

In order to elaborate this further, consider a signal from industrial data oscillating due to the non-linearity. The power spectrum of the signal is shown in first row of Figure 5.6. The fundamental frequency at 0.06min^{-1} and higher harmonics (peaks are shown with red circles). The spectra of extracted IMFs are shown in second row of Figure 5.6. The fundamental harmonic is shown by the black curve (third IMF). There is some contribution from the second harmonic while there is negligible energy in higher harmonics. The second harmonic primarily rests in the second IMF (blue curve) with some signatures of the fundamental and third harmonics. The third harmonic resides in the first IMF (red curve), with significant contribution also from the second, fourth and fifth harmonics.

Thus the IMF with greater harmonic content will show greater variation in IF within one cycle. This effect is evident from Figure 5.7 where the IF for 1^{st} IMF, plotted in red, contains third and higher harmonics, thus showing higher intra-wave frequency modulation. The black curve gives IF for the 2^{nd} IMF containing predominantly second harmonic but have lesser intra-wave frequency fluctuation

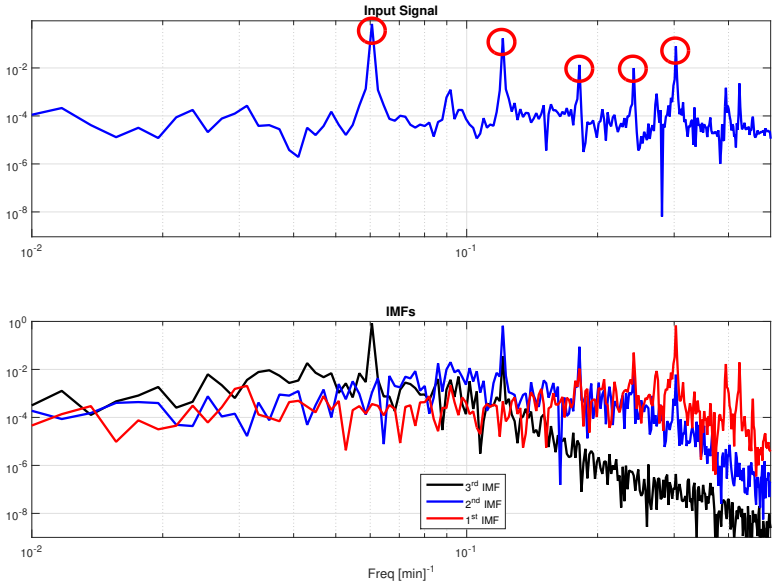

Figure 5.6: Harmonic content of input signal and corresponding IMFs.

as compared to red curve. The signal having a higher level of harmonic content will thus have a higher level of non-linearity. The quantification of the extent of non-linearity is given by the degree of non-linearity measure which is explained next.

5.4.3 Degree of Non-Linearity (DNL) Index

As discussed in previous section that variation of fluctuation of the IF within one wave cycle increases with increase in the extent of non-linearity. Therefore, Huang et al. (2014) has suggested to use the variation of IF from its mean value for the quantification of non-linearity. The index **Degree of Non-Linearity (DNL)** is defined to quantify the non-linearity in the signal given by:

$$DNL \quad = std\left[\frac{IF-IF_z}{IF_z}\right] \tag{5.19}$$

where 'std' is the standard deviation, IF is the instantaneous frequency and IF_z is the constant mean frequency over one wave cycle. The DNL, weighted by the amplitude,for an i^{th} IMF can be defined as (Huang et al. 2014, Aftab et al. 2016).

$$DNL_i = std\left[\left\{\frac{IF_i - IF_{z_i}}{IF_{z_i}}\right\}.\frac{a_{z_i}}{\bar{a}_{z_i}}\right] \tag{5.20}$$

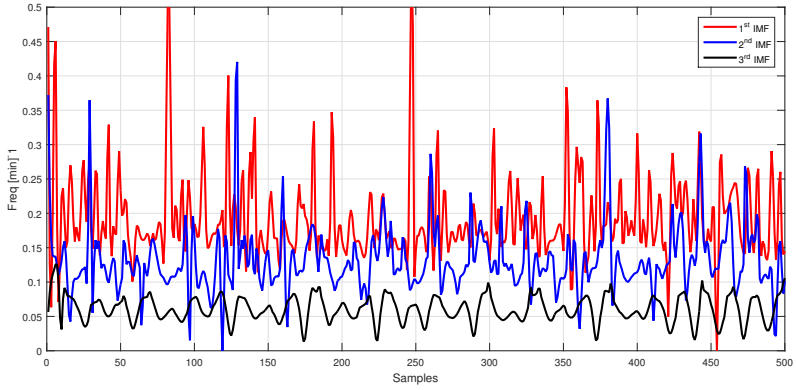

Figure 5.7: Instantaneous frequency plot for IMFs in Figure 5.6.

where a_z is the zero crossing amplitude; defined as the absolute value of the extrema between successive zero crossings, $\bar{a}_z$ is the mean of a_z. Equation (5.20) gives the extent of Non-Linearity in individual IMF; but the **Total Degree of Non-linearity (TDNL)** for the complete signal consisting of N IMFs can be given as sum of individual DNLs weighted by the energy of each IMF. The TDNL is given by (Huang et al. 2014, Aftab et al. 2016).

$$TDNL = \sum_{j=1}^{N} \left[DNL_j \frac{|c_j|^2}{\sum_{k=1}^{N} |c_k|^2} \right] \tag{5.21}$$

Here $|c_j|^2$ is the 2-norm of the j^{th} IMF.

5.5 Isolating the Source of Non-Linearity

Once the control loops suffering from the non-linearity induced oscillations are identified, the next step is to isolate the source for targeted maintenance and remedial actions. In a multi loop environment non-linearity induced oscillations at one point may propagate to other variables so the correct diagnosis will reduce the shut down time and cost of repair. Thornhill (2005) pointed out that the different parts of a plant tend to behave as low pass mechanical filters and thus filter out the higher harmonics as we move away from the source of non-linearity.

The TDNL measure discussed in section 5.4.3 can be used to compare the extent of non-linearity in different variables. The loop with maximum TDNL value is taken to be the source of non-linearity. In section 5.8, the proposed method is illustrated using the industrial data and results are compared with the existing methods.

5.6 Proposed Method

The detailed steps for the proposed method for the identification and isolation of the non-linearity induced oscillations are listed below.

Step I Follow steps given in Table 5.1 to enforce the dyadic filter bank property.

Step II Retain the IMFs that are correlated to the original signal and are sparse as per the criterion in Section 5.3.2.

Step III Calculate the ACF of the retained IMFs.

Step IV Identify the zero crossings of individual ACFs and calculate the mean period of oscillation $\bar{T}_p$ and regularity statistics r.

Step V Report the $\bar{T}_p$ of all the IMFs with $r > 1$.

Step VI Check for the presence of harmonics using the steps described in Section 5.3.5.

Step VII Report non-linearity if harmonics are present.

Step VIII Compute the DNL and TDNL measures of the IMFs representing fundamental and harmonics

Step IX The loop with maximum TDNL value is taken as the source of non-linearity induced oscillation.

Step X In case of multiple sources of non-linearity induced oscillations, steps VII to IX are repeated for each fundamental/ harmonics pair.

5.7 Simulation Example

The simulation example is taken form our recent work Aftab et al. (2016); where oscillations from different sources in a SISO feedback system are analyzed using the proposed method to look for signatures of harmonics or non-linearities. The feedback system is shown in Figure 5.8. The non-linearity is modelled by stiction using the two parameter model by Choudhury et al. (2008) with $S = 7$ and $J = 5$. The plant dynamics are given by

$$G(s) = \frac{2.25}{4.54s + 1} e^{-3s} \tag{5.22}$$

Nominal PI controller gains are $K_c = 0.1$ and $K_i = 0.05$. Noise with variance 0.1 is added to test the robustness of the proposed scheme.

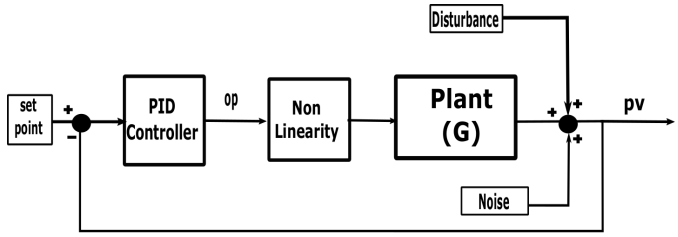

Figure 5.8: Closed loop system (simulation example).

In total five test cases, using plant output or process variable (pv) data, are considered for analysis and the results are explained next. It is to be highlighted that the transient effects are included in the analysis that induce non-stationarity. Moreover, a special test case with time varying drift is also introduced to test the robustness of the proposed scheme against stronger non-stationarity. The results for all test case are summarized in Table 5.5. The DNL and TDNL measures for the simulation example are given in Table 5.6.

5.7.1 External Disturbance

In this case the feedback system is subjected to an external sinusoidal disturbance. The system response is shown in the first row of Figure 5.9. The results show absence of harmonics thereby attributing the oscillations to a linear cause.

5.7.2 External Disturbance and Poor Tuning

In this scenario the oscillations resulting from the combination of poor controller tuning and external sinusoidal disturbance are analyzed. The time trend for this case is shown in the second row of Figure 5.9. The results show absence of any harmonics thus concluding that oscillation are caused by linear source.

5.7.3 Non-Linearity/Stiction

In this case oscillations induced due to non-linearity effects, modelled by stiction in the control valve, are analyzed. The time trend for this scenario is shown in the third row of Figure 5.9. The results clearly indicate presence of harmonics in the oscillatory signal. It is to be noted that the oscillations are characterized by odd harmonics as third (3^{rd}) harmonic is identified along with the fundamental. This case of stiction induced oscillation cannot be captured by the HOS based non-linearity detection method. The value of Non Gaussianity Index $NGI = -0.01$ confirms this fact.

5.7.4 Stiction and External Sinusoidal Disturbance

In this case an external sinusoidal disturbance is added to the system with stiction non-linearity. The same case is reported in the illustrative example of Section 5.3.5. The results clearly show that the proposed method has been able to separate the external sinusoidal disturbance from the stiction induced oscillation. Moreover the proposed rules make it possible to treat the sinusoidal disturbance as separate oscillation from the stiction.

5.7.5 Sinusoidal Disturbance with Time Varying Set Point

In order to test the robustness of the proposed scheme against non stationary effects, the system is subjected to a time varying set point in the form of ramp signal and an external sinusoidal disturbance. The system response, given in the last row of Figure 5.9, clearly shows the non-stationary effects. The proposed scheme separates out the sinusoidal disturbance from the time varying component, with the latter extracted in the residue $b(t)$ (5.2). The results for this case show the presence of only one oscillation void of any harmonics.

Table 5.5: Non-linearity detection simulation example

Case	IMF	λ	$\bar{\Omega}$	Ω_{min}	Ω_{max}	Harmonics	Type
Ext Dist	1	1.0	0.0159	0.0156	0.0163	No	Linear
	2	0.35	0.010	0.0085	0.013		
Ext Dist + Poor Tuning	1	1.0	0.083	0.083	0.083	No	Linear
	2	0.5	0.0159	0.0157	0.0162		
Stiction	1	0.54	0.0353	0.027	0.0506	Yes (3^{rd})	Non-Linear
	2	1.0	0.0135	0.0134	0.0137		
Ext Dist Plus Stiction	1	0.53	0.0348	0.0278	0.0466	Yes (3^{rd})	Non-Linear
	2	1.0	0.0134	0.0132	0.0137		
	3	0.30	0.0067	0.0062	0.0072	No	Linear
Ext Dist + Time Varying Set Point	1	1.0	0.0159	0.0156	0.0161	No	Linear

5.7.6 Robustness with Increasing Noise Variance

The robustness of the proposed method against noise levels is studied by varying the noise variance for the five simulation test cases. The noise variance σ_v^2 is varied from $\sigma_v^2 = 0.2$ to $\sigma_v^2 = 1.0$ in steps of 0.2. A graphical representation of the results are provided in Figure 5.10. It can be seen that proposed method is quite robust and can work well even with increased noise variance. In only two cases the reported results are erroneous, both occurring at a large noise variance of $\sigma_v^2 = 1.0$ with SNR of around 2dB.

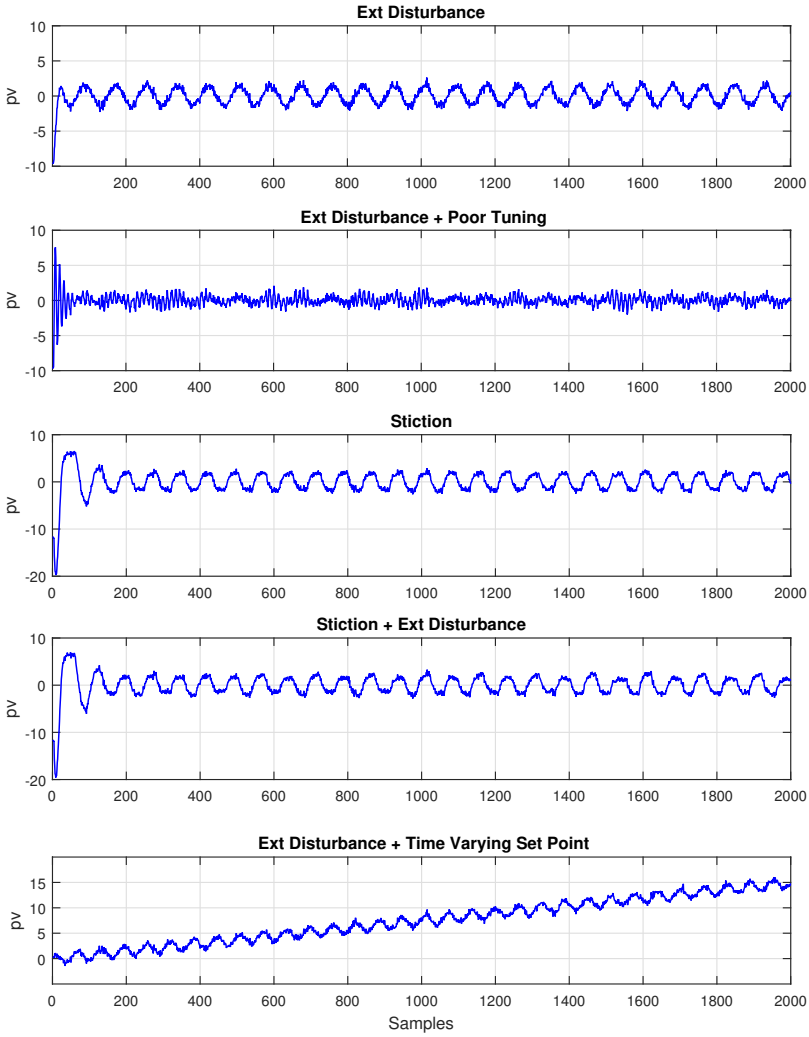

Figure 5.9: Time trends for simulation example.

Table 5.6: DNL and TDNL for simulation example

Case	Type	IMF	DNL	$TDNL$
Ext Dist	Linear(no harmonics)	–	–	–
Ext Dist Poor Tuning	Linear (no harmonics)	–	–	–
Stiction	Non-Linear	1 2	0.29 0.1	0.11
Stiction + Ext Dist	Non-Linear	1 2	0.33 0.1	0.11
Ext Dist + Time Varying Set Point	Linear(no harmonics)	–	–	–

5.8 Industrial Case Study

The case study is taken from Thornhill (2005) and Zang and Howell (2005), where a group of variables from a South East Asian refinery are found to be oscillating with the same fundamental frequency of 0.06 min $^{-1}$. It has been reported that the plant-wide oscillation stems from the the sticking valve in one of the control loops. Thus the aim is to find the variables that exhibit non-linearity induced oscillations and hence can be regarded as the source of oscillations. The results of the proposed method are compared with those obtained by the HOS based method and surrogate based methods for the same case study. The time trends (plotted in Figure 5.11) recorded at 1 minute sample rate, are used for the analysis.

5.8.1 Non-Linearity Detection in Individual Loops

The group consisting of 12 variables (shown in first column of Table 5.7) is analyzed using the proposed method and the results are summarized in Table 5.7; whereas comparison with HOS based method ((Choudhury et al. 2008),(Choudhury 2006)) and surrogate based method (Thornhill 2005) are given in Table 5.8.

The proposed method has identified four loops containing the signatures of non-linearity, namely tags 11, 24, 33 and 34. In addition to the identification of harmonics the proposed method has been able to identify the presence of multiple oscillations in different tags. For instance low frequency oscillations, other than plant-wide oscillation, in tags 11, 19 and 25 are also detected.

5.8.2 Isolating the Source of Non-Linearity

Once the variables oscillating due to a non-linear cause are identified the next logical step is to isolate the non-linearity to reduce the critical maintenance and shut down time. As discussed in section 5.5 the non-linearity signature diminishes as we move away from the source of the non-linearity due to mechanical filtering

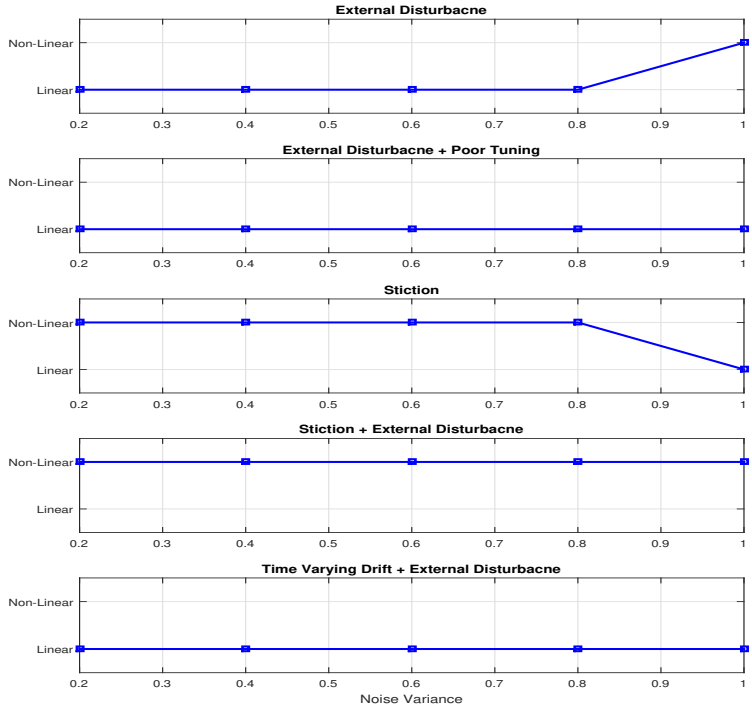

Figure 5.10: Performance of proposed method for increasing noise levels.

of the higher order harmonics. Therefore the loop with the highest value of the non-linearity measure is candidate for the source of non-linearity.

The TDNL measure given in Table 5.8 shows that tag 34 exhibits greatest non-linearity and hence is the most probable candidate for the root cause. Similar results are reported by both HOS based method Choudhury (2006) and surrogate method Thornhill (2005). The proposed method places tag 33 as the second most non-linear one due to greater energy of higher order harmonics (same as in HOS based method); whereas the surrogate analysis ranks it lower than tags 11 and 13.

The provision of flow diagrams and P&IDs can make the analysis more accurate by taking considerations based on the physical structure of the plant into account. However, some companies may be reluctant to make such documentation available to external consultants, and the aim of this work has been to develop an analysis technique based on the on-line measurements alone.

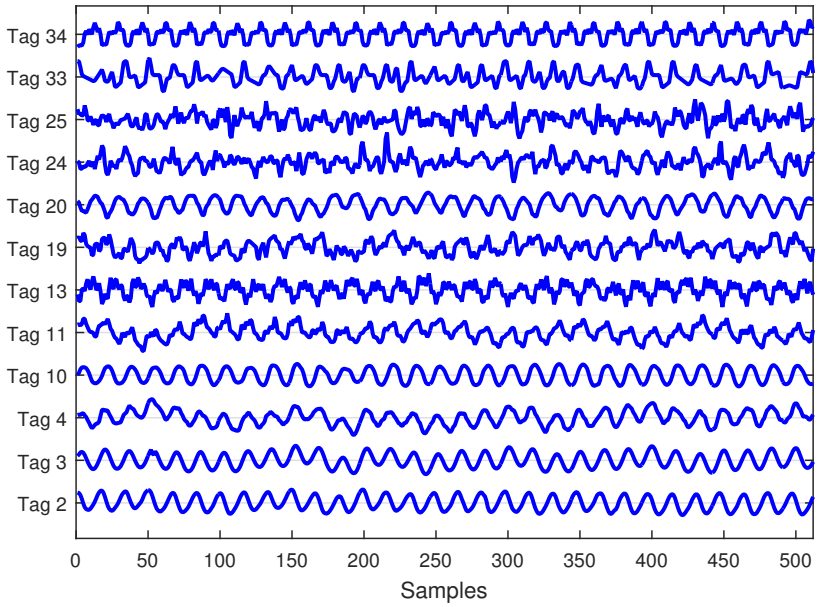

Figure 5.11: Time trends for the group of variables oscillating with frequency 0.06 min^{-1} for SEA refinery.

5.8.3 Effect of Changing Correlation and Sparseness Thresholds

The effect of changing the thresholds both for normalized correlation coefficient η and sparseness index S_{Thresh} are analyzed and the results for both cases are given in Figures 5.12 and 5.13 respectively. The correlation threshold is varied from 0.15-0.35 whereas the sparseness threshold is varied from 0.55-0.62. It can be seen that the results are mostly stable with respect to these variations. The only exception are Tags 11 and 24 whose harmonic content is lost if the sparseness threshold increases beyond 0.60.

5.8.4 Comparison with other Methods

The results are quite comparable with the surrogate based non-linearity detection method with the exception of only tags 13 and 24. Tag 13 comes out to be linear in the proposed method because the oscillations in the IMF representing the second harmonic were not regular; giving the regularity index $r = 0.7$. The reason is the presence of some other oscillations with the second harmonic in that frequency range. Tag 24, because of presence of harmonics, is declared non-linear, similar to the results of the HOS method which reports it to have quite significant non-

Table 5.7: Non-linearity detection industrial case study

Tag	IMF	λ	$\bar{\Omega}$	Ω_{min}	Ω_{max}	Harmonics	Type
2	1	1.0	0.0604	0.0572	0.0641	No	Linear
3	1	1.0	0.0604	0.0572	0.0641	No	Linear
4	1	0.49	0.0696	0.0548	0.0954	no	Linear
	2	1.0	0.0604	0.0572	0.0641		
	3	0.32	0.0201	0.0181	0.0226		
10	1	1.0	0.0604	0.0572	0.0641	no	Linear
11	1	0.30	0.1833	0.1565	0.2212	Yes	Non-Linear
	2	0.47	0.1222	0.0993	0.1589		
	3	1.0	0.0604	0.0572	0.0641		
	4	0.59	0.0120	0.0104	0.0142		
13	1	0.66	0.2750	0.2250	0.3537	No	Linear
	2	1.0	0.0604	0.0572	0.0641		
19	1	1.0	0.0604	0.0572	0.0641	No	Linear
	2	0.49	0.0163	0.0129	0.0219		
	3	0.41	0.0121	0.0111	0.0132		
20	1	1.0	0.0604	0.0572	0.0641	No	Linear
24	1	0.80	0.1122	0.1005	0.1272	Yes	Non-Linear
	2	1.0	0.0604	0.0572	0.0641		
25	1	1.0	0.0604	0.0572	0.0641	No	Linear
	2	0.32	0.0275	0.0264	0.0287		
	3	0.29	0.0168	0.0139	0.0212		
33	1	0.75	0.122	0.1138	0.1319	Yes	Non-Linear
	2	1.0	0.0604	0.0572	0.0641		
34	1	0.42	0.2895	0.2279	0.3968	Yes	Non-Linear
	2	0.52	0.1222	0.1080	0.1408		
	3	1.0	0.0604	0.0572	0.0641		

linearity with NLI$=0.76$. The amplitude of tag 24 is very small and both tag 13 and tag 24 are considered ambiguous in the analysis by Zang and Howell (2005) as well.

The HOS based method declares all the loops with the exception of tag 25 as non-linear. This finding is substantiated by neither the method proposed in this paper nor the surrogate analysis. Thus apart from not being able to capture the non-linearities characterized by odd harmonics, the HOS based method appears to classify too many signals as non-linear.

As far as the comparison with harmonic distortion factor by Thornhill et al. (2001) is concerned, it was pointed out by the authors themselves that the method is based on manual detection of harmonics and can be misleading in presence of noise and multiple oscillations. Tag 25 is shown to have high distortion factor owing to the presence of noise. This shortcoming is totally removed in the proposed method

Table 5.8: Comparison of proposed and existing non-linearity measures

Tag	Proposed Method			HOS			N
	IMF	DNL	TDNL	NGI	NLI	TNLI	(surrogate)
Tag 2	–	–	–	0.15	0.99	2.71	–
Tag 3	–	–	–	0.14	0.94	2.65	–
Tag 4	–	–	–	0.06	0.81	0.81	–
Tag 10	–	–	–	0.04	0.79	0.79	–
Tag 11	1	0.32	0.09	0.20	0.96	2.84	2.74
	2	0.22					
	3	0.06					
Tag 13	–	–	–	0.15	0.96	1.80	2.64
Tag 19	–	–	–	0.13	0.88	0.88	–
Tag 20	–	–	–	0.13	0.94	1.76	–
Tag 24	1	0.197	0.12	0.01	0.76	0.76	–
	2	0.099					
Tag 25	–	–	–	0.0	0.0	0.0	–
Tag 33	1	0.23	0.13	0.080	0.87	3.32	2.57
	2	0.09					
Tag 34	1	0.53	0.31	0.20	0.99	7.63	4.91
	2	0.27					
	3	0.27					

where automatic harmonic detection is performed and tag 25 is correctly diagnosed as not containing any non-linearity signatures.

5.9 Conclusions

This paper presents a method, based on noise assisted MEMD, to identify the harmonics and hence non-linearity in control loops. The method is adaptive in nature with no *a priori* assumptions about the underlying signal or the process itself. The method can extract harmonics in the presence of noise and multiple oscillations and hence identify the presence of non-linearity. The extent of non-linearity is then calculated via intra-wave frequency modulation and quantified using the total degree of non-linearity measure (TDNL). Results of the proposed method are compared with the HOS and surrogate based methods and are comparable with the surrogate analysis method both in terms of detection and isolation of non-linearity.

5.10 Acknowledgements

This work is being funded by Siemens AS, Norway. The fruitful discussions with Prof Nina F. Thornhill (Imperial College London) regarding the calculation of $N-$ measure are gratefully acknowledged. Moreover, the authors would like to thank Prof A.K. Tangirala (IIT Madras) for the provision of industrial data.

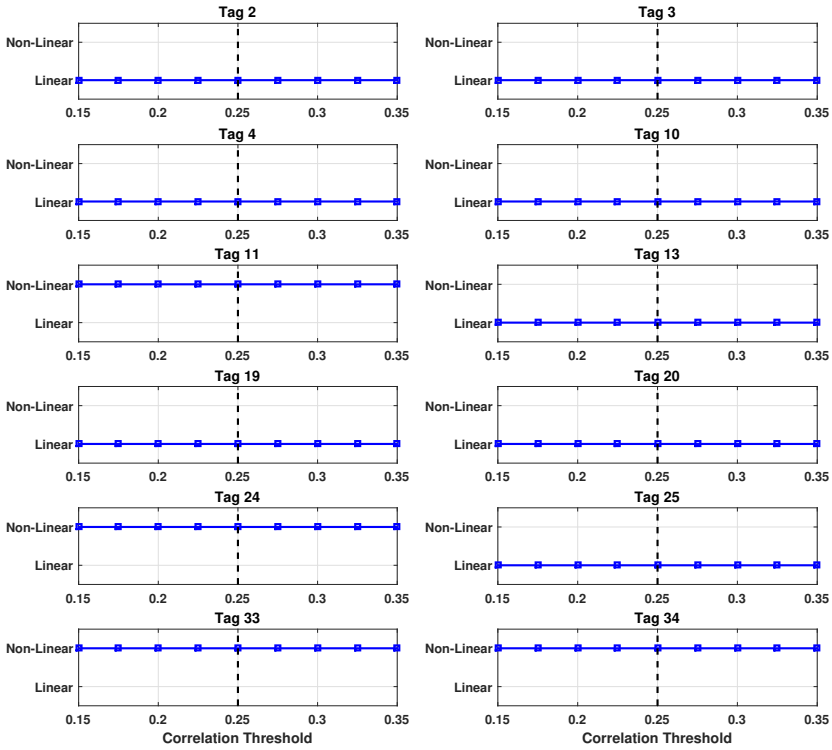

Figure 5.12: Effect of changing correlation threshold (η) for industrial case study; vertical dashed line shows default value.

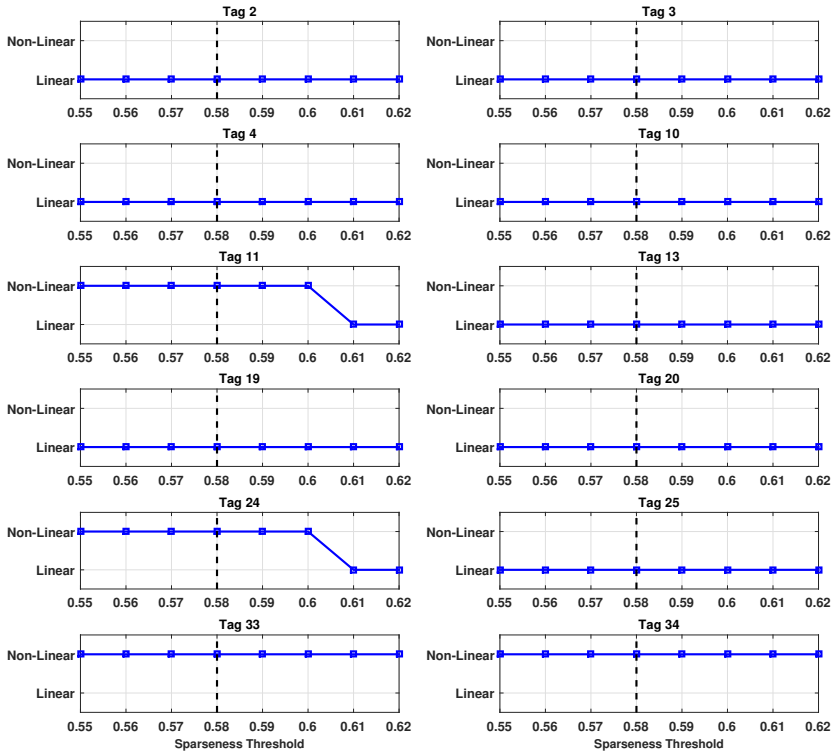

Figure 5.13: Effect of changing sparseness index threshold (S_{Thresh}) for industrial case study; vertical dashed line shows default value.

Chapter 6

Combination of Delay Vector Variance and Multivariate Empirical Mode Decomposition

Paper D: Diagnosis of Plant-Wide Oscillations by Combining Multivariate Empirical Mode Decomposition and Delay Vector Variance

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

Journal of Process Control (Submitted 2017)

Abstract

A combination of multivariate empirical mode decomposition (MEMD) and delay vector variance (DVV) method is explored for the diagnosis of the plant-wide oscillations. The proposed algorithm improves upon the already presented method based solely on the DVV approach. The proposed modification makes it possible for the DVV method to work not only for the non-stationary data but also the case where multiple oscillations sources are present in the control loop. The focus of this work has been to distinguish between linear and non-linear causes of oscillations and isolating the source of non-linearity. A noise-assisted multivariate empirical mode decomposition is used to sift out different oscillating components in the plant-wide data and the nature of the oscillation is then ascertained using the already presented DVV approach. In case of oscillations due to non linear causes the method can isolate the source using the extent of non-linearity. The method is tested on simulated and industrial case studies and results are compared with

existing methods based on higher order statistics (Choudhury et al. 2008) and surrogate based methods (Thornhill 2005).

6.1 Introduction

Diagnosis of plant-wide oscillations has gained considerable research attention in the past decade. There are hundreds of control loops running in a large scale industrial plant. A disturbance or oscillation originating at one place tends to propagate to other parts of the plant owing to underlying interactions and process flows. The plant-wide oscillations, if allowed to continue unchecked, can have detrimental effects on the plant operation. Therefore the diagnosis of plant-wide oscillation is an essential part of any control performance monitoring (CPM) paradigm.

An important aspect of plant-wide oscillation diagnosis is the ability to classify the oscillations arising due to non-linear and linear causes. This is so because non-linearity induced oscillations are considered as one of the major performance issues in industrial control systems (Srinivasan et al. 2005a). Both model based and data driven techniques are utilized in literature to achieve this task. However, the data driven techniques are more popular and flexible as they need minimal process knowledge and can work in the absence of plant model, that is seldom known accurately. The popular data driven methods include cross correlation method (Horch 1999), higher order spectra (HOS) (Chaudhry et al. 2004, Zang and Howell 2003), Lyapunov exponents (Zang and Howell 2005). Shape analysis formalism has also been used for detecting valve non-linearities (Srinivasan et al. 2005a, Hägglund 2011, Yamashita 2005). Aftab et al. (2016) and Babji et al. (2009) applied methods based on the Hilbert Huang Transform (HHT) to identify non-linearities in the control loop data.

Methods based on the phase space reconstruction via time delayed embedding (Takens's theorem Takens (1981)) are also proposed (Thornhill 2005), (Aftab et al. 2017b) to indicate the presence of non linearity as the source of plant-wide oscillations. These methods originate from the non linear time series analysis, where an attractor manifold $\mathbf{M}$ is constructed using the time series data from the control loops under analysis. Similar manifolds $\mathbf{M}_s$ are constructed for the surrogate counter parts, where surrogates are time series that exhibit the same power spectrum as the original time series but with randomized phase. The difference in the predictability of the surrogate and the original time series indicates the presence of non linearity in the control loop. An important limitation of these surrogate based methods is the assumption of stationarity in the loop data which cannot be guaranteed in presence of time varying disturbances and changing set points.

The present work is aimed at removing this hurdle caused by the assumption of

stationarity so that the scope of analysis can be broadened to time varying signals while keeping the advantages of surrogate based methods intact. The said objective is achieved by incorporating the noise-assisted multivariate empirical mode decomposition (NA-MEMD) (Rehman and Mandic 2011) as a pre processing step for the DVV method. The dyadic filter bank property of NA-MEMD makes it possible to sift out different frequency components from the control loop data as well as the time varying components. The output of NA-MEMD then can processed using the DVV analysis to ascertain the presence or absence of the non-linearity in the specific variable. Similarly the case of multiple oscillations due to combination of linear and non-linear causes can also be analysed. The NA-MEMD has also been used to assess the harmonic content of the control loop data, as the indication of non-linearity, in our previous work Aftab et al. (2017c).

The proposed changes in the previous work Aftab et al. (2017b) make it possible to process non stationary data under the surrogate analysis framework. Moreover, the case of multiple oscillations originating from different sources can also be handled.

The paper is organized as follows. Section 6.2 gives and overview of the EMD, Multivariate EMD and associated dyadic filter bank property. Section 6.3 outlines the steps involved in phase space reconstruction, delay vector variance (DVV) method and testing the Null hypothesis using rank based statistics. Automatic determination of embedding dimensions is explained in section 6.4. Section 6.5 gives the detailed algorithm, followed by simulation and industrial case studies in section 6.6 and 6.7 respectively followed by conclusions.

6.2 Multivariate Empirical Mode Decomposition (MEMD)

6.2.1 Empirical Mode Decomposition (EMD)

EMD caters for the univariate (consisting of one variable) time series data and is a data driven procedure that adaptively sifts out the fast modes (high frequency) from slower (low frequency) ones, in a signal, iteratively. The components so obtained are termed as intrinsic mode functions (IMFs). An IMF is a zero mean signal in which number of extrema and zero crossings at most differ by one. The sifting process is carried out by identifying the extrema in the signal and forming an envelope via cubic splines. The local mean of the envelope for time series $x(t)$ gives the slow frequency signal $m(t)$ while the fast frequency signal $d(t)$ is formed by [Rilling et al. (2003), Huang et al. (1998)]

$$d(t) = x(t) - m(t) \tag{6.1}$$

This process is iterated until $d(t)$ is an IMF. The fast component $d(t)$ is then subtracted from the $x(t)$ and the procedure is iterated again on the residue and contin-

ues till no more IMFs are left to be extracted. The outcome of EMD process with N IMFs $c_i(t)$ and and residue $b(t)$ is given by

$$x(t) = \sum_{i=1}^{N} c_i(t) + b(t) \tag{6.2}$$

6.2.2 Multivariate EMD (MEMD)

The multivariate empirical mode decomposition (MEMD) is the extension of the univariate case to the n-dimensions i.e. for signals comprising n variables. The generalization to the $n-$dimension necessitates the generation of envelope in the n-dimensional space, identification of extrema and calculation of the associated local mean. Rehman and Mandic (2010a) proposed that the envelopes in the n-dimensional space can be generated by extrema sampling of multiple signal projections in n-dimensional space. The projections are calculated using direction vectors from the center of the unit n-dimensional sphere to the uniformly distributed points on its surface. The uniform sampling of the unit nn-dimensional sphere is accomplished using Halton Sequence.

The the kth sample of the q-dimensional, Halton sequence is given by

$$\left(\Phi_{p_1}, \Phi_{p_2}, \ldots \Phi_{p_q}\right) \tag{6.4}$$

where $p_1, p_2 \ldots p_q$ are q prime numbers and k has the prime-base-p representation given by (Wong et al. 1997)

$$k = a_0 + a_1 p + a_2 p^2 \ldots a_r p^r \tag{6.5}$$

where each a_i is an integer in $[0, p-1]$ and $\Phi_p(k)$ is written as

$$\Phi_p(k) = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} \ldots \frac{a_r}{p^{r+1}} \tag{6.6}$$

The envelope is generated for all the projection curves, using cubic splines, and are averaged to calculate the $n-$dimensional mean envelope $\mathbf{m(t)}$. The mean is subtracted from the $n-$dimensional signal $\mathbf{X(t)}$, using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$, to extract the fast oscillation as in univariate case. The iterative procedure goes on till all IMFs (n-dimensional) are extracted. The detailed algorithm is given in the Table 6.1.

6.2.3 Dyadic Filter bank Property of MEMD

The MEMD effectively behaves as the sequence of band pass filters in the presence of white noise similar to the wavelet decomposition effect (Rehman and Mandic 2011). This property is referred to as the dyadic filter bank characteristics of

Table 6.1: Multivariate empirical mode decomposition algorithm

Algorithm Multivariate EMD

Step I Consider a multivariate n-dimensional signal $\mathbf{X}$ given in equation 2.3.

Step II Setup K direction vectors u^k with $k = 1 \dots K$ by choosing uniformly spaced points on n dimensional sphere.

Step III Find the projections $p^k(t)$ of the input signal $\mathbf{X}$ along the direction vectors u^k using

$$p^k(t) = \mathbf{X}u^k \quad \forall \quad k = 1 \dots K$$

Step IV Identify the maxima of projections $p^k(t)$ and corresponding time instants t^k.

Step V Generate n-dimensional envelope curve $\mathbf{e}^k(t)$ by interpolating $[t^k, \mathbf{X}(t^k)]$.

Step VI The n-dimensional mean of envelope curve is then given by

$$\mathbf{m} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{e}^k(t) \tag{6.3}$$

Step VII Extract n-dimensional detail $\mathbf{d(t)}$ using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$.

Step VIII Repeat the steps II-VII using $\mathbf{d(t)}$ as input till it fulfils the criteria for an IMF.

Step IX Calculate the residue $\mathbf{b(t)} = \mathbf{X(t)} - \mathbf{d(t)}$ and iterate the procedure on $\mathbf{b(t)}$ till there are no more IMFs left to be extracted.

MEMD. In order to elaborate this concept an average power spectrum of IMFs from 1000 different realizations of three-channel (three variables) white Gaussian noise is shown in Figure 6.1 (Aftab et al. 2017c). The IMFs can be seen as the output of the series of band pass filters, with frequency of each band decreasing with the IMF index.

The dyadic filter bank property of the MEMD can be enforced by adding two or more noise channels or sequences with the input multivariate signal (Rehman and

Mandic 2011), the so called noise assisted multivariate empirical mode decomposition (NA-MEMD). The addition of noise channels would tend to arrange the IMFs of the input signal according to the dyadic filter bank structure as shown in Figure 6.1. The IMFs corresponding to the noise channels are then discarded while the ones corresponding to the input signal are retained for analysis. In order to extract the different oscillations from the signals suffering from plant-wide oscillations, all the times series data of measurements, affected by the plant-wide oscillations, are stacked to make a $n-$ dimensional multivariate signal X given by:

$$\mathbf{X} = \begin{Bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(l) & x_2(l) & \dots & x_n(l) \end{Bmatrix} \tag{6.7}$$

Where x_i is the data from the i_{th} measurement consisting of l samples. This multivariate signal $\mathbf{X}$ is then processed via NA-MEMD via following steps (Rehman and Mandic 2011).

Step I Generate two uncorrelated white Gaussian noise sequences with k samples.

Step II Add the two noise sequences with the original signal $\mathbf{X}$ to make $n + 2$ variable/channel[1] signal, with n channels of original data and two channels of noise.

Step III Process the signal using MEMD. The resulting IMFs will have $n + 2$ channels.

Step IV Retain the IMFs in the channels corresponding to the original signal ($\mathbf{X}$), and discard the IMF components corresponding to noise channels.

Step V Each column of the multivariate signal $\mathbf{X}$ will have N IMFs. The IMFs for the individual channel will be then separated for further analysis.

6.2.4 Discarding Spurious and Noisy IMFs

Discarding Pseudo IMFs

The MEMD process may produce pseudo IMFs that are poorly correlated with the input signal due to spline fitting issues (Peng et al. 2005, Srinivasan and Rengaswamy 2012, Aftab et al. 2016; 2017c). Therefore it is important retain only

[1] here channel is same as variable

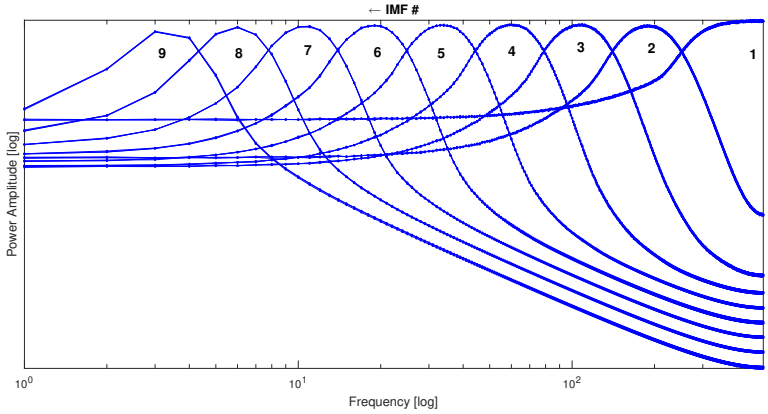

Figure 6.1: Dyadic filter bank property of multivariate EMD (MEMD) (Aftab et al. 2017c).

those IMFs that have good correlation with the original signal. The correlation coefficient for each IMF (c_i) with input signal x is given by

$$\rho_i = \frac{Cov(c_i, x)}{\sigma_x \sigma_{c_i}}, \quad i = 1, 2, 3 \ldots N \tag{6.8}$$

where Cov denotes the covariance; σ_x and σ_{c_i} are the standard deviations of the signal and the IMF, respectively, and N is the total number of IMFs. The normalized correlation coefficient λ_i for each IMF is given by:

$$\lambda_i = \frac{\rho_i}{max(\rho_i)}, \quad i = 1, 2, 3 \ldots N \tag{6.9}$$

Only the IMFs with ($\lambda > \eta$) are retained, where the value of threshold η used in this work is discussed in subsequent sections.

Discarding Noisy IMFs

In case of large noise content the IMFs containing only noise may have higher correlation with the input signal. These IMFs are removed using a sparseness index (Hoyer 2004) and (Srinivasan and Rengaswamy 2012, Aftab et al. 2017c) given by.

$$SI(x) = \frac{\sqrt{I} - \left(\sum_{i=1}^{I} |X_i| / \sqrt{\sum_{i=1}^{I} |X_i|^2} \right)}{\sqrt{I} - 1} \tag{6.10}$$

where X is the power spectrum and I is total number of frequency bins. The oscillatory signals will exhibit distinct peaks in frequency domain whereas the

noisy signals will have power spread across a broad frequency range. Equation (6.10) shows that the sparseness index SI of power spectrum $X(f)$ of signal $x(t)$, will be almost zero for noisy signal whereas it will attain the value near one for the oscillatory ones.

In order to analyse the IMFs for the signatures of non linearity using DVV method, only those IMFs are considered that have sparseness index SI greater than certain threshold SI_{Thresh}.

6.2.5 Default Parameter Settings

Correlation Threshold η

It is discussed in section 6.2.4 that MEMD process can generate spurious IMFs and only the one having significant correlation are retained for further analysis. The threshold η used in this work is 0.25. The detailed discussion about the selection of this threshold value is given in Section 5.3.4. Selection of this value ensures that enough harmonic content is retained for the non-linearity analysis.

Sparseness Threshold (S_{Thresh})

The IMFs extracted from the MEMD process are to be analysed by the DVV method for the presence or absence of non-linearity. Therefore, the IMFs having large noise content and little information are discarded via sparseness index. The IMFs having sparseness index greater than certain Sparseness Threshold (S_{Thresh}) are retained for the analysis. The Sparseness Threshold (S_{Thresh}) used in this analysis is 0.58. The detailed discussion about the selection of this threshold is given in Sections 4.5.3 and 5.3.4.

6.3 Delay Vector Variance (DVV) Method

The delay vector variance (DVV) method, proposed by Gautama et al. (2004a), identifies the linear or non-linear nature of the time series via time delayed embedding (Takens's theorem (Takens 1981)). The time series $\mathbf{x}(k) = \{x(k) \mid k = 1 \ldots N\}$ is converted to a state space representation by means of delay vectors (DVs) with the embedding dimension m and time delay τ, given by $\mathbf{y}(\mathbf{k}) = [x_{k-m\tau}, \ldots, x_{k-\tau}]$ (Kantz and Schreiber 2004). The next sample $x(k)$ of each delay vector is designated as the target (Gautama et al. 2004b) for calculation of target variance σ^* using the following steps (Gautama et al. 2004a;b, Aftab et al. 2017b).

Step I The distance between DVs is calculated using the Euclidean norm, denoted by $||\mathbf{y}(\mathbf{i}) - \mathbf{y}(\mathbf{j})||$ for $i \neq j$.

Step II The mean μ_d and standard deviation σ_d are computed for the calculated DV distances.

Step III All the DVs that lie within a certain distance τ_d are collected in sets Ω_k , written mathematically as

$$\Omega_k = \{\mathbf{y(i)} \mid \|\mathbf{y(i)} - \mathbf{y(j)}\| \leq \tau_d\} \tag{6.11}$$

Step IV The threshold τ_d is taken from the uniformly sampled interval $[\mu_d - n_d\sigma_d; \mu_d + n_d\sigma_d]$; with n_d specifying the span over which the analysis is performed.

Step V The target variance σ_k^{*2}, is calculated for each set Ω_k that contains at least N_0 DVs.

Step VI The average of these target variances σ_k over all sets divided by the variance of time series σ_x^2 gives the overall target variance σ^{2*}, given by (6.12).

$$\sigma^{*2} = \frac{(1/K)\sum_{k=1}^{K}\sigma_k^{*2}}{\sigma_x^2} \tag{6.12}$$

The target variance σ^* is the inverse measure of the predictability of the time series. The target variance is plotted against the standardized distance measure r_d given by

$$r_d = \frac{\tau_d - \mu_d}{\sigma_d} \tag{6.13}$$

Target variance not only gives an indication of determinism in the time series but also helps in classifying the nature of time series. Moreover, as for larger spans the set Ω_k contains all the DVs, the target variance converges to the variance of the time series i.e. converges to unity. The span parameter n_d needs to be increased if this is not the case. Figure 6.2 illustrates this effect graphically for two data sets, a non linear case (square wave) and a linear case (sine wave).

6.3.1 Detection of Non-linearity using DVV Method

The DVV method makes use of the surrogate framework to identify the signatures of non-linearity in time series. The surrogate time series have the same power spectrum as the original time series but with random phase. This phase randomization destroys any kind of phase coupling present due to the non-linearity. Hence the difference in the target variances of the original signal and the surrogate counterpart is taken as an indication of non-linearity. For illustration purposes target

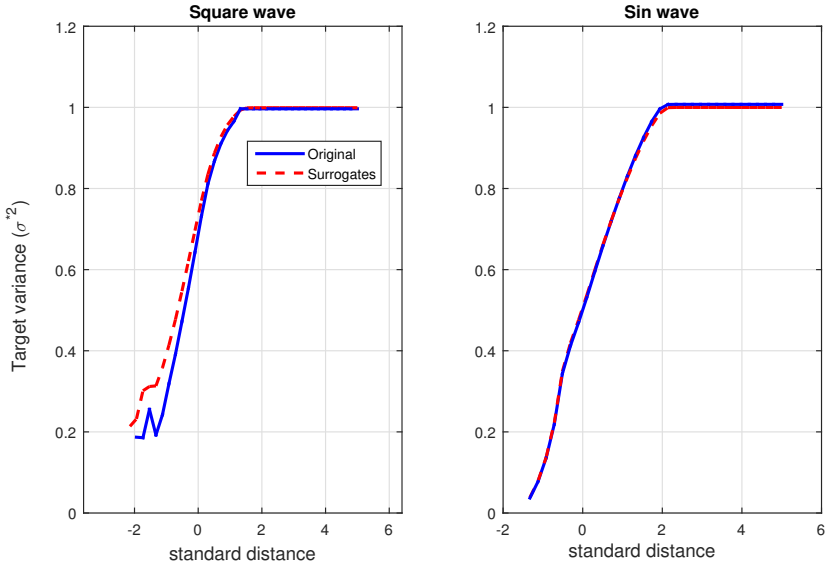

Figure 6.2: An overview of the target variance σ^{*2}.

variances of two signals are compared in Figure 6.2. The signal with non linearity (square wave) shows difference in its target variance from its surrogate pair, whereas the target variances are the same for both in case of linear signal (sine wave).

DVV scatter plot provides an another visual tool to compare the target variances of the signal and its surrogates. The DVV scatter diagram plots target variance of original signal (horizontal axis) against the target variance of surrogates (vertical axis). In case the target variances of signal and surrogates are same then all the points will confine to the bisector line while the plot will deviate from this bisector line in case of difference in target variance. The scatter plot of the two cases, that of square wave and sine wave are shown in Figure 6.3. In case of square wave (non-linearity) the difference in the target variances of the signal and its surrogate leads to diversion of scatter plot from bisector line, whereas for the sine wave (linear case) both target variances are same and thus conform closely with the bisector line.

6.3.2 Measuring the Extent of Non-linearity

The difference in the target variance of original signal and its surrogate counter part can be used to quantify the extent of non-linearity in the signal. Therefore the root mean square error e_{rms} (RMSE), given in Equation 6.14, between the target

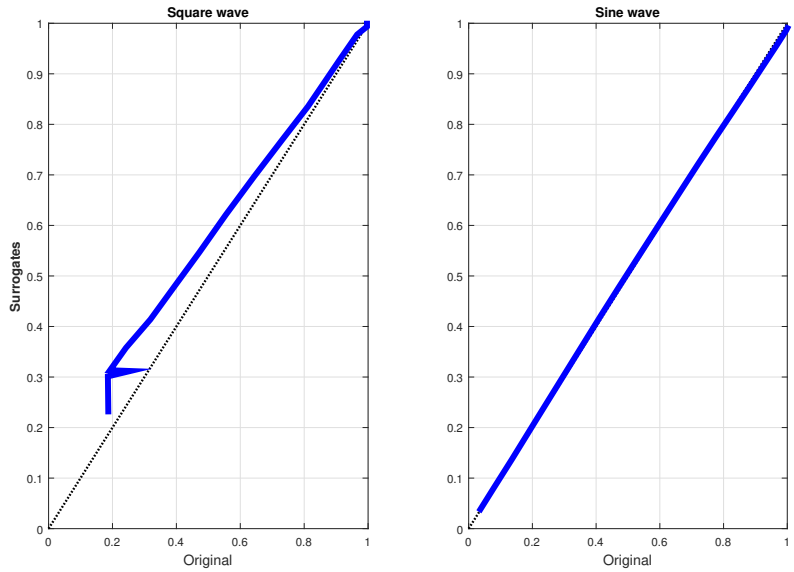

Figure 6.3: DVV scatter plot for non-linear (square wave) and linear (sine wave) case.

variance of input signal σ^{*2} and average target variance of surrogates ($\bar{\sigma}_s^{*2}$) gives the numerical measure of the extent of non-linearity. Greater the RMSE greater will be the non-linearity and vice versa.

$$e_{rms} = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\sigma^{*2} - \bar{\sigma}_s^{*2})^2} \tag{6.14}$$

where J are total number of target variances and $\bar{\sigma}_s^{*2}$ is the average target variance of B surrogates.

6.3.3 Isolating the Source of Non-Linearity

The source of the non-linearity can be identified using the fact that the signatures of non-linearity are stronger near the source and gets weaker away from origin due to low pass filtering characteristics of different plant components Thornhill et al. (2001). Thus the maximum value of the RMSE measure, discussed in section 6.3.2, will be closest to the source of the non-linearity.

Table 6.2: Rank statistics for rejecting NULL hypothesis Aftab et al. (2017b)

if rank $> (B+1)(1-\alpha)$
then reject Null hypothesis (non-linear source)
else
accept Null hypothesis (linear source)

Table 6.3: Rank statistics and RMSE

Signal	rank	Type	RMSE
sine wave	17	Linear	–
square wave	99	Non-Linear	0.047

6.3.4 Testing of the Null Hypothesis: The Test Statistics

The NULL hypothesis that the time series is an outcome of a linear Gaussian process has been tested via non parametric rank based statistics. Kantz and Schreiber (2004) had recommended the non parametric statistics on the pretext that the probability distribution of the non-linearity measure is not known in advance. RMSE has been chosen as a discriminating statistics with μ_0 being the RMSE value for the original time series. Then μ_k (RMSE of surrogates) is calculated for B surrogates with $k = 1, \ldots, B$. For one sided test the null hypothesis rejected at level α, if η_0 is among largest $(B+1)\alpha$ in the sorted list containing μ_0 and all $\mu_k s$.

In this work, to test the Null hypothesis, $B = 99$ surrogates and $\alpha = 0.1$ are chosen. Therefore the Null hypothesis is rejected if rank of μ_0 is greater than 90 in the sorted list of all μ_{ks}. For illustration purposes the rank statistics for two cases a linear and non-linear generated by sine and square wave respectively are tabulated in Table 6.3.

6.4 Determining the Embedding Dimensions

The phase space reconstruction procedure, used in DVV method, makes use of the time delayed embedding to construct an attractor manifold $\mathbf{M}$. The most important parameter in this process is the embedding dimension m of the attractor manifold. Too small m will mask the dynamics, whereas larger than required value will result in unnecessary computation overload. An automatic determination of the embedding dimension from Cao (1997) is used in this work to calculate the minimum required embedding dimension.

This method makes use of the fact an embedding dimension m is said to be true embedding dimension if two points which are close in the m-dimensional recon-

structed space stay close in the $(m+1)$-dimensional reconstructed space (Sauer et al. 1991, Cao 1997). Therefore a distance measure between near neighbours in m and $m+1$ dimensional space is used to calculate the required dimension m.

For a time series $\mathbf{x(k)} = \{x(k) \mid k = 1 \dots N\}$ and i^{th} delay vector $y_i(m)$ with embedding dimension m is given by (Cao 1997, Aftab et al. 2017b)

$$y_i(m) = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}) \qquad i = 1, 2, \dots, N - m\tau \tag{6.15}$$

Now a distance measure $a(i,m)$ is defined as

$$a(i,m) = \frac{\|y_i(m+1) - y_{n(i,m)}(m+1)\|}{\|y_i(m) - y_{n(i,m)}(m)\|} \qquad i = 1, 2, \dots, N - m\tau \tag{6.16}$$

where $\|.\|$ represents vector norm operation; $y_i(m+1)$ is the i^{th} delay vector with $m+1$ dimensions and $n(i,m)$ is the set containing nearest neighbours of i_{th} vector in the m dimensional space.

The mean $E(m)$ of all $a(i,m)$s is defined as

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a(i,m) \tag{6.17}$$

The variation in E for the increase in embedding dimension to $m+1$ is given by the quantity E_1 defined as

$$E_1(m) = \frac{E(m+1)}{E(m)} \tag{6.18}$$

Therefore when E_1 stops changing for some $m \geq m_0$; $m_0 + 1$ is taken as the minimum embedding dimension for the reliable phase space reconstruction [Cao (1997)]. The E_1 measure for two time series a square wave and a sine wave is plotted in Figure 6.4. It can be seen that the E_1 approaches the unity value after some initial transients. In this work the embedding dimension is chosen once E_1 value is greater than 0.99 and stays that way for all the higher embedding dimensions.

6.5 Proposed Method

The proposed method for the diagnosis or plant-wide oscillations to identify non-linearity induced oscillations and isolation of the source using combination of NA-MEMD and DVV is listed below.

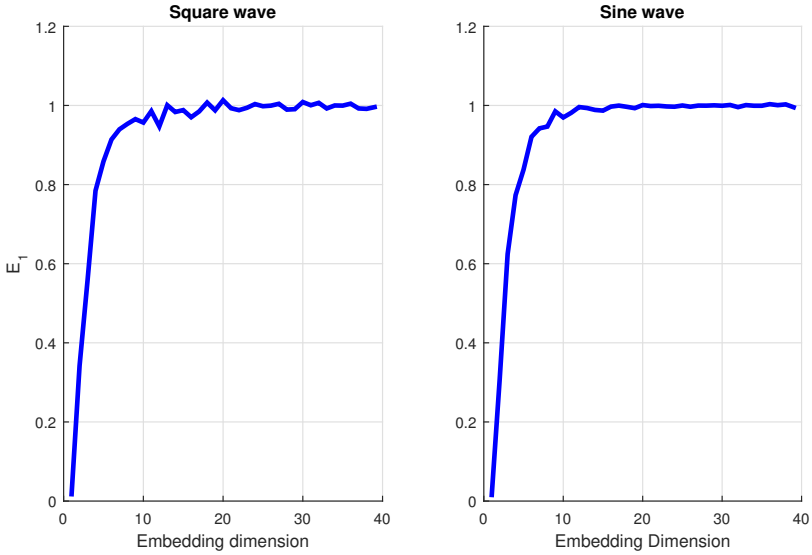

Figure 6.4: Determining the embedding dimensions.

Step I Make a multivariate signal comprising all the variables suffering from plant-wide oscillations.

Step II Process the multivariate signal using the NA-MEMD as discussed in section 6.2.3.

Step III Retain the IMFs fulfilling the correlation and sparseness criteria (section 6.2.4).

Step IV Mean center and normalize the all retained IMFs corresponding to each variable to unit standard deviation.

Step V End match the data to avoid spurious modes in the surrogate. End matching requires that the difference between both initial and final values d_0 and difference between initial and final gradient d_1 is minimized. d_0 and d_1 are calculated using (6.19) (Thornhill 2005).

$$\begin{aligned} d_0 &= \frac{(x_i - x_{i+n-1})^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2} \\ d_1 &= \frac{[(x_{i+1} - x_i) - (x_{i+n-1} - x_{i+n-2})]^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2} \end{aligned} \tag{6.19}$$

where x_i is the i^{th} element and $\bar{x}$ is the mean of time series $x_i \ldots x_{i+n-1}$.

Step VI Determine the minimum embedding dimension for the phase space reconstruction.

Step VII Perform the DVV analysis on normalized, end matched IMF using section 6.4.

Step VIII Generate B surrogates of the each IMF.

Step IX Perform DVV analysis on all the surrogates.

Step X Perform the rank based statistical test via procedure explained in section 6.3.4.

Step XI Loops with RMSE rank $r > (B+1)(1-\alpha)$ are classified as non-linear at confidence level α.

Step XII Compute the RMSE using the target variance of input data and average target variance of surrogates as given in section 6.3.2 (equation 6.14).

Step XIII Perform steps IV-XII for all the IMFs of each variable/measurement under analysis.

Step XIV Sum the RMSE of individual IMFs with rank ≥ 90 for each variable.

Step XV The measurement with maximum RMSE value is designated as the source of the non-linearity.

6.6 Simulation Example

The proposed method is applied to a simulation example where two non stationary time series, with a common oscillation at the frequency of 0.25 sec^{-1}, are considered. One of the time series (first row Figure 6.5) exhibit multiple oscillation i.e. square wave and a pure sinusoidal signal (0.075sec^{-1}) while the other (second row Figure) contains sinusoidal signal only. Both time series have a time varying drift in opposite directions along with a added white noise with $\sigma_v^2 = 0.2$. The square wave exhibits non linearity characterised by the presence of odd harmonics.

The four significant IMFs are obtained for the signal containing a combination of square and sine waves (Figure 6.6). The first and the second IMFs contain higher order harmonics whereas the fundamental harmonic is the primary oscillation in third IMF. The sine wave resides in the fourth IMF.

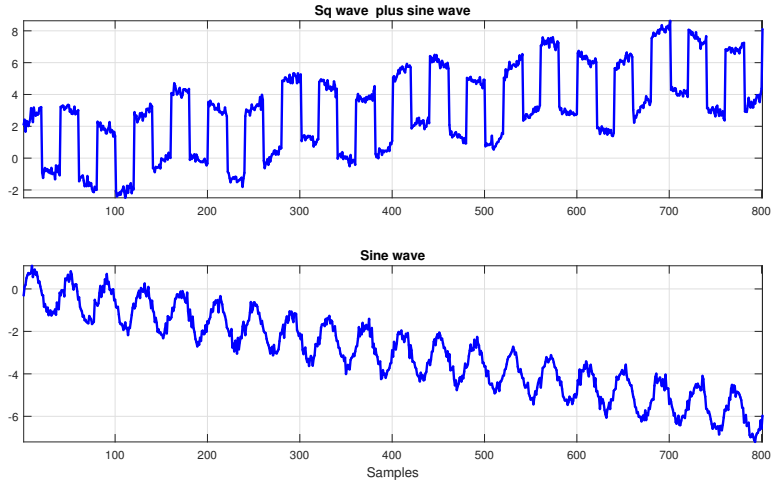

Figure 6.5: Simulation example:: square plus sine wave (first row), sine wave (second row).

Similarly there is only one IMF for the second signal. The time varying drifts, causing non-stationary effects are extracted in the residue.

The results for the extracted IMFs using the DVV analysis are presented in Table 6.4. For the square wave it can be seen that the IMFs containing the third and higher order harmonics in square wave have been designated as non-linear whereas the other IMFs do not show any signature of non-linearity. The only IMF for the sinusoidal signal is rightly detected as void of any non linearity. The proposed method therefore highlighted the nature of different oscillatory components in the signals as well as catering to the non stationary effects.

Table 6.4: DVV analysis (Simulation example)

Case	IMF	rank	type	RMSE
square	1	100	Non Linear	0.02
plus	2	97	Non Linear	0.004
	3	8	Linear	–
sinusoidal	4	36	Linear	–
sinusoidal	1	59	Linear	–

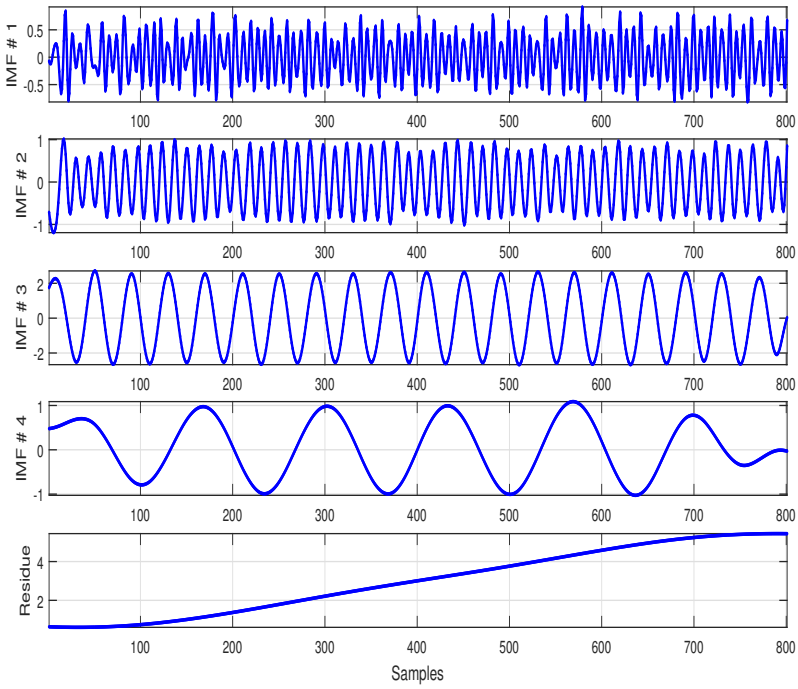

Figure 6.6: IMFs and residue for square plus sine wave (simulation example).

6.7 Industrial Case Study

Industrial data from a hydrogen reformer of the South East Asian refinery containing 37 Tags, recorded at sampling rate of 1 minute is used for the analysis. The same data has been analysed for the diagnosis of plant-wide oscillations in number of studies Aftab et al. (2017b), Aftab et al. (2017c), Choudhury (2006), Zang and Howell (2005), Thornhill (2005). The plant is reported to be suffering from the plant-wide oscillations with period 16.7min^{-1} due to valve non-linearity. It is also found that 12 tags, namely $2, 3, 4, 10, 11, 13, 19, 20, 24, 25, 33, 34$ (shown in Figure 6.8), are being affected by this plant-wide oscillation. The objective is to analyse the data from these tags using the proposed method and identify the tags that exhibit signatures of non-linearity and to isolate the origin of this plant-wide oscillation.

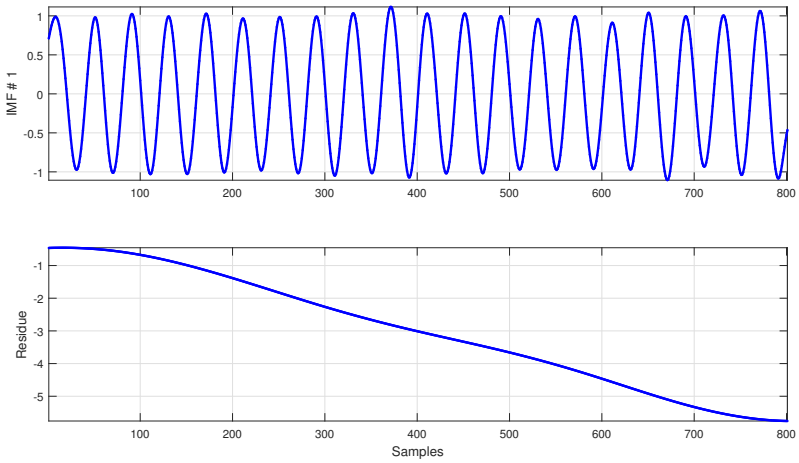

Figure 6.7: IMFs and residue for sine wave (simulation example).

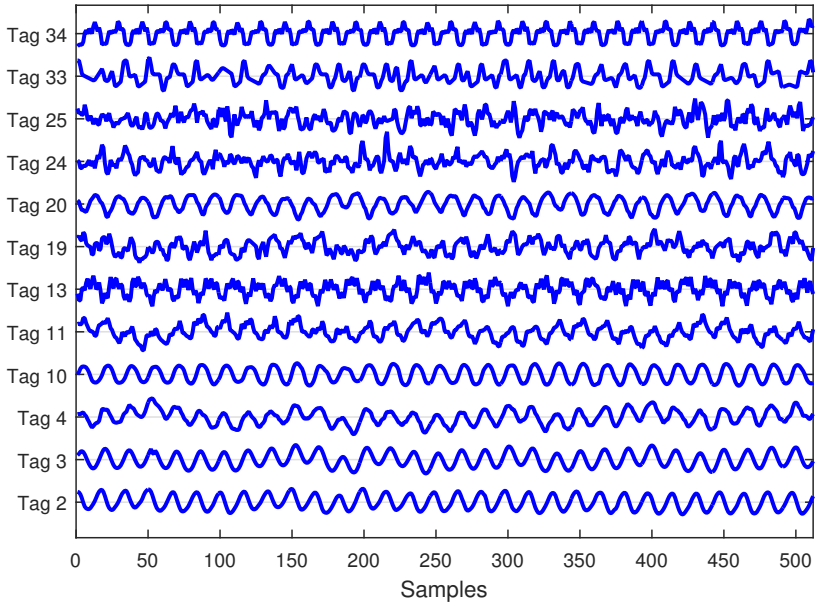

Figure 6.8: Time trends for the group of variables oscillating with frequency $0.06min^{-1}$ for SEA refinery (Aftab et al. 2017c).

Table 6.5: Industrial case study results

Tag	IMF	rank	Type	RMSE	Total RMSE
2	1	96	Non Linear	0.03	0.03
	2	61	Linear	–	
3	1	69	Linear	–	–
	2	67	Linear	–	
4	1	91	Non-Linear	0.007	0.007
	2	52	Linear	–	
	3	32	Linear	–	
10	1	49	Linear	-	-
11	1	99	Non-Linear	0.018	0.031
	2	97	Non-Linear	0.013	
	3	5	Linear	–	
	4	17	Linear	–	
13	1	79	Linear	–	–
	2	5	Linear	–	
19	1	95	Non-Linear	0.019	0.019
	2	16	Linear	–	
	3	55	Linear	–	
	4	83	Linear	–	
20	1	8	Linear	–	–
24	1	60	Linear	–	–
25	1	59	Linear	–	–
	2	60	Linear	–	
33	1	60	Linear	–	0.014
	2	98	Non-Linear	0.014	
	3	52	Linear	–	
34	1	100	Non-Linear	0.044	0.0511
	2	82	Linear	-	
	3	100	Non-Linear	0.007	

6.7.1 Detection of Non-Linearity

In all 12 tags are analysed using the proposed method and the results are tabulated in table 6.5. Six tags namely $2, 4, 11, 19, 33, 34$ are found to have signatures of non-linearity. The non-linearities are exhibited by the IMFs that represent fundamental oscillation at 16.7min^{-1} or its higher order harmonics. Tag 4 is a borderline case with very small RMSE value. Moreover, the proposed method has also highlighted the multiple oscillations in Tags $3, 4, 11, 19$ giving more insight into the oscillation pattern in different control loops.

The results are also compared with the existing non-linearity detection methods , the HOS method by Chaudhry et al. (2004) and N measure by Thornhill (2005) (Table 6.6). The results are quite similar to the N measure with the exception of

Table 6.6: Comparison of proposed and existing non-linearity measures

Tag	Proposed Method	HOS Chaudhry et al. (2004)			N (Thornhill 2005)
	Total RMSE	NGI	NLI	TNLI	(surrogate)
Tag 2	0.03	0.15	0.99	2.71	–
Tag 3	–	0.14	0.94	2.65	–
Tag 4	0.007	0.06	0.81	0.81	–
Tag 10	–	0.04	0.79	0.79	–
Tag 11	0.031	0.20	0.96	2.84	2.74
Tag 13	–	0.15	0.96	1.80	2.64
Tag 19	0.019	0.13	0.88	0.88	–
Tag 20	–	0.13	0.94	1.76	–
Tag 24	–	0.01	0.76	0.76	–
Tag 25	–	0.0	0.0	0.0	–
Tag 33	0.014	0.080	0.87	3.32	2.57
Tag 34	0.051	0.20	0.99	7.63	4.91

Tags $2, 4$ and 19. As already discussed that Tag 4 is a border line case. IMFs of Tag 2 and 19 have significant presence of second and third order harmonics of 16.7min^{-1} oscillation and therefore designated as non linear. HOS method has also declared Tag 2 highly non-linear with total degree of non-linearity (TNLI) around 2.71.

6.7.2 Isolating the Source of Non-linearity

The analysis of the case study reveals that the plant-wide oscillation with period 16.7min^{-1} originates due to non-linearity in one of control loops. As discussed earlier the extent of non-linearity will be greatest near the source and tends to get weaker away from the origin. Therefore, the tag with greatest signature of non-linearity, tag 34 in this case having the highest RMSE value can be regarded as the most probable source of the non-linearity. The same tag has been reported to be the most probable root cause of oscillation in number of other studies Aftab et al. (2017c;b), Thornhill (2005), Choudhury (2006).

6.8 Conclusions

In this paper the noise assisted MEMD and delay vector variance methods are integrated for the diagnosis of plant-wide oscillations. The proposed algorithm has enhanced the capability of original dvv version (Aftab et al. 2017b) to process the non stationary signals. The case of multiple sources of oscillation in a control loops can also be analysed via the proposed method. The NA-MEMD sifts out different oscillatory components from the control loop data and the nature of these oscillations are then ascertained using the DVV method that can also isolate the

source of non-linearity induced oscillations.

6.9 Acknowledgements

This work is being funded by Siemens AS, Norway. The authors would like to thank Prof A.K. Tangirala for the provision of industrial data.

Chapter 7

Identifying the Source of Plant-wide Oscillations via Causality Analysis

Paper E: Convergent Cross Mapping (CCM) based Approach for Isolating the Source of Plant-wide Disturbances

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

In Proceedings of 2017 IEEE Conference on Control Technology and Applications (CCTA), 2017, 1492-1498

Abstract

Disturbances originating in one control loop of a large industrial plant can propagate far from the source, giving rise to plant-wide oscillations. The underlying interactions among the different control loops make it hard to identify the origin of such large scale disturbances. This paper studies the application of the convergent cross mapping (CCM) based technique to isolate the source of a plant-wide disturbance. The proposed scheme exploits the cause and effect relationships among the affected variables to find the source of disturbance. The states of the causative factors are estimated from the effect variable and the directionality of information flow is established using the correlation between the original and estimated signal. The method is applied to the industrial case study and is shown to be effective in isolating the disturbance origin.

7.1 Introduction

In an industrial control system plant-wide disturbances can result in product variability and excessive use of precious resources, with both economically and environmentally detrimental consequences. Thus there is clear motivation to look for the root cause/source of these disturbances, so that targeted remedial action can be taken. Disturbances in a large scale plant may arise due to number of reasons such as, a) poor controller tuning, b) process degradation, and c) equipment wear and failure. Once started these disturbances travel away from their origin due to mass and energy flows as well as control loop interactions, thus giving rise to plant-wide disturbances (Choudhury et al. 2008, Thornhill 2005). In such a scenario, the early and accurate identification of the origin of the disturbance is a key to reduce maintenance and shut down time and thus improving the overall economics.

In order to find the source of plant-wide oscillations different approaches, both model based and data driven, have been considered by researchers over the past decade or so. The data driven methods are more popular owing to the fact that model based methods rely on process information/model, expert knowledge and analysis of P&IDs that may not be available or lack desired level of accuracy (Duan et al. 2014). Whereas data driven methods are based on the historical data alone and are quite flexible to use. The detailed description of both these approaches can be found in Duan et al. (2014).

The data driven approaches are mostly based on the analysis of cause and effect relationship, the so called causality analysis. The causality is defined as: *if prediction of time series y is improved by using the knowledge of another time series x, then x has a causal influence on y* (Wiener 1956). This definition of causality is taken up by Granger Granger (1969) who argued that x "Granger causes (G-Causes)" y if removing x from the universe of all possible causative variables U decreases the predictability of y i.e $\sigma^2(y|U) < \sigma^2(y|(U-x))$ (σ^2 is the variance). The key assumptions in the Granger Causality (GC) paradigm are, a) the cause occurs prior to effect, and b) the causative information of a cause variable is independently unique to that variable, the so called "separability" assumption i.e if x G-causes y then information about x is uniquely present in x only and can be removed from the system . The latter assumption may be difficult to ensure as in dynamic systems the information about x may be redundantly present in effect variable y and impossible to separate (Sugihara et al. 2012). The GC has been applied for diagnosis of plant-wide oscillations by Yuan and Qin (2014). Moreover, though GC is a powerful concept and has both time and frequency domain variants, it is originally applicable to purely linear stochastic systems only, although extensions to nonlinear systems have been proposed (Bressler and Seth 2011).

Bauer et al. (2007a) have studied the determination of disturbance propagation path in process systems by employing the "Transfer Entropy" (TE) methodology to measure the extent of information transfer from one variable to other. The method requires tuning of certain parameters like prediction horizon, time interval and embedding dimensions that can significantly effect the results (Duan et al. 2014). Moreover, estimation of joint and conditional probability density functions (pdfs) is computationally intensive and needs sufficiently large data sets.

Phase space reconstruction of the attractor manifold based on time delayed embedding is an important and popular concept for analysis of nonlinear dynamical systems. The time series $x(t)$ from a dynamical system is used for reconstruction of the attractor manifold $\mathbf{M}$, with embedding dimension m , using the time lagged vectors $\mathbf{x} = [x(t) \quad x(t-\tau) \quad x(t-2\tau) \ldots (t-(m-1)\tau)$; where τ is the time lag (Huffaker 2010), (Takens 1981), (Sauer et al. 1991) and (Kantz and Schreiber 2004). Interdependence between two time series $x(t)$ and $y(t)$ using the "nearest neighbors" (NN) from their respective attractor manifolds has been used in number of studies for example Pereda et al. (2001), Chicharro and Andrzejak (2009), and Arnhold et al. (1999). The similar approach has been adopted by Bauer et al. (2007b) for diagnosis of plant-wide disturbances. This method is also sensitive to tuning parameters like embedding dimension, prediction horizon and number of nearest neighbors.

In this paper we use a relatively new method called Convergent Cross Mapping or CCM (Sugihara et al. 2012), to locate the source of plant-wide disturbances. The method is also based on the nearest neighbor concept. The method exploits the fact that if $x(t)$ drives $y(t)$ then the effect variable $y(t)$ will contain signatures of $x(t)$, as the effect cannot influence the cause there will be no signatures of $y(t)$ in $x(t)$ for unidirectional causality. The procedure is rather simple and is claimed to work well with the short time series data as well (Sugihara et al. 2012). The extent of the causal influence is determined by estimating the states of time series x, denoted by $\hat{x}$, from near neighbors of y and vice versa. The correlation between estimated and original time series then gives the directionality of information flow. Moreover, the proposed method is integrated with automatic determination of embedding dimension to make it more robust and reliable. The proposed method is applied to an industrial data set suffering from plant-wide oscillations and is able to identify the source of disturbance.

The paper is organized as follows. Section 7.2 gives the detailed description of the proposed CCM based source detection method and Section 7.3 outlines the statistical test to check the significance of the correlation between estimated and original time series. Automatic determination of embedding dimensions is presented in Section 7.4. Section 7.5 describes the steps involved in the proposed method for

identifying the root source followed by an industrial case study and conclusions in Sections 7.6 and 7.7 respectively.

7.2 Convergent Cross Mapping (CCM)

Convergent Cross Mapping (CCM) is developed by Sugihara et al. (2012) for the causality analysis of ecological time series. The method make use of the fact that time series from the same dynamical system share a common attractor manifold and hence can be used to estimate the state of each other. In CCM causal relationship $x \rightarrow y$ is established by looking at the signatures of causative factor x in the effect variable y and in case of $x \rightarrow y$, y can give reliable estimates of x.

In order to explain the CCM, consider two discrete time series $x(k)$ and $y(k)$ each with N samples, from a dynamical system that share a common attractor manifold $\mathbf{M}$. Making use of the time delayed embedding (Takens (1981)) shadow attractor manifolds $\mathbf{M_x}$ and $\mathbf{M_y}$, with embedding dimension m, can be constructed from the time series $x(k)$ and $y(k)$ respectively. The time delayed embedding for manifold $\mathbf{M_x}$ is represented as $\mathbf{x_i} = [x(i) \quad x(i-\tau) \quad x(i-2\tau) \ldots x(i-(m-1)\tau)]$ $\forall i = (m-1)\tau + 1, (m-1)\tau + 2, \ldots N$; construction of $\mathbf{M_y}$ follows the same procedure.

CCM looks for how well the local neighbourhood of $\mathbf{M_x}$ maps to the local neighbourhood of $\mathbf{M_y}$. In order to determine the causative influence of $x(k)$ on $y(k)$, for each vector i in $\mathbf{M_y}$, $m+1$ nearest neighbors are identified and their corresponding entries are marked in $x(k)$. The weighted mean (weights calculated using equations 7.5 and 7.7) of these these $m+1$ values in $x(k)$ are used to estimate the state of $x_i(k)$ ($\hat{x}_i(k)|\mathbf{M_y}$); given by

$$\hat{x}_i = \sum_{j=1}^{m+1} w_j x_j \tag{7.1}$$

where $\sum w_j = 1$. The estimate of $\hat{y}|M_x$ can be calculated in an analogous way to infer the causative influence of y on x. The root mean square error($e_{\hat{x}|M_y}$) and correlation ($\rho_{\hat{x}|M_y}$), given in (7.2), can then be used to determine the extent of causative influence of $x \rightarrow y$.

$$\begin{aligned} \rho_{\hat{x}|M_y} &= \frac{\sum \hat{x}_i x_i - N^* \bar{x}_i \bar{\hat{x}}_i}{\sqrt{(\sum x_i^2 - N^* \bar{x}^2)}\sqrt{(\sum \hat{x}_i^2 - N^* \bar{\hat{x}}^2)}} \\ e_{\hat{x}|M_y} &= \sqrt{\frac{1}{N^*}(\sum_{k=1}^{k=N^*} \hat{x}_k|M_y - x_k)^2} \end{aligned} \tag{7.2}$$

where N^* is the number of embedded vectors. The greater the correlation $\rho_{\hat{x}|M_y}$ (lower rms error), the greater is the causative influence of x on y and vice versa. The extent of causative influence of y on x can be determined similarly. In the case of unidirectional causality from x to y, $x(k)$ contains no information about $y(k)$ and thus cannot reliably estimate the states of $y(k)$ giving lower values of correlation and higher rms error. The correlation coefficient, as calculated in (7.2), is used in this work to determine the direction of information flow. The correlation coefficient is the preferred choice as it can be transformed to normally distributed z-scores using Fisher's z-transformation (Fisher 1915) and can thus be checked for significance against some null hypothesis. The significance test will be explained in section 7.3.

7.2.1 Convergence in CCM

Another important aspect of CCM is the convergence of ρ and rms error (e in equation 7.2) with increase in length of time series N. By convergence we mean that ρ increases (e_{rms} decreases) as length increases. This is so because with increased information the trajectories forming the attractor fill in, thus giving closer neighbors and higher correlation (lower rmse error) (Sugihara et al. 2012). The convergence will be limited by observational error, noise and time series length [1]. Nevertheless the convergence of the estimates is a key to establish the directionality of information flow. In case of bidirectional coupling, apart from the absolute value of the correlation ρ, its convergence rate can also be used to judge the extent of causation. The greater the causality effect the greater will be the convergence rate and vice versa. This can be explained from the following illustrative example of CCM.

7.2.2 Illustrative Example

The interpretation of results from CCM analysis can be explained using the following example from Sugihara et al. (2012). Consider two interacting time series given by

$$\begin{aligned} x(k+1) &= x(k)[3.8 - 3.5x(k) - \beta_{xy}y(k)] \\ y(k+1) &= y(k)[3.8 - 3.8y(k) - \beta_{yx}x(k)] \end{aligned} \tag{7.3}$$

where β_{xy} (β_{yx}) controls the contribution of $y(k)$ ($x(k)$) on $x(k)$ ($y(k)$).

Time series and results of CCM analysis for the case $\beta_{yx} > \beta_{xy}$ are shown in Figure 7.1 and Figure 7.2 respectively. It has been shown by Sugihara et al. (2012) that the causal relationship in this example cannot be captured by Granger causality as the "separability" assumption cannot be fulfilled. The estimated state of x has

[1] In perfect deterministic and noise free settings $\rho \to 1$ as length $L \to \infty$ Sugihara et al. (2012)

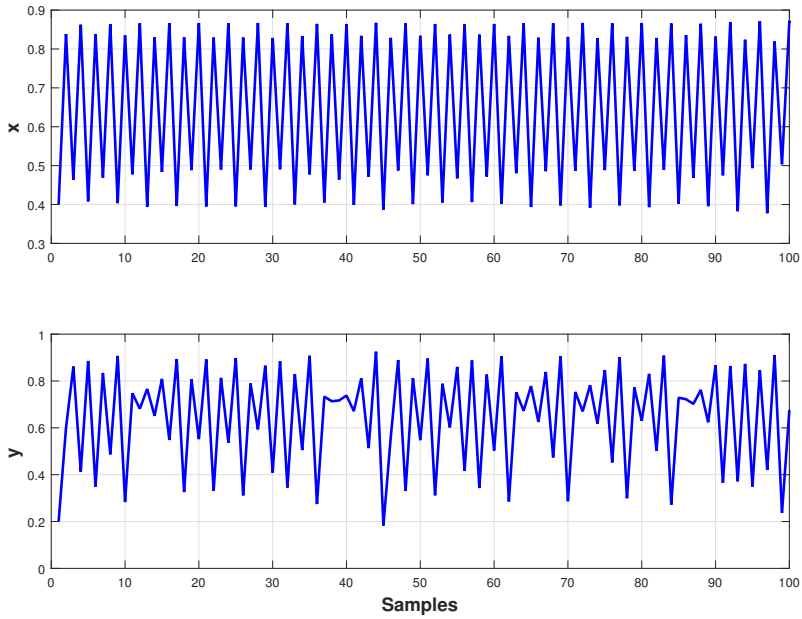

Figure 7.1: Time series x and y for CCM illustrative example, $\beta_{yx} = 0.1, \beta_{xy} = 0.02$.

greater correlation with the original time series and $\hat{x}|M_y$ converges faster than $\hat{y}|M_x$ thus confirming that x has larger influence on y.

7.2.3 Convergent Cross Mapping Algorithm

The causation detection from $x \rightarrow y$ using the CCM method is summarized in the following steps (Sugihara et al. 2012)

1. Consider two time series $x(k)$ and $y(k)$ of length N, mean centred and normalized to unit standard deviation.
2. Construct time delayed embedding vectors $Y_i(t)$, the shadow manifold $\mathbf{M_y}$, for time series $y(t)$ with given embedding dimension m and time delay τ using

$$Y_i = [y(i), y(i-\tau), y(i-2\tau), \ldots y(i-(m-1)\tau)] \tag{7.4}$$

   where $i = (m-1)\tau + 1, (m-1)\tau + 2, \ldots N$.
3. For each vector Y_i find the $m+1$ nearest neighbors on the manifold M_y denoted by $n_j \; \forall j = 1, \ldots (m+1)$.

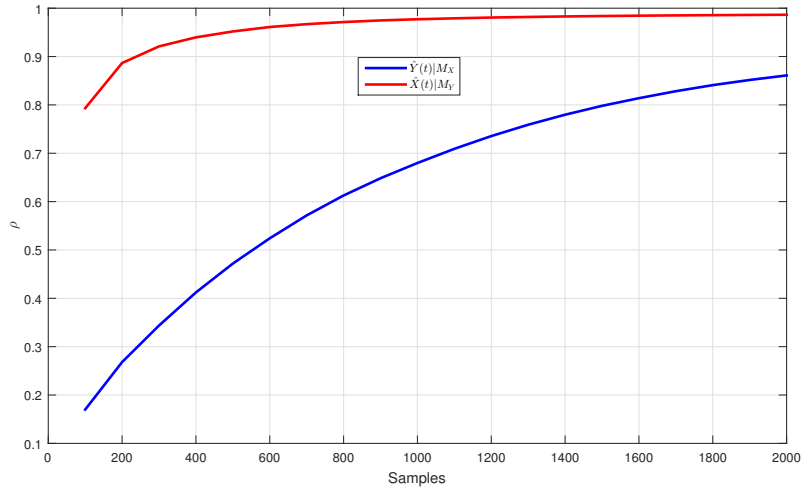

Figure 7.2: CCM results (illustrative example).

4. Assign weights γ_j based on the distance of each $m+1$ nearest neighbors from Y_i with the weight of the nearest neighbor getting maximum value; given by

$$\gamma_j = exp(\frac{-||Y_i - Y_j||}{||Y_i - Y_{n_1}||}) \tag{7.5}$$

where $||Y_i - Y_j||$ and $||Y_i - Y_{n_1}||$ are the Euclidean distance in $\mathbf{M_y}$ between i^{th} and j^{th} vector and nearest neighbor respectively. In case distance to the nearest neighbor is zero then $\gamma_1 = 1$ and $\gamma_j = 0\ \forall j = 2 : m+1$.

5. Using the nearest neighbors $n_j\, \forall j = 1, \ldots (m+1)$ from $\mathbf{M_y}$ manifold locate the corresponding values in $x(t)$ and mark them $x_j, x_{j+1}, \ldots x_{m+1}$

6. Generate cross mapped estimate of $x(t)$ given by $\hat{x}|M_y$ from weighted mean of x_{js} using

$$\hat{x}_i = \sum_{j=1}^{m+1} w_j x_j \tag{7.6}$$

where

$$w_j = \frac{\gamma_j}{\sum \gamma_k} \quad k = 1, \ldots m+1 \tag{7.7}$$

7. Calculate the correlation coefficient $\rho_{\hat{x}|M_y}$ using Equation 7.2.

8. Repeat Steps 1-6 for different time series length N to check for convergence of $\rho_{\hat{x}|M_y}$ as a function of time series length.

The cross mapping from x to y can be evaluated similarly. the directionality or the strength of causality can be determined from the difference in $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$. If $\rho_{\hat{x}|M_y} > \rho_{\hat{y}|M_x} \implies x \to y$ and vice versa.

7.3 Significance Test

It is discussed in the preceding section that the directionality or the extent of causative effects can be determined from the relative value of correlation (ρ) between the original and estimated time series. However, one should question whether the difference between $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$ is significant to reliably infer the causal relationship and direction of information flow. The same is true for inferring the convergence of the $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$. To this end, it is necessary to test the null hypothesis

H_0 :: *There is no significant difference in correlations* i.e. $\rho_{\hat{x}|M_y} = \rho_{\hat{y}|M_x}$

The sample correlation does not follow the normal distribution so in order to test the null hypothesis, i.e, the significance between the relative correlation, Fisher's $z-$ transformation is used. Fisher's $z-$ transformation maps the correlation coefficient, to normally distributed z-scores by the relation.

$$z_{\hat{x}|\mathbf{M_y}} = 0.5 \ln \left[\frac{1 + \rho_{\hat{x}|M_y}}{1 - \rho_{\hat{x}|M_y}} \right] \tag{7.8}$$

where $\ln$ is the natural logarithm. Similarly $z_{\hat{y}|\mathbf{M_x}}$ can be calculated from $\rho_{\hat{y}|M_x}$ The difference between $z-$ transformed correlation coefficients can be given by

$$Z^* = \frac{z_{\hat{y}|\mathbf{M_x}} - z_{\hat{x}|\mathbf{M_y}}}{\sqrt{\frac{1}{N_1 - 3}} + \sqrt{\frac{1}{N_2 - 3}}} \tag{7.9}$$

where N_1 and N_2 are sample sizes for two correlations. In this case we have $N_1 = N_2 = N^*$. The NULL hypothesis is rejected at the confidence level α, for a two tailed-test, if $|Z^*| > Z_{\alpha/2}$. A confidence level of $\alpha = 0.01$ is used in this work, which gives $Z_{\alpha/2} = 2.58$.

7.3.1 Testing the Significance of Convergence

It has been discussed in Section 7.2.1 that the correlation estimates tends to increase with increase in time series length. Therefore, the significance of convergence can be used to distinguish the case when two variables are driven by a common cause but have no influence on each other. In such a scenario increasing the

time series length will not improve the correlation estimates (Sugihara et al. 2012) and thus can be excluded from the cause-effect analysis.

Thus, to test convergence of the correlation, between original and estimated time series, the CCM is performed for increasing lengths of time series in interval $[N_{min}N_{max}]$ with N_{max} can be at most equal to N, the length of input time series.

The convergence is inferred if difference between $\rho_{N_{min}}$ and $\rho_{N_{max}}$, using Equations (7.8) and (7.9), is found to be significant. The only difference would be to plug in $\rho_{N_{min}}$ and $\rho_{N_{max}}$ to calculate $z-$scores in Equation (7.8) and finding whether the corresponding Z^* gives $|Z^*| > Z_{\alpha/2}$ for confidence level α.

7.4 Determining Embedding Dimension

The approach used in this work is based on the phase space reconstruction of attractor manifolds $\mathbf{M_x}$ and $\mathbf{M_x}$ via delayed embedding. An important parameter in this approach is the embedding dimension of the attractor. The embedding dimension is to be chosen such that the attractor unfolds enough to describe the dynamic behavior of the underlying dynamical system. Moreover, if the embedding dimension m is less than the required minimum m_0 i.e $m < m_0$, then the points appearing as near neighbor of a point in state space may not be true neighbors in actual state space (Kennel et al. 1992). They only appear to be near neighbors in low dimensional phase space because the attractor may not have unfolded fully.

Therefore any analysis based on the nearest neighbor approach is prone to error. This is true for the work presented here and other related methods such as given in Bauer et al. (2007b). The adverse impact of choosing a larger embedding dimension than required is the increased computational effort without improving the result much. Thus, it is quite essential to have an automatic way of determining the minimum embedding dimension to make any method based on phase space reconstruction reliable with minimum computational effort. In order to address this concern an automatic method for determining the minimum embedding dimensions based on false near neighbors is incorporated with the CCM based method to make it robust and effective.

The minimum embedding dimensions needed to for phase space reconstruction are calculated using the method of false nearest neighbors proposed by Cao (1997). The method relies on the fact that if m is the true embedding dimension of the phase space, then points near in $m-$ dimensional space will remain close in the $m+1$ dimensional space. Consider and time series $z(t) = z_1, z_2, \ldots z_N$ of N samples. Then delayed vectors, with embedding dimension m and time delay τ

are given by

$$Z_i = (z_i, z_{i+\tau}, \ldots z_{i+(m-1)\tau}) \quad \forall i = 1, 2, \ldots (N - m)\tau \tag{7.10}$$

For i^{th} vector Z_i distance measure $a(i, m)$ is defined as

$$a(i,m) = \frac{\|Z_i(m+1) - Z_{n(i,m)}(m+1)\|}{\|Z_i(m) - Z_{n(i,m)}(m)\|} \qquad i = 1, 2, \ldots, N - m\tau \tag{7.11}$$

where $\|.\|$ represents vector norm operation; $Z_i(m+1)$ and $Z_i(m)$ are i^{th} delay vectors in m and $m+1$ dimensional space. $Z_n(m)$ is the nearest neighbor of $Z_i(m)$ in the m dimensional space, where $n(i, m)$ is integer in set $(1 \geq n \geq N - m\tau)$.

The value of $a(i, m)$ is calculated for all the delay vectors and its mean $E(m)$ is given by

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a(i,m) \tag{7.12}$$

The variation in E for the increase in embedding dimension to $m + 1$ is given by the quantity $E1$ defined as

$$E_1(m) = \frac{E(m+1)}{E(m)} \tag{7.13}$$

The quantity $E_1(m)$ in (7.13) approaches unity and stops changing for some $m > m_0$ when the near neighbors in m dimensions stay near in $m+1$ dimension as well. If this is the case then the minimum embedding dimension is given by $m_0 + 1$.

In order to illustrate the procedure described above for determination, the $E1(m)$ quantity for one of the data sets, discussed in Section 7.6, is plotted in Figure 7.3. It can be seen that E_1 approaches unity and stops changing for $m > 8$, thus giving $m = 9$ as the minimum embedding dimension for phase space reconstruction.

7.5 Proposed Method

The method proposed looks for the causal relationship and direction of information flow between two time series $x(k)$ and $y(k)$ each of length N. Both time series are recorded from a dynamical system and share a common manifold $\mathbf{M}$ in phase space. The procedure involves cross mapping /estimation of one time series from the other thereby giving the direction of maximum influence or information flow. It is assumed that the variables suffering from plant-wide disturbance/oscillations have been identified and grouped before hand. The main steps involved in the proposed method are as follows:

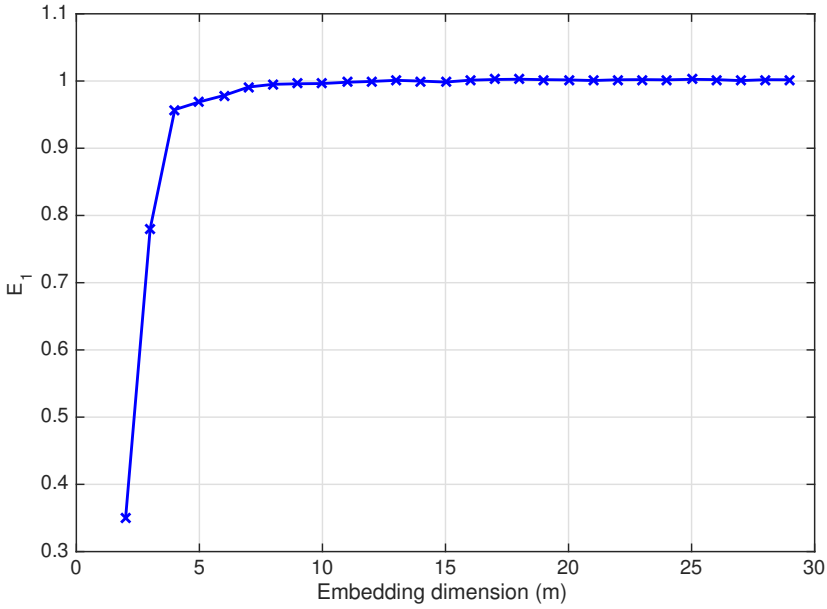

Figure 7.3: Automatic determination of embedding dimension (m).

Step I Mean center and normalize the time series, for variables suffering from the plant-wide disturbance, to unit standard deviation.

Step II Determine the minimum embedding dimension for the time series using the method given in section 7.4. All the time series can have different embedding dimension.

Step III Consider two variables at a time and label them x and y .

Step IV Compute $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$ using CCM method described in section 7.2.

Step V Test the NULL hypothesis using significance test given in section 7.3.

Step VI Repeat steps III-V for increasing length of time series starting from N_{min} to N_{max}. Variables passing the convergence test, given in Section 7.3.1, are considered for analysis.

Step VII In case the the NULL hypothesis is rejected and there is significant difference in $\rho_{N_{Max}}$ and $\rho_{N_{Min}}$, then $\rho_{\hat{x}|M_y} > \rho_{\hat{y}|M_x}$ gives $x \rightarrow y$ otherwise $y \rightarrow x$.

Step VIII Repeat steps III-VII for all pairs of variables affected by the plant-wide disturbance.

Step IX A variable influencing all other affected variables is designated as the source of disturbance.

Step X Alternatively, an information flow graph can be constructed with variables as nodes and arrows directing the flow of information as determined in step VII. This is the same concept as is employed in Duan et al. (2014).

7.6 Industrial Case Study

The proposed CCM based method is applied to industrial data from a South East Asian refinery. The same data set has been used previously for the detection of plant-wide oscillations(Thornhill et al. (2002),Tangirala et al. (2005)) and root cause analysis (Thornhill (2005), Choudhury (2006) and Zang and Howell (2005)). The data consists of 37 different tags each having $N = 512$ samples recorded at sampling rate of 1 min. The plant schematic is shown in Figure 7.4.

It has been reported that the plant suffers from a plant-wide oscillation, due to valve non-linearity in pressure swing absorption (PSA) unit, with oscillation period of $16.7\ \text{min}^{-1}$. Nine tags are identified to be affected by this oscillation (Thornhill et al. (2002)) (shown in Figure 7.5). The non-linearity signatures are found to be present in four tags, namely $11, 13, 33$ and 34 (Thornhill 2005). Tag 34 is found to be the most nonlinear and is designated as the source of this plant-wide oscillation. The non-linearity indices calculated by Thornhill Thornhill (2005) are also shown in the Figure 7.5.

The same data set is studied using the proposed CCM method to analyze how well it can identify the origin of the plant-wide disturbance. Two scenarios are considered here, a) in the first scenario the known reason for disturbance is taken into account and tags showing non-linearity are analyzed with the proposed method to identify the origin of the non-linearity, and b) the second scenario assumes that the cause is not known in advance and all 9 tags affected by plant-wide oscillation are analyzed. This scenario will test the broader scope of the proposed method.

7.6.1 Parameter Settings

The following parameters settings are used to analyze the industrial data.

- The analysis is started with sample size of $N_{min} = 100$ samples that is increased in steps of 100. The maximum number of samples used are $N_{max} =$

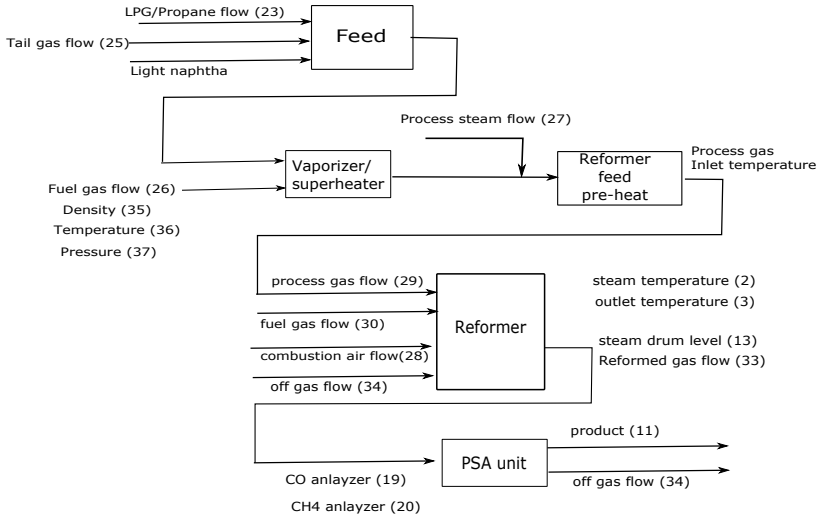

Figure 7.4: Process schematic for industrial case study (Thornhill 2005).

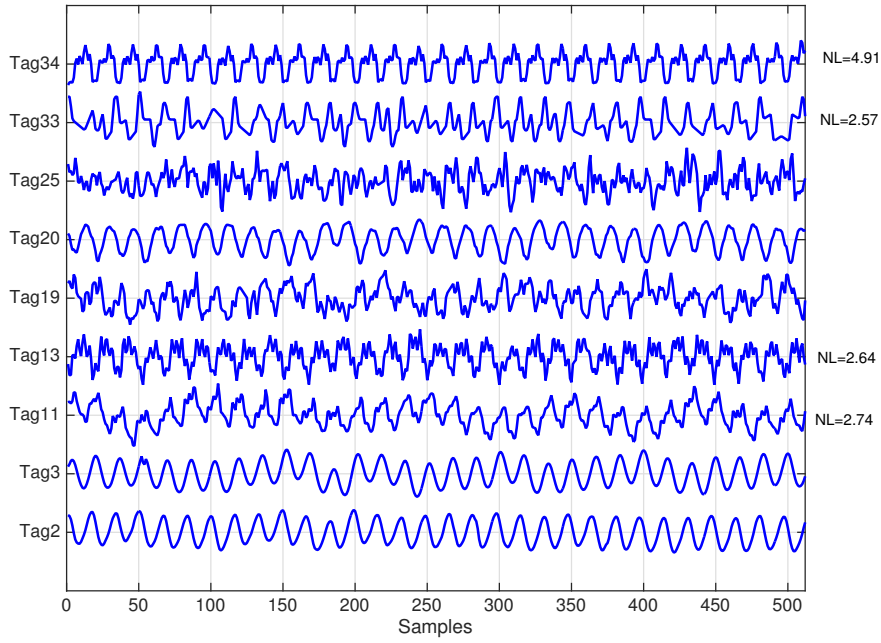

Figure 7.5: Time trends and non-linearity indices for SEA refinery data (case study).

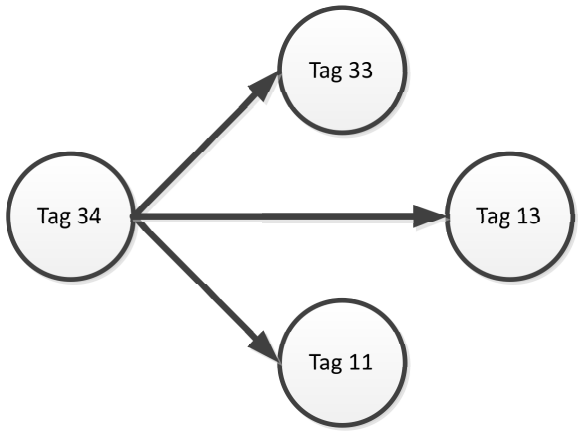

Figure 7.6: CCM results (information flow graph): industrial case study (known root cause).

500

- Minimum embedding dimensions are computed using the method given in section 7.4 and it is found that $m = 20$ fits well for all the time series. One may choose different m for different tags, but in this example same m is used for all the loops.

7.6.2 Analysis with Known Root Cause

As discussed the plant-wide oscillations in the industrial data under study are due to valve non-linearity. Only four tags are identified as the ones with signatures of nonlinearity. These four tags $11, 13, 33$ and 34 are analyzed using the proposed CCM based method and results are summarized in the Table 7.1. The corresponding correlation plots are given in Figure 7.7.

It can be seen from the Table 7.1 that tag 34 has influence on all the other loops.

Table 7.1: CCM results (known root cause):: (✓) shows the presence of causative effect

Cause	Influenced Tags			
	Tag 11	Tag 13	Tag 33	Tag 34
Tag 11		–	–	–
Tag 13	–		–	–
Tag 33	–	–		–
Tag 34	✓	✓	✓	

Table 7.2: CCM results for all loops suffering from plant-wide oscillation (industrial case study) :: (✓) shows the presence of causative effect

Cause	Influenced Tags								
	Tag 2	Tag 3	Tag 11	Tag 13	Tag 19	Tag 20	Tag 25	Tag 33	Tag 34
Tag 2		–	–	–	–	–	✓	–	–
Tag 3	–		–	–	–	–	✓	–	–
Tag 11	–	–		–	–	–	✓	–	–
Tag 13	–	–	–		–	–	–	–	–
Tag 19	–	–	–	–		–	–	–	–
Tag 20	–	✓	–	–	–		–	–	–
Tag 25	–	–	–	–	–	–		–	–
Tag 33	–	–	–	–	–	–	–		–
Tag 34	✓	✓	✓	✓	✓	✓	✓	✓	

The same effect is shown in the form of information flow graph in Figure 7.6. The disturbance is due to valve non-linearity in the PSA section and both tags 34 and 11 are output of this section. Moreover, the tag 34 is recycled to the reformer section thereby effecting the flows emanating from reformer down to the PSA. The information flow graph shows only those connections that fulfil the significance criteria given in section 7.3 and show $\Delta\rho > 0.02$. Thus the proposed method has been successful in designating the origin of plant-wide disturbance.

7.6.3 Analysis with Unknown Root Cause

In this scenario it is assumed that the root cause is not known and all the loops affected by plant-wide oscillation are analyzed using proposed method. The results are summarized in Table 7.2. Here again tag 34 is found to be influencing all the other variables and thus designated as the source of plant-wide disturbance. It is interesting to note that tag 25 has been influenced by tags other than tag 34. It might be so because the tag 25 is recycled waste gas from another unit and the disturbance propagated through that unit (Thornhill 2005).

7.7 Conclusions

In this paper convergent cross mapping based approach for isolating the source of plant-wide disturbance is presented. The method is simple and effective and requires tuning of embedding dimension only. The proposed method is made reliable by appending it with automatic determination of embedding dimension. Moreover, the application of the proposed approach to an industrial case study gives promising results in identifying the origin of disturbance.

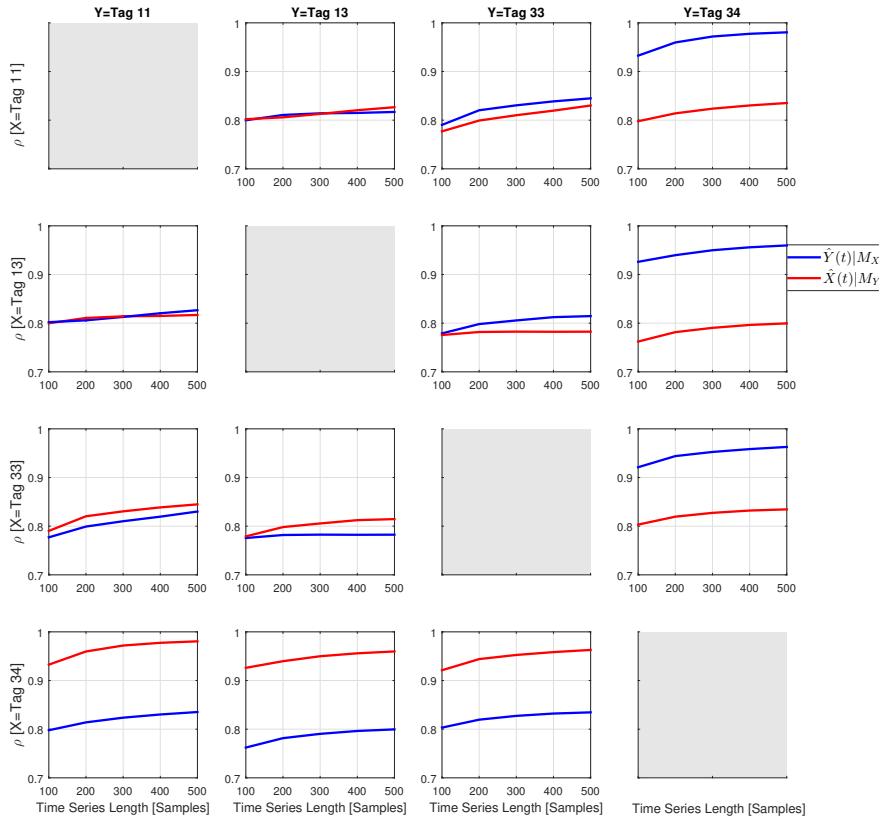

Figure 7.7: CCM results industrial case study (with known root cause).

7.8 Acknowledgments

The financial support from Siemens AS, Trondheim is gratefully acknowledged. The authors would like to thank Prof A.K. Tangirala for providing the industrial data.

Chapter 8

Conclusions and Future Work

The thesis is aimed at contributing to controller performance monitoring and diagnosis by introducing adaptive data processing tools. Primarily multivariate empirical mode decomposition (MEMD) and associated properties employed for detection and characterisation of control loops oscillations have been presented in this thesis. The core work can be divided into two parts namely. *a*) oscillation detection, and *b*) oscillation diagnosis .

The detection aspect is covered by introducing improved methods for automatic detection of oscillations both in individual loops and in a plant-wide sense. The effectiveness and robustness of the proposed methods have been established via simulation as well as industrial case studies. The proposed methods are fully data driven and can handle non-stationary effects and do not require any kind of pre filtering.

The diagnosis part consists of detection of non-linearity induced oscillations and isolating the source of disturbance via causality analysis. An automated way of identifying the harmonics in case of non-linearity induced oscillations has been proposed, using the dyadic filter bank property of MEMD. Harmonics analysis can differentiate between the oscillations caused by linear and non-linear sources without any pre-filtering while catering for time varying disturbances as well. A degree of non-linearity (DNL) index is used to quantify the level of non-linearity and the variable with the highest DNL is designated as the most probable source of non-linearity induced oscillations. Moreover, delayed vector variance (DVV) has been integrated with MEMD to make it applicable for the non-stationary data as well as identifying the sources of multiple oscillations in the control loops.

Cause and effect relationship among different plant variables is also exploited via

convergent cross mapping (CCM) to identify the source of plant-wide disturbance. This simple and effective method uses the phase space reconstruction and near neighbours to establish the causal relationship among different variables. The method can work for both non-linear and short term time series data. Moreover, the automatic determination of embedding dimension (the most important parameter) has rendered it more robust and reliable.

Though the work presented in this thesis has been backed by the industrial case studies and results are compared with the existing methods, further verification using industrial data should be undertaken. The aim of such further verification should be both to document the applicability and effectiveness of the methods proposed, and to compare the methods proposed herein to methods proposed previously. Different aspects like effects of data compression and data loss also need to be studied in future work. An important step would be to compare results of transfer entropy and Granger causality with the CCM based approach using more simulation and industrial case studies, so that pros and cons of the different techniques can be highlighted.

The mode mixing problem associated with standard empirical mode decomposition has been considerably reduced via noise assisted MEMD, but is not removed completely. The future work can also be directed in finding the ways and methods to further reduce or remove the mode mixing effects, thereby improving the robustness and effectiveness of the MEMD.

The work in the thesis is aimed at catering for more general control performance issues like oscillation detection and non-linearity detection. Future work may also entail the investigation of specific faults occurring in different system components.

Bibliography

Aftab, M. F., Hovd, M., Huang, N. E., Sivalingam, S., 2016. An adaptive non-linearity detection algorithm for process control loops. IFAC-PapersOnLine 49 (7), 1020 – 1025.

Aftab, M. F., Hovd, M., Sivalingam, S., Aug 2017a. Convergent cross mapping (ccm) based approach for isolating the source of plant-wide disturbances. In: 2017 IEEE Conference on Control Technology and Applications (CCTA). pp. 1492–1498.

Aftab, M. F., Hovd, M., Sivalingam, S., 2017b. A delay vector variance based approach for detecting and isolating the non-linearity induced oscillations in control loops. IFAC-PapersOnLine 50 (1), 7975 – 7980, 20th IFAC World Congress.

Aftab, M. F., Hovd, M., Sivalingam, S., 2017c. Detecting non-linearity induced oscillations via the dyadic filter bank property of multivariate empirical mode decomposition. Journal of Process Control 60, 68–81.

Aftab, M. F., Hovd, M., Sivalingam, S., 2017d. Diagnosis of plant-wide oscillations by combining multivariate empirical mode decomposition and delay vector variance. Journal of Process Control(Submitted).

Aftab, M. F., Hovd, M., Sivalingam, S., 2017e. Plant-wide oscillation detection using multivariate empirical mode decomposition. Computers and Chemical Engineering(Submitted).

Aftab, M. F., Hovd, M., Sivalingam, S., 2018. Improved oscillation detection via noise assisted data analysis. Control Engineering Practice(Submitted).

Andersson, L. E., Aftab, M. F., Scibilia, F., Imsland, L., Aug 2017. Forecasting using multivariate empirical mode decomposition; applied to iceberg drift forecast. In: 2017 IEEE Conference on Control Technology and Applications (CCTA). pp. 1097–1103.

Arnhold, J., Grassberger, P., Lehnertz, K., Elger, C. E., 1999. A robust method for detecting interdependences: application to intracranially recorded eeg. Physica D: Nonlinear Phenomena 134 (4), 419–430.

Babji, S., Gorai, P., Tangirala, A. K., 2009. Detection and quantification of control valve nonlinearities using Hilbert–Huang transform. Advances in Adaptive Data Analysis 1 (03), 425–446.

Bacci di Capaci, R., Scali, C., 02 2018. Review and comparison of techniques of analysis of valve stiction: from modeling to smart diagnosis. Chemical Engineering Research and Design 130.

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

Bauer, M., Cox, J. W., Caveness, M. H., Downs, J. J., Thornhill, N. F., 2007b. Nearest neighbors methods for root cause analysis of plantwide disturbances. Industrial & Engineering Chemistry Research 46 (18), 5977–5984.

Bauer, M., Horch, A., Xie, L., Jelali, M., Thornhill, N., 2016. The current state of control loop performance monitoring–a survey of application in industry. Journal of Process Control 38, 1–10.

Bressler, S. L., Seth, A. K., 2011. Wiener–Granger causality: a well established methodology. Neuroimage 58 (2), 323–329.

Cao, L., 1997. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena 110 (1), 43–50.

Chaudhry, M., Shah, S. L., Thornhill, N. F., 2004. Diagnosis of poor control-loop performance using higher-order statistics. Automatica 40, 1719–1728.

Chicharro, D., Andrzejak, R. G., 2009. Reliable detection of directional couplings using rank statistics. Physical Review E 80 (2), 026217.

Choudhury, A. A. S., Shah, S. L., Thornhill, N. F., 2008. Diagnosis of process nonlinearities and valve stiction: data driven approaches. Springer Science & Business Media.

Choudhury, M. S., 2006. Troubleshooting plantwide oscillations using nonlinearity information. Journal of Chemical Engineering 24, 50–60.

Cui, J., Freeden, W., 1997. Equidistribution on the sphere. SIAM Journal on Scientific Computing 18 (2), 595–609.

Desborough, L., Miller, Y., 2002. Increasing customer value of industrial control performance monitoring Honeywell's experience. In: 6th International Conference Chemical Process Control, AIChE Symp., Series 326. American Institute of Chemical Engineers.

di Capaci, R. B., Scali, C., 2015. Review on valve stiction. part i: From modeling to smart diagnosis. Processes 3 (2), 422.
URL http://www.mdpi.com/2227-9717/3/2/422

Duan, P., Chen, T., Shah, S. L., Yang, F., 2014. Methods for root cause diagnosis of plant-wide oscillations. AIChE Journal 60 (6), 2019–2034.

Eberly, D., 1999. Quaternion algebra and calculus. Tech. rep., Geometric Tools, LLC http://www.geometrictools.com/.

El-Ferik, S., Shareef, M. N., Ettaleb, L., 2012. Detection and diagnosis of plant-wide oscillations using GA based factorization. Journal of Process Control 22 (1), 321 – 329.

Fisher, R. A., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10 (4), 507–521.

Flandrin, P., Gonçalves, P., Rilling, G., 2014. EMD equivalent filter banks, from interpretation to applications. Hilbert-Huang Transform and Its Applications 16, 99–116.

Flandrin, P., Rilling, G., Goncalves, P., 2004. Empirical mode decomposition as a filter bank. Signal Processing Letters, IEEE 11 (2), 112–114.

Gautama, T., Mandic, D. P., Van Hulle, M. M., 2004a. The delay vector variance method for detecting determinism and nonlinearity in time series. Physica D: Nonlinear Phenomena 190 (3), 167–176.

Gautama, T., Mandic, D. P., Van Hulle, M. M., 2004b. A novel method for determining the nature of time series. IEEE Transactions on Biomedical Engineering 51 (5), 728–736.

Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424–438.

Hägglund, T., 1995. A control-loop performance monitor. Control Engineering Practice 3(11), 1543–1551.

Hägglund, T., 2011. A shape analysis approach for diagnosis of stiction in control valves. Control Engineering Practice 40, 782–789.

Horch, A., Oct. 1999. A simple method for detection of stiction in control valves. Control Engineering Practice 7(10), 6708–6718.

Hoyer, P. O., 2004. Non-negative matrix factorization with sparseness constraints. The Journal of Machine Learning Research 5, 1457–1469.

Huang, N. E., Lo, M.-T., Zhao-Hua, W., Xian-Yao, C., May 20 2014. Method for quantifying and modeling degree of nonlinearity, combined nonlinearity, and nonstationarity. US Patent 8,732,113.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 454 (1971), 903–995.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., Blank, K., 2009. On instantaneous frequency. Advances in Adaptive Data Analysis 1 (02), 177–229.

Huffaker, R., 2010. Phase space reconstruction from time series data: where history meets theory. Proceedings in Food System Dynamics, 1–9.

Jelali, M., 2008. Estimation of valve stiction in control loops using separable least-squares and global search algorithms. Journal of Process Control 18 (7-8), 632–642.

Jiang, H., Choudhury, M. S., Shah, S. L., 2007. Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method. Journal of Process Control 17 (2), 143–155.

Kantz, H., Schreiber, T., 2004. Nonlinear time series analysis. Vol. 7. Cambridge University Press.

Karra, S., Jelali, M., Karim, M. N., Horch, A., 2010. Detection of Oscillating Control Loops. Springer London, London, Ch. Chapter 4, pp. 61–100.
URL https://doi.org/10.1007/978-1-84882-775-2_4

Karra, S., Karim, M. N., 2009. Comprehensive methodology for detection and diagnosis of oscillatory control loops. Control Engineering Practice 17 (8), 939–956.

Karra, S., Karim, M. N., 2010. DOscillation Root-cause Detection and Quantification Under Multiple Faults. Springer London, London, Ch. Chapter 12, pp. 61–100.
URL https://doi.org/10.1007/978-1-84882-775-2_4

Kennel, M. B., Brown, R., Abarbanel, H. D., 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A 45 (6), 3403.

Kocis, L., Whiten, W. J., Jun. 1997. Computational investigations of low-discrepancy sequences. ACM Transactions on Mathematical Software 23 (2), 266–294.
URL http://doi.acm.org/10.1145/264029.264064

Lang, X., Zhang, Z., Xie, L., Horch, A., Su, H., 2018. Time-frequency analysis of plant-wide oscillations using multivariate intrinsic time-scale decomposition. Industrial & Engineering Chemistry Research 57 (3), 954–966.

Lee, K. H., Ren, Z., Huang, B., 2008. Novel closed-loop stiction detection and quantification method via system identification. In: In ADCONIP Conference, Jasper, Canada. pp. 4–7.

Li, X., Wang, J., Huang, B., Lu, S., 2010. The DCT-based oscillation detection method for a single time series. Journal of Process Control 20 (5), 609–617.

Miao, T., Seborg, D. E., 1999. Automatic detection of excessively oscillatory feedback control loops. In: Proceedings of the IEEE International Conference on Control Applications, 1999. Vol. 1. IEEE, pp. 359–364.

Naghoosi, E., Huang, B., 2014. Automatic detection and frequency estimation of oscillatory variables in the presence of multiple oscillations. Industrial & Engineering Chemistry Research 53 (22), 9427–9438.

Niederreiter, H., 1992. Random number generation and quasi-Monte Carlo methods. Vol. 63. Society For Industrial and Applied Mathematics (SIAM).

Olsson, H., 1996. Control Systems with Friction: PhD thesis. Lund Institute of Technology, Lund, Sweden.

Peng, Z., Peter, W. T., Chu, F., 2005. A comparison study of improved Hilbert Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. Mechanical Systems and Signal Processing 19 (5), 974–988.

Pereda, E., Rial, R., Gamundi, A., Gonzalez, J., 2001. Assessment of changing interdependencies between human electroencephalograms using nonlinear methods. Physica D: Nonlinear Phenomena 148 (1), 147–158.

Rehman, N., Mandic, D. P., 2010a. Multivariate empirical mode decomposition. In: Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 466. The Royal Society, pp. 1291–1302.

Rehman, N. U., Mandic, D. P., 2010b. Empirical mode decomposition for trivariate signals. Signal Processing, IEEE Transactions on 58 (3), 1059–1068.

Rehman, N. U., Mandic, D. P., 2011. Filter bank property of multivariate empirical mode decomposition. Signal Processing, IEEE Transactions on 59 (5), 2421–2426.

Rehman, N. U., Park, C., Huang, N. E., Mandic, D. P., 2013. EMD via MEMD: Multivariate noise-aided computation of standard emd. Advances in Adaptive Data Analysis 5 (02), 1350007.

Rilling, G., Flandrin, P., Gonçalves, P., Lilly, J. M., 2007. Bivariate empirical mode decomposition. Signal Processing Letters, IEEE 14 (12), 936–939.

Rilling, G., Patrick, F., Paulo, G., 2003. On empirical mode decomposition and its algorithms. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing 3, 8–11.

Rossi, M., Scali, C., Sep. 2012. A comparison of techniques for automatic detection of stiction: simulation and application to industrial data. Journal of Process Control 6(3), 689–695.

Sauer, T., Yorke, J. A., Casdagli, M., 1991. Embedology. Journal of Statistical Physics 65 (3), 579–616.

Schreiber, T., Schmitz, A., 1996. Improved surrogate data for nonlinearity tests. Physical Review Letters 77 (4), 635.

Srinivasan, B., Rengaswamy, R., 2012. Automatic oscillation detection and characterization in closed-loop systems. Control Engineering Practice 20 (8), 733–746.

Srinivasan, R., Rangaswamy, R., Miller, R., 2005a. Control loop performance assessment. 1. a qualitative approach for stiction diagnosis. Industrial Engineering Chemistry Research 44, 6708–6718.

Srinivasan, R., Rangaswamy, R., Narasimhan, S., Miller, R., 2005b. Control loop performance assessment. 2. a Hammerstein model approach for stiction diagnosis. Industrial Engineering Chemistry Research 44(17), 6719–6728.

Srinivasan, R., Rengaswamy, R., Miller, R., 2007. A modified empirical mode decomposition (emd) process for oscillation characterization in control loops. Control Engineering Practice 15 (9), 1135–1148.

Starr, K. D., Petersen, H., Bauer, M., 2016. Control loop performance monitoring ABB's experience over two decades. IFAC-PapersOnLine 49 (7), 526–532.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338 (6106), 496–500.

Takens, F., 1981. Detecting strange attractors in turbulence. In: Dynamical systems and turbulence, Warwick 1980. Springer, pp. 366–381.

Tangirala, A., Shah, S., Thornhill, N., 2005. PSCMAP: A new tool for plant-wide oscillation detection. Journal of Process Control 15 (8), 931–941.

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

Theiler, J., Prichard, D., 1996. Constrained-realization Monte-Carlo method for hypothesis testing. Physica D: Nonlinear Phenomena 94 (4), 221–235.

Thornhill, N., Shah, S., Huang, B., 2001. Detection of distributed oscillations and root-cause diagnosis. In: Proceedings of CHEMFAS 4, june 7-8, Jejudo (Chejudo) Island, Korea. pp. 167–172.

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

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

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

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

Torres, M. E., Colominas, M., Schlotthauer, G., Flandrin, P., et al., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, pp. 4144–4147.

Wang, Y., Wang, H., Zhang, Q., Aug. 2012. Non-linear distortion identification based on intra-wave frequency modulation. An International Journal of Applied Mathematics and Information Sciences 15(5), 505–514.

Wardana, A. N. I., Oct 2015. A method for detecting the oscillation in control loops based on variational mode decomposition. In: 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA). pp. 181–186.

Wiener, N., 1956. The theory of prediction. Modern mathematics for engineers.

Wong, T.-T., Luk, W.-S., Heng, P.-A., 1997. Sampling with Hammersley and Halton points. Journal of graphics tools 2 (2), 9–24.

Wu, Z., Huang, N. E., 2004. A study of the characteristics of white noise using the empirical mode decomposition method. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 460. The Royal Society, pp. 1597–1611.

Wu, Z., Huang, N. E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis 1 (01), 1–41.

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

Xia, C., Howell, J., Thornhill, N. F., 2005. Detecting and isolating multiple plant-wide oscillations via spectral independent component analysis. Automatica 41 (12), 2067–2075.

Xia, C., Zheng, J., Howell, J., 2007. Isolation of whole-plant multiple oscillations via non-negative spectral decomposition. Chinese Journal of Chemical Engineering 15 (3), 353–360.

Xie, L., Lang, X., Chen, J., Horch, A., Su, H., 2016a. Time-varying oscillation detector based on improved lmd and robust lempel–ziv complexity. Control Engineering Practice 51, 48–57.

Xie, L., Lang, X., Horch, A., Yang, Y., 2016b. Online oscillation detection in the presence of signal intermittency. Control Engineering Practice 55, 91–100.

Yamashita, Y., 2005. An automatic method for detection of valve stiction in process control loops. Control Engineering Practice 14, 503–510.

Yeh, J.-R., Shieh, J.-S., Huang, N. E., 2010. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. Advances in Adaptive Data Analysis 2 (02), 135–156.

Yuan, T., Qin, S. J., 2014. Root cause diagnosis of plant-wide oscillations using granger causality. Journal of Process Control 24 (2), 450–459.

Zang, X., Howell, J., 2003. Discrimination between bad tuning and non-linearity induced oscillations through bispectral analysis. In: SICE 2003 Annual Conference. Vol. 1. IEEE, pp. 896–900.

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

Zang, X., Howell, J., 2007. Isolating the source of whole-plant oscillations through bi-amplitude ratio analysis. Control Engineering Practice 15 (1), 69–76.

Part III

Appendices

Appendix A

Paper F

Paper F: An Adaptive Non-Linearity Detection Algorithm for Process Control Loops

Muhammad Faisal Aftab, Morten Hovd, Norden E. Huang, Selvanathan Sivalingam

IFAC-PapersOnLine vol 49(7), 1020 - 1025

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online
CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-7 (2016) 1020–1025

An Adaptive Non-Linearity Detection Algorithm for Process Control Loops ⋆

Muhammad Faisal Aftab * **Morten Hovd** **
Norden E. Huang *** **Selvanathan Sivalingam** ****

* *Department of Engineering Cybernetics, NTNU, Trondheim Norway (e-mail: muhammad.faisal.aftab@itk.ntnu.no).*
** *Department of Engineering Cybernetics, NTNU, Trondheim, Norway (e-mail: morten.hovd@itk.ntnu.no)*
*** *Research Center for Adaptive Data Analysis, National Central University, Taiwan, (e-mail: Norden@ncu.edu.tw)*
**** *Seimens AG, Trondheim, Norway*

Abstract: Non-linearities are considered to be a major source of oscillations and poor performance in industrial control systems, as 20-30 % of loops are reported to be oscillating due to valve non-linearities (Srinivasan et al. (2005)). This fact has led to a significant effort aimed at the detection and diagnosis of non-linearities; in particular for valve non-linearities in the control loops. The current paper presents an adaptive algorithm, based on HHT (Hilbert Huang Transform), for non-linearity detection and isolation in process systems. The HHT is an adaptive data analysis technique that is applicable to non-linear and non-stationary time series. An index termed the *Degree of Non-Linearity* (DNL), based on intra-wave frequency modulation, is used to identify the presence of non-linearity in the signal generating system. The proposed method is shown to be more robust in differentiating between linear and non-linear causes of oscillations when compared to existing methods, and can handle non-stationary effects.

Keywords: Control non-linearities, adaptive algorithm, intra-wave frequency modulation, instantaneous frequency

1. INTRODUCTION

Control loop performance assessment (CLPA) is an important concept owing to the fact that only about one-third of all industrial control loops are reported to be giving satisfactory performance (Srinivasan et al. (2005)). These performance issues can have adverse effect on the productivity and profitability of any industrial process. Poor performance may manifest itself as poor set point tracking, excessive control actions and presence of oscillations. Among these manifestations of poor performance, the detection of oscillations and identification of their root cause has attracted significant research recently, as oscillations are the most common indicator of performance degradation. There can be a variety of different sources of oscillations, such as poor controller tuning, presence of non-linearities, disturbances etc. The high complexity of industrial systems necessitates the provision of robust monitoring mechanism that can detect the root cause of the oscillations, hence minimizing the critical maintenance time, energy consumption and ensuring product quality.

Of all sources of oscillations in control loops, valve non-linearities (backlash, stiction, hysteresis) are considered to be one of the major causes of oscillations. Therefore, distinguishing between linear and non-linear causes of oscillations is an important aspect of performance monitoring.

The oscillation characterization process can be divided into sub-categories like oscillation detection, grouping the loops oscillating due to (apparently) the same cause, and oscillation diagnosis. Oscillation detection is mainly concerned with the detection of oscillating loops and related characteristics such as the frequency and the amplitude of oscillations. Different methods to detect oscillation in individual loops as well as plant wide oscillations are reported in literature. Hägglund (1995) proposed a procedure based on monitoring the Integral Absolute Error (IAE) to detect the oscillations, whereas Thornhill et al. (2003) proposed to use the Auto Covariance Function (ACF) for detection and characterization of oscillations. A Modified Empirical Model Decomposition (EMD) method is used by Srinivasan et al. (2007). For grouping loops with similar oscillation patterns (and presumably also a common cause of oscillations), the use of tools like the Power Spectral Correlation Map (PSCMAP) (Tangirala et al. (2005)) and Principal Component Analysis (PCA) (Thornhill et al. (2002)) have been proposed.

The diagnosis part, that deals with the identifying the root cause of oscillations, is rather tricky owing to complex underlying plant dynamics and presence of multiple sources of oscillations. Most of the research revolves around differentiating between linear and non-linear causes of oscilla-

⋆ Financial support to the first author from Siemens AS, Norway, is gratefully acknowledged.

10.1016/j.ifacol.2016.07.336

tions; with non-linearities in process systems often being associated with valves, in particular due to valve stiction. Data driven approaches to detect non-linearities in the control loops are more practical as they do not require an accurate model of the process dynamics, which is seldom available. Some important non-linearity detection methods will be discussed in this paper; while readers are referred to Thornhill and Horch (2007) and Capaci and Scali (2015) for a more complete of oscillation detection and grouping methods.

A simple approach is to use the cross correlation between controller and plant output to differentiate between linear and stiction caused oscillations, for stable and non-integrating plant, is presented in Horch (1999). An odd correlation indicates the presence of stiction, whereas an even correlation indicates other causes for oscillation.

Different shape analysis formalism methods are proposed in Hägglund (2011) and Srinivasan et al. (2005) where the shape of the process output is compared with different patterns like square, triangular and sinusoidal waves to assess the presence (or absence) of stiction. Another similar shape analysis method is given in (Yamashita (2005)) where the phase portrait of the input and output of the controller is used to detect the presence of sticton.

It is well known that non-linearities in valves can generate limit cycles that have waveforms different from ordinary sinusoids, thereby giving rise to harmonics in the power spectrum. These harmonics can be used as a signature of non-linearity (Thornhill and Horch (2007)), but the high frequency peaks in the power spectrum due to fast disturbances may be mistaken for the non-linearity. Chaudhry et al. (2004) proposed a non-linearity detection method based on higher order statistics. The method claims to distinguish between oscillations caused by linear or non-linear sources, with a prior assumption of stationary plant data; the assumption of stationarity may not hold in actual practice. The limitations of cross correlation and bi-spectrum based techniques are highlighted in Rossi and Scali (2012). All the existing methods for detecting non-linearity in control loops thus rely on certain prior assumptions or conditions that may not be fulfilled in actual practice, therefore limiting the practical scope and application. This is not intended to say that these methods cannot be useful in practice, but rather that some care and expertise is required in their application.

In an effort to overcome the shortcomings in the listed approaches, an adaptive non-linearity detection method based on the Hilbert Huang Transform (HHT) Huang et al. (1998) is proposed in this paper. The essence of the concept is that the non-linearity manifests itself as intra-wave frequency modulation in the instantaneous frequency (IF). Babji et al. (2009) used the same concept to detect non-linearity in the control loops, but the procedure required manual inspection of instantaneous frequency plot and is therefore not applicable for automated non-linearity detection.

In this paper we propose the use of a non-linearity index called the degree of non-linearity (DNL), based on intra-wave frequency modulation, to detect the presence of non-linearity in the process control loop. This method is widely applicable and not limited to valve non-linearities. Furthermore, in this work, it is assumed that oscillations are being detected and objective is restricted to identify the linear or non-linear source of these oscillations. The proposed method can handle non-stationary time series, which is an inherent benefit of the HHT. This paper is organized as follows. Section 2 gives a brief overview of HHT and instantaneous frequency concepts. Section 3 presents the proposed non-linearity measurement index, DNL, that can be used to detect the non-linearity. Simulation studies are presented in Section 4.

2. HILBERT-HUANG TRANSFORM (HHT)

2.1 Overview

The Hilbert Huang Transform (HHT) is an adaptive data processing tool, recently developed by Huang et al. (1998). The HHT is finding applications in many different areas due to its adaptive nature and ability to handle non-stationary and non-linear time series. The procedure involves the decomposition of the time series into components called Intrinsic Mode Functions (IMFs). IMFs can be used to calculate the instantaneous frequency (IF) through the application of the Hilbert transform. In order to get a meaningful IF the IMFs must fulfil certain criteria, i.e. they must be symmetric, zero mean and their number of extrema and zero crossings should at-most differ by one. The component IMFs are generated through a decomposition method called Empirical Mode Decomposition (EMD).

The EMD process sifts out high frequency components from the data by iteratively subtracting low frequency components. These low frequency components are local means $m(t)$ of the envelope defined by spline fitting of the extrema.

$$d(t) = x(t) - m(t) \tag{1}$$

where $d(t)$ represents the local high frequency component Rilling et al. (2003). The sifting process is iterated on d until it qualifies as an IMF; and named as $c_1(t)$. Once the IMF is extracted it is subtracted from the original signal and the sifting procedure is started again on the residue. This continues until there are no more IMFs to be extracted. If $c_i(t)$ is the i^{th} IMF and $r(t)$ is residue, the sifting procedure gives

$$x(t) = \sum_{i=1}^{N} c_i(t) + r(t) \tag{2}$$

where $x(t)$ is the input time series and N is the total no of IMFs. The details of the procedure can be seen in Huang et al. (1998) and Rilling et al. (2003).

2.2 Instantaneous Frequency (IF)

An important consequence of the HHT is the Instantaneous frequency (IF) for time-frequency analysis. The Instantaneous frequency is calculated by the application of the Hilbert Transform to the IMF. A brief overview of the IF calculation is provided here. The Hilbert transform $Y(t)$ of a signal $X(t)$, also regarded as convolution of $x(t)$ and $1/\pi t$, is given by

$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(\tau)}{t-\tau} d\tau = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(t-\tau)}{\tau} d\tau \quad (3)$$

where P indicates Cauchy's principal value of the integral. The analytic signal $Z(t)$ is therefore defined as

$$\begin{aligned} Z(t) &= X(t) + jY(t) = a(t)e^{j\theta t} \\ a(t) &= \sqrt{X^2(t) + Y^2(t)}, \quad \theta(t) = \arctan \frac{Y(t)}{X(t)} \end{aligned} \quad (4)$$

Here the amplitude and phase are functions of time and the instantaneous frequency (IF) is defined as the time derivative of the phase function $\theta(t)$; given by (Wu and Huang (2009))

$$\omega(t) = \frac{d\theta(t)}{dt} = \frac{1}{A^2}[X\dot{Y} - Y\dot{X}] \quad (5)$$

The original signal $x(t)$ can be represented as

$$x(t) = \Re(a(t)e^{j\int \omega t}) \quad (6)$$

where $\Re$ is the real part of the analytic signal. The instantaneous frequency gives insight into the characteristics of the signal under analysis and can be used to detect non-linearity, as explained in the next section.

2.3 Discarding Spurious IMFs

The EMD process may generate spurious IMFs, as large swings at the ends of spline fitting may creep inwards and generate additional IMFs than necessary (Peng et al. (2005)). These spurious IMFs can be regarded as pseudo-components of the input signal and must be eliminated. As the IMFs generated through the EMD process are nearly orthogonal components, the significant IMFs will be more correlated with the input signal than the pseudo-components (Peng et al. (2005)). The correlation coefficient, therefore can be used to extract significant IMFs as reported in (Peng et al. (2005)) and (Sirinavasan and Rengasawamy (2012)). The correlation coefficient ρ_i of normalized i^{th} IMF c_i with normalized signal $x(t)$ is calculated from

$$\rho_i = \frac{Cov(c_i, x)}{\sigma_x \sigma_{c_i}}, \quad i = 1, 2, 3 \ldots n \quad (7)$$

where Cov is the covariance of i^{th} IMF and input signal $x(t)$; σ_x and σ_{c_i} are standard deviations of signal and IMF respectively and n is total number of IMFs. IMFs with normalized coefficient $\lambda > 0.5$ are retained , while the others are eliminated and added to the residue.

$$\lambda_i = \frac{\rho_i}{max(\rho_i)}, \quad i = 1, 2, 3 \ldots n \quad (8)$$

3. NON-LINEARITY DETECTION

3.1 Intra-wave Frequency Modulation

It is an established fact that non-linearity, in HHT, manifests itself as an intra-wave frequency modulation, i.e fluctuation of the IF within one period of oscillation (Huang et al. (1998),Babji et al. (2009),Wang et al. (2012)). The same fact is used in this paper to detect non-linearity in process control loops. The intra-wave frequency modulation in a non-linearly distorted wave form can further be elaborated by a simple example.

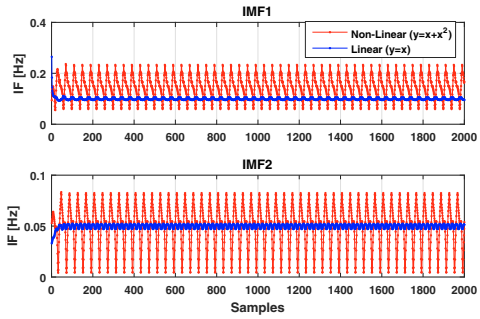

Fig. 1. Non-linearity effect: Intra-wave frequency modulation

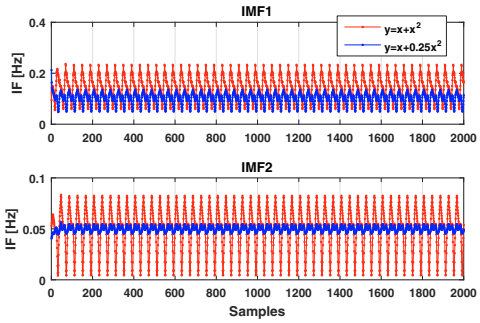

Fig. 2. Intra-wave frequency modulation spread increases with increased non-linearity

Example I: Intra-wave Frequency Modulation A signal is generated by sum of two sinusoids with frequencies $f_1 = 0.05Hz$ and $f_2 = 0.1Hz$. This signal is added to its square to create a non-linearly distorted waveform as given in (9).

$$\begin{aligned} x(k) &= \sin(2\pi f_1 k) + \sin(2\pi f_2 k) \\ y(k) &= x(k) + x^2(k) \end{aligned} \quad (9)$$

Signals $x(k)$ and $y(k)$ are subjected to EMD process to generate 2 IMFs each. The corresponding IF plots, given in figure 1, show significant intra-wave frequency modulation for the non-linearly distorted waveform $y(k)$).

3.2 Non-linearity Measurement Index

To develop the non-linearity index, it is being highlighted that the presence of non-linearity effects can be identified by intra wave frequency modulation. The difficult issue is to have a quantitative measure that can automatically detect the presence of non-linearity. The spread of the intra wave frequency modulation is proportional to the extent of non-linearity in the signal (IMF) as highlighted in figure 2.

As for the non-linearly distorted waveforms the IF changes within one oscillation cycle with variation proportional to the extent of non-linearity; mathematically we can define the degree of non-linearity in an IMF as (Huang et al. (2014))

$$DNL \propto var(IF)$$
$$DNL \propto \left\langle \left\{ \frac{IF - IF_z}{IF_z} \right\}^2 \right\rangle^{1/2} \quad (10)$$

where IF is the instantaneous frequency and IF_z is the full wave zero crossing frequency. However, it would be more appropriate for the measure to also include the effect of the amplitude in the DNL; as the distorted wave (non-linearity) with large amplitude should contribute more towards the degree of non-linearity. Therefore DNL, weighted by the amplitude,for an i^{th} IMF can be defined as (Huang et al. (2014))

$$DNL_i = std\left\langle \left\{ \frac{IF_i - IF_{z_i}}{IF_{z_i}} \right\} . \frac{a_{z_i}}{\bar{a}_{z_i}} \right\rangle \quad (11)$$

where a_z is the zero crossing amplitude; defined as the absolute value of the extremum between successive zero crossings, $\bar{a}_z$ is the mean of a_z and *std* represents the standard deviation of the resulting vector. Equation (11) gives the measure of Non-Linearity for individual IMF; but the actual signal may consists of number of IMFs. Therefore a **Total non-linearity (TDNL)** for the complete signal consisting of N IMFs can be given as sum of individual DNLs weighted by the energy of each IMF. The TDNL is given by (Huang et al. (2014))

$$TDNL = \sum_{j=1}^{N} \left\langle DNL_j \frac{|c_j|^2}{\sum_{k=1}^{N} |c_k|^2} \right\rangle \quad (12)$$

Where $|c_j|^2$ is 2-norm of j^{th} IMF.

3.3 Non-linearity Measure Threshold

Ideally the linear signal will have no intra-wave frequency modulation and hence zero DNL and TDNL; but in actual practice the linear signal may have slight fluctuations in IF due to spline fitting issues. This will be further elaborated by application of the non-linearity measure to the signal in Example-I. Though the exact threshold value to distinguish between linearity and otherwise is under investigation; for preliminary analysis threshold value of 0.1 is assumed in this work. Therefore detection of linear or non-linear source of oscillations is based on the rule summarized in Table 1.

Table 1. Non-linearity detection threshold

if DNL > 0.1
then non-linear source
else
linear source

3.4 Non-Linearity Measure for Example I

The DNL defined in (11) and (12) is applied to the signals in Example-I; the results are summarised in Table 2. The indices calculated in Table 2 show that the proposed measures can detect the varying non-linearity quite well. The strength of the proposed scheme in differentiating between linear and non-linear causes of oscillations are further established by simulation results of closed loop system with First Order plus Delay Time (FOPDT) plant model and PI controller in Section 4.

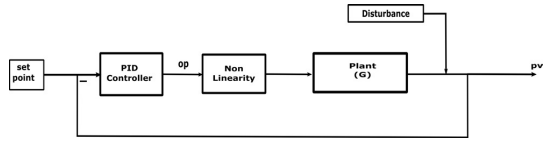

Fig. 3. Closed loop model

4. SIMULATION EXAMPLE

The objective is to distinguish between linear and non-linear causes of oscillation using the non-linearity measure given in section 3. The same scenario is discussed in the bi-spectrum method proposed by Chaudhry et al. (2004); but the scheme proposed in this paper can work even in case of transient effects and non-stationary data. The simulation example is single input single output closed loop system with a FOPDT plant and a PI controller. The plant dynamics are given by

$$G(s) = \frac{2.25}{4.5s + 1} e^{-3s} \quad (13)$$

Nominal PI controller gains are $K_c = 0.1$ and $K_i = 0.05$. The block diagram of the simulation setup is shown in Figure 3. The non-linearity is modelled as stiction in the valve; whereas for linear cases the non-linear block is removed. Simulation data is recorded for 4000 seconds with a sampling rate of 1 sec. The transient effects are also included to prove the effectiveness of the proposed scheme in the case of non-stationary data. In the following, oscillations are induced in the closed loop, with both linear and non-linear causes, in order to illustrate the ability of the HHT to differentiate between these causes.

4.1 Inappropriate controller tuning

The simulation is done with the stiction block removed and with increased integral gain in the controller ($K_i = 0.19$), thereby inducing oscillations in the system. The EMD process generates four IMFs but only the first one is significant and hence retained for non-linearity analysis; the corresponding oscillatory response and IMF is shown in row 1 of Figure 4. The Degree of Non-linearity (DNL), is calculated to be 0.057. The value establishes the absence of any non-linear element in the closed loop system. It is to be noted that the data also contains transient effects that can cause false detection of non-linearity if analysed by other tools such as the bi-spectrum based method.

4.2 External Sinusoidal Disturbance

The closed loop system is subjected to external sinusoidal disturbance with amplitude 5 and frequency $0.25 rad.sec^{-1}$. Here again only the first IMF is significant and is therefore analysed. The plant output and IMF are shown in row 2 of Figure 4. The DNL value of 0.023 again confirms the absence of any non-linearity as a source of oscillations.

Table 2. Non-linearity Example I

Signal	DNL (IMF1)	DNL (IMF2)	TDNL
$x(k)$	0.04	0.04	0.04
$x(k) + 0.25x^2(k)$	0.21	0.05	0.13
$x(k) + x^2(k)$	0.29	0.47	0.37

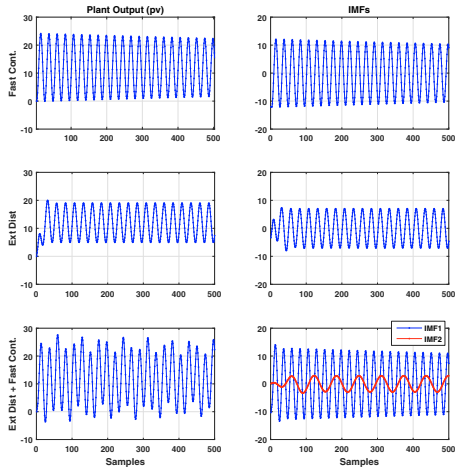

Fig. 4. Plant output and significant IMFs for oscillations from linear sources

4.3 Oscillations due to Multiple Linear Effects

In this case both increased integral action and sinusoidal disturbance of amplitude 10 and frequency $0.1 rad.sec^{-1}$ are applied to generate oscillatory response of the system. The two IMFs showing two distinct frequency bands, one because of compromised loop stability and other due to external disturbance, and plant output are shown in row 3 of Figure 4. The non-linearity measures for both the IMFs are 0.036 and 0.022 ; indicating linear cause of oscillation. Again the presence of transient/ non-stationary effects are emphasized here that will hinder the diagnostics in other methods. The instantaneous frequencies (IF), given in row 1 of Figure 6, also shows the absence of any non-linearity as there are no intra-wave frequency modulations in the IF.

4.4 Stiction Non-linearity

In order to assess whether the proposed scheme can identify the presence of non-linearity as source of oscillations, the stiction non-linearity is introduced in the closed loop system as shown in Figure 3. Stiction is simulated using both data driven and physical models to illustrate the robustness of the proposed scheme. The data driven model used is the two parameter model, taken from Chaudhry et al. (2008); with stick-band S=7 and slip jump J=5, while the physical model is the LuGre model presented in Olsson (1996); the details and parameters of LuGre model are given in appendix A. The process output for both the models is shown in Figure 11. The parameters for both models are tuned to get similar response.The non-linearity measures for data driven model and LUGRe model are found to be 1.6 and 1.38 respectively; thereby indicating presence of similar extent of non-linearity in the system. The IF plot for both models, given in row 2 of Figure 6, shows significant intra-wave frequency modulation.

The non-linearitry meaures for different cases are summarized in Table 3.

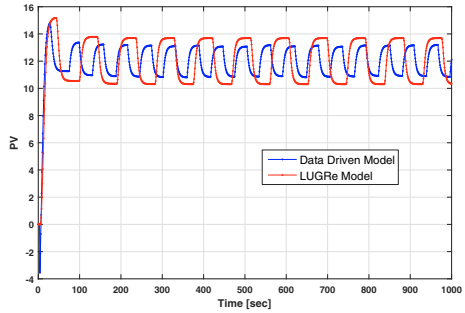

Fig. 5. Plant output: oscillations from non-linear sources

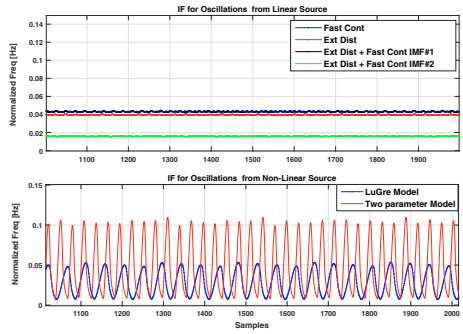

Fig. 6. Instantaneous frequencies (IF) for linear and non-linear sources of oscillations

Table 3. Non-linearity measures (closed loop FOPDT process)

Signal	DNL(IMFs)	TDNL
Fast Cont	0.057	0.057
Ext Dist	0.023	0.023
Ext Dist + Fast Cont	0.036 & 0.022	0.024
stiction (data driven model)	1.6	1.6
stiction (physical model)	1.38	1.38

5. CONCLUSIONS

In this paper an adaptive non-linearity detection algorithm based on the HHT is presented. The method can differentiate between linear and non-linear causes of oscillations solely from the recorded data and doesn't require any prior knowledge about the underlying process dynamics. The proposed algorithm can handle transient and non stationary effects that can limit the effectiveness of existing performance monitoring tools to a great extent. The application of the scheme to actual plant data is in progress and results will be published later.

REFERENCES

Babji, S., Gorai, P., and Tangirala, A. (2009). Detection and quantification of control valve non linearities using hilbert huang transform. *Advances in Adaptive Data Analysis*, 1 No 3, 425–446.

Capaci, R.D. and Scali, C. (2015). An review on valve stiction. part i: From modelling to smart diagnosis. *processes 2015*, 3, 422–451.

Chaudhry, M., Shah, S.L., and Thornhill, N.F. (2004). Diagnosis of poor control-loop performance using higher-order statistics. *Automatica*, 40, 1719–1728.

Chaudhry, M., Shah, S.L., and Thornhill, N.F. (2008). *Diagnosis of process nonlinearities and valve stiction; data driven approaches*. Springer, Berlin.

Hägglund, T. (1995). Control-loop performance monitor. *Control Engineering Practice*, 3(11), 1543–1551.

Hägglund, T. (2011). A shape analysis approach for diagnosis of stiction in control valves. *Control Engineering Practice*, 40, 782–789.

Horch, A. (1999). A simple method for detection of stiction in control valves. *Control Engineering Practice*, 7(10), 6708–6718.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., and Liu, H.H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *R. Soc. Lond. A*, 454, 903–995.

Huang, N.E., Lo, M.T., Zhao-Hua, W., and Xian-Yao, C. (2014). Method for quantifying and modeling degree of nonlinearity, combined nonlinearity, and nonstationarity. *US Patent 8,732,113*.

Olsson, H. (1996). *Control Systems with Friction: PhD thesis*. Lund Institute of Technology, Lund, Sweden.

Peng, Z.K., Tse, P.W., and Chu, F. (2005). A comparison study of improved hilbert-huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, 19(5), 974–988.

Rilling, G., Flandrin, P., and Goncalves, P. (2003). On empirical mode decomposition and its algorithms. *IEEE-EURASIP workshop on nonlinear signal and image processing*, 3, 8–11.

Rossi, M. and Scali, C. (2012). A comparison of techniques for automatic detection of stiction: simulation and application to industrial data. *Journal of Process Control*, 6(3), 689–695.

Sirinavasan, B. and Rengasawamy, R. (2012). Automatic oscillation detection and characterization inc ontrol loop. *Control Enginnering Practice*, 20, 733–746.

Srinivasan, R., Rangaswamy, R., and Miller, R. (2005). Control loop performance assessment. 1. a qualitative approach for stiction diagnosis. *Industrial Engineering Chemistry Research*, 44, 6708–6718.

Srinivasan, R., Rengaswammy, R., and Miller, R. (2007). A modified empirical mode decomposition (emd) process for oscillations characterization in control loops. *Control Engineering Practice*, 15, 1135–1148.

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

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

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

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

Wang, Y., Wang, H., and Zhang, Q. (2012). Non-linear distortion identification based on intra-wave frequency modulation. *An International Journal of Applied Mathematics and Information Sciences*, 15(5), 505–514.

Wu, Z. and Huang, N.E. (2009). On instantaneous frequency. *Advances in Adaptive Data Analysis*, 1(2), 177–229.

Yamashita, Y. (2005). An automatic method for detection of valve stiction in process control loops. *Control Engineering Practice*, 14, 503–510.

Appendix A. LUGRE FRICTION MODEL

The LuGre model used in the simulation is based on the work by Olsson (1996). A brief overview of the model is discussed here. The distance x travelled by the valve stem under the influence of the applied force F_e is governed by Newton's second law as

$$m\ddot{x} = F_e + F_f + F_k \tag{A.1}$$

where F_f and F_k are friction and spring forces respectively. The spring force given by $-kx$, with spring constant k acting as a proportional control action, and the valve stops when the spring force becomes equal to applied force F_e. The frictional force F_f arises due to the friction the stem experiences while travelling through the packaging element, the details can be seen in Chaudhry et al. (2008). The frictional force $F_f$is modelled as

$$\begin{aligned} \dot{z} &= v - \frac{|v|}{g(v)}z \\ F_f &= \sigma_0 z + \sigma_1 \dot{z} + F_v v \\ g(v) &= \frac{1}{\sigma_0}(F_c + (F_s - F_c)e^{-(v/v_s)}) \end{aligned} \tag{A.2}$$

where F_s is the static friction; F_c is Coulomb friction and F_v is the coefficient of viscous friction, v_s is the Stribeck velocity, $\sigma_0 > 0$ is stiffness and $\sigma_1 > 0$ is velocity dependant damping coefficient. z is an intermediate variable representing relative average deflection of bristles (friction is modelled as a contact among bristles); (Olsson (1996)) The values of these parameters used in simulation are given in table A.1. The simulations are done with unit mass

Table A.1. LuGre model parameters

F_s	F_c	F_v	σ_0	σ_1	v_s
1	0.5	0.3	10^4	$0.2\sqrt{10^5}$	0.01

actuator and spring constant $k = 1$.

Appendix B

Paper G

Paper G: A Delay Vector Variance based Approach for Detecting and Isolating the Non-linearity Induced Oscillations in Control Loops

Muhammad Faisal Aftab, Morten Hovd, Selvanathan Sivalingam

IFAC-PapersOnLine vol 50, 7975 - 7980

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 7975–7980

A Delay Vector Variance based Approach for Detecting and Isolating the Non-linearity Induced Oscillations in Control Loops ⋆

Muhammad Faisal Aftab * **Morten Hovd** *
Selvanathan Sivalingam **

* *Department of Engineering Cybernetics, NTNU, Trondheim Norway (e-mail: muhammad.faisal.aftab;morten.hovd@itk.ntnu.no).*
** *Seimens AG, Trondheim, Norway*

Abstract: Non-linear time series analysis based methods are a popular choice for industrial control loop data analysis. In this paper a delay vector variance (DVV) based approach is presented to analyze the source of oscillations in an industrial control loop. The method is capable of differentiating between the linear and non-linear causes of oscillations and can also help in isolating the source of non-linearity. The automatic determination of embedding dimensions is augmented with the DVV analysis to make it more robust and reliable. The efficacy of the proposed method is established using simulation as well as industrial case studies.

Keywords: Delay vectors, embedding dimensions, rank statistics, target variance, scatter plot

1. INTRODUCTION

Data driven approaches for control loop performance monitoring (CPM) offer considerable advantages over conventional model based diagnosis. These data driven methods, dealing exclusively with recorded data, are more flexible, robust and are easy to automate. There are numerous factors affecting the performance of industrial control systems, such as poor controller tuning, external disturbances, equipment non-linearities etc. Therefore differentiating among these factors leads to enhanced profitability and reduced shut down time. The fact that only a third of industrial control loops are reported to be functioning correctly, while rest are suffering from one problem or the other [Choudhury et al. (2008)], makes performance monitoring an important part of any industrial control systems.

Non-linearities, whether stemming from a sticking valve or a faulty sensor, are reported to be one of the major sources of performance issues [Srinivasan et al. (2005)]. Therefore much focus has been on the detection and isolation of non-linearities in the control loops. Different methods are being explored and reported in literature [Thornhill and Horch (2007) and di Capaci and Scali (2015)], a brief overview of some procedures are discussed here.

Some methods, like the one proposed by Srinivasan et al. (2005), Hägglund (2011) and Yamashita (2005), try to find similarity of the waveform with predetermined sets.

A popular method to distinguish between non-linear and linear sources of oscillations in control loops based on higher order spectra is proposed by Choudhury et al. (2008). Inability to detect non-linearities in symmetric waveforms(exhibiting odd harmonics) such as triangular or square waves are major limitations of this method [Thornhill (2005) and Zang and Howell (2003)].

Quite recently non-linearity detection methods based on the Hilbert Huang Transform (HHT) have been presented [Babji et al. (2009) and Aftab et al. (2016)]. These method use the intra-wave frequency modulation to detect the presence of non-linearity induced oscillations. Though these methods are applicable to non-stationary time series and are fully data driven, the mode mixing in the empirical mode decomposition (the precursor of HHT) in the presence of noise, make these methods prone to falsely report non-linearities.

Non-linear time series analysis methods make use of the fact that time series from nonlinear systems exhibit phase coupling and hence are more predictable. Analysis based on surrogate time series is therefore used to differentiate between linear and nonlinear nature of time series. Surrogate time series are time series that share the same power spectrum with the original time series, but contain random phase. One such method, to detect the presence of non-linearity in control loops, is proposed by Thornhill (2005). The algorithm compares the predictability of the time series from control loops with its surrogate counterparts. The accuracy of this method depends on the tuning of different parameters like embedding dimensions, prediction horizon, number of nearest neighbors, etc. Moreover the method relies on a test statistic that assumes the prediction error to follow the Gaussian distribution, that may not be the case in practice [1] .

⋆ Financial support to the first author from Siemens AS, Norway, is gratefully acknowledged.

[1] Details can be seen in Thornhill (2005)

10.1016/j.ifacol.2017.08.992

This paper presents the delay vector variance method (DVV) for detecting the presence of non-linearity in the control loops. This method is also based on surrogate data, but is more general and doesn't require the tuning of different parameters. Surrogate data is generated using the iterative amplitude adjusted fourier transform method (IAFFT) proposed by Schreiber and Schmitz (1996). Moreover a rank based test statistic is adopted to test the Null hypothesis, i.e. whether the data is from a linear source, as recommended by Kantz and Schreiber (2004). Another important feature is the automatic determination of embedding dimensions that is augmented with this analysis to make it more robust and reliable.

The paper is organized as follows. Section 2 gives and overview of the delay vector variance (DVV) method. The steps involved in testing the Null hypothesis using rank based statistics are outlined in section 3. Automatic determination of embedding dimensions is explained in section 4. Section 5 gives the detailed algorithm. Simulation and industrial case studies are presented in sections 6 and 7 followed by conclusions.

2. DELAY VECTOR VARIANCE (DVV) METHOD

The delay vector variance method, developed recently [Gautama et al. (2004a)], tries to ascertain the nature of the time series via estimation of predictability and determinism. The method, as the name suggests, is based on the time delay embedding representation of time series $\mathbf{X(k)} = \{\mathbf{x(k)} \mid k = 1 \dots N\}$. The time series is represented as a set of delay vectors (DVs), the method of so called *phase space reconstruction* [Kantz and Schreiber (2004)], with embedding dimension m and time delay τ, given by $\mathbf{y(k)} = [x_{k-m\tau}, \dots, x_{k-\tau}]$. The each DV has a corresponding target, i.e. the next sample $x(k)$ [Gautama et al. (2004b)].

The delay vectors so obtained are the state space representation of the actual time series. The target variance σ^* is then calculated using the following steps [Gautama et al. (2004a),Gautama et al. (2004b)].

(1) The distance between DVs is calculated using the Euclidean norm, denoted by $||\mathbf{y(i)} - \mathbf{y(j)}||$ for $i \neq j$.
(2) The mean μ_d and standard deviation σ_d are computed for the calculated DV distances.
(3) All the DVs that lie within a certain distance τ_d are collected in sets Ω_k , written mathematically as
$$\Omega_k = \{\mathbf{y(i)} \mid \|\mathbf{y(i)} - \mathbf{y(j)}\| \leq \tau_d\} \quad (1)$$
(4) The threshold τ_d is taken from the uniformly sampled interval $[\mu_d - n_d\sigma_d; \mu_d - n_d\sigma_d]$; with n_d specifying the span over which the analysis is performed.
(5) The target variance σ_k^{*2}, is calculated for each set Ω_k that contains at least N_0 DVs.
(6) The average of these target variances σ_k over all sets divided by the variance of time series σ_x^2 gives the overall target variance σ^{2*}, given by Equation 2.
$$\sigma^{*2} = \frac{(1/K)\sum_{k=1}^{K} \sigma_k^{*2}}{\sigma_x^2} \quad (2)$$

The target variance σ^* is the inverse measure of the predictability of the time series. The target variance is plotted against the standardized distance measure r_d given by

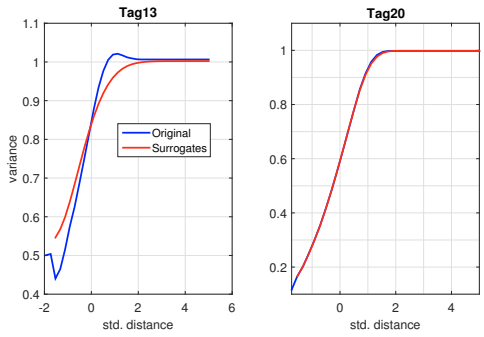

Fig. 1. A snapshot of target variance

$$r_d = \frac{\tau_d - \mu_d}{\sigma_d} \quad (3)$$

The greater the determinism in the time series the lower would be the value of target variance at smaller spans and vice versa. The target variance in (2) converges to unity for larger spans as for the maximum span all DVs belong to same set Ω_k and variance of Ω_k matches the variance of the data itself. The span parameter n_d must be increased if the target variance does not converge to unity. Figure 1 illustrates this effect graphically.

This target variance parameter not only gives the estimate of the determinism in the time series but also helps in classifying the time series as an output of linear or non-linear systems. This fact can be used to detect presence of non-linearity induced oscillations in industrial control loops as explained in subsequent sections.

2.1 Detection of Non-linearity using DVV Method

The surrogates of the input time series are used within the DVV framework to assess the linear or non-linear nature of the data from the oscillating control loops. The difference in the predictability of the original time series and its surrogate counterpart is taken as an indication of the non-linearity in the time series.

Figure 1 shows the target variance of two control loops from an industrial plant, along with the corresponding target variance of surrogate time series, to illustrate the DVV analysis results graphically. It can be seen that the target variance of data from Tag 13 is significantly different from its surrogate and is also more predictable; whereas for Tag 20 the loop data and its surrogates exhibit almost similar target variance. Moreover, another way of visualizing this relationship is using the so called DVV scatter plot; wherein the target variance of the time series is plotted versus that of its surrogates. For the linear case the target variance of the original time series and its surrogates would be similar and the scatter plot will be in close proximity of the bisector line whereas in the case of the non-linear signal this plot will diverge from the bisector line. This fact is shown in Figure 2 where Tag 13 scatter plot deviates clearly from the bisector line whereas for Tag 20 it is in close match with the bisector line.

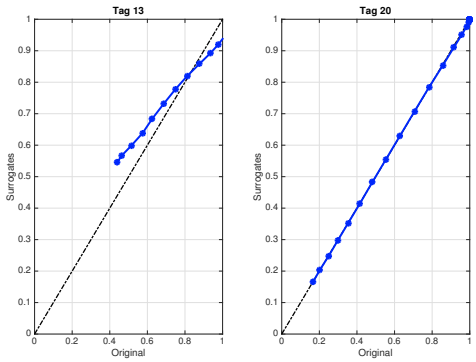

Fig. 2. DVV scatter plot ::A glance

2.2 Measuring the Extent of Non-linearity

The extent of non-linearity can be adjudged using the deviation of the target variance from the bisector line in the scatter plot. For this purpose the root mean square error (RMSE) from the bisector line seems to be an automatic choice to compare the extent of non-linearity in different time series. The rank based statistical test (described next) is used to differentiate between the linear and non-linear nature of time series and then the extent of non-linearity is calculated using the RMSE value.

2.3 Isolating the Source of Non-Linearity

Once the control loops suffering from the non-linearity induced oscillations are identified, the next step is to isolate the source for targeted maintenance and remedial actions. In a multi-loop environment, non-linearity induced at one point may propagate to other variables so the correct identification of the origin of the oscillations can significantly reduced the effort required in maintaining the control system. It is a well known fact that the different parts of plant act as low pass mechanical filter and tend to filter out the higher harmonics as we move away from the source of non-linearity [Thornhill et al. (2001)].

The RMSE measure discussed in section 2.2 can be used to compare the extent of non-linearity in different variables. The loop with maximum RMSE value is taken to be the source of non-linearity. The results, using RMSE, for the industrial case study are discussed in section 7.

3. TESTING OF THE NULL HYPOTHESIS: THE TEST STATISTICS

The rank based statistical test, as recommended by Kantz and Schreiber (2004), is adopted here to test the Null hypothesis that the time series is an output of linear Gaussian process. The rank based test is chosen as the probability distribution of the non-linearity measure is not known in advance; so a non-parametric rank based statistical test from Theiler and Prichard (1996) is adopted here to accept or reject the Null hypothesis.

The discriminating statistics chosen here is the RMSE value of the target variance. The RMSE η_0 is computed for the original time series. Then B surrogates are generated and corresponding statistics η_k are calculated for each $k = 1, \ldots, B$. Then it is checked whether the η_0 is on the tail of distribution. For one sided test the null hypothesis rejected at level α, if η_0 is among largest $(B+1)\alpha$ in the sorted list containing η_0 and all $\eta_k s$.

Table 1. Rank statistics for rejecting NULL hypothesis

if rank $> (B+1)(1-\alpha)$
then reject Null hypothesis (non-linear source)
else
accept Null hypothesis (linear source)

In this work, to test the Null hypothesis, $B = 99$ surrogates and $\alpha = 0.1$ are chosen. Therefore the Null hypothesis is rejected if the RMSE of the original time series in among largest 10 RMSEs out of total $100(r > 90)$ and time series is classified as originating from a non-linear source (Table 1). The rank based test and the RMSE values for the two signals in Figure 2 are given in Table 2. This procedure will be applied to simulation and industrial examples later in the paper to show how effective this method is in detecting the non-linearities in the control loops and isolating the root source.

Table 2. Rank statistic and RMSE

Signal	rank	Type	RMSE
Tag 20	4	Linear	–
Tag 13	99	Non-Linear	0.041

4. DETERMINING THE EMBEDDING DIMENSIONS

The construction of delay vectors, for the phase space reconstruction, is carried out using delayed embedding. The embedding dimension m of the attractor manifold is the most important parameter in such a reconstruction. Too small m will mask some of the dynamics and will lead to erroneous results whereas too large m will result in unnecessary computation overload. An automatic method for determination of the embedding dimension, proposed by Cao (1997), is augmented with the DVV to make it more robust and reliable.

According to the embedding theorems [Sauer et al. (1991), Cao (1997)] m is the true embedding if any two points which are close in the m-dimensional reconstructed space stay close in the $(m+1)$-dimensional reconstructed space as well. Therefore some kind of a distance measurement between two points in m and $m+1$ dimensional space can be used to identify the required dimension m.

Here again we consider a time series $\mathbf{X(k)} = \{\mathbf{x(k)} \mid k = 1 \ldots N\}$ and i^{th} delay vector $y_i(m)$ with embedding dimension m is given by

$$y_i(m) = (x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}) \qquad i = 1, 2, \ldots, N - m\tau \tag{4}$$

Now a distance measure $a(i, m)$ is defined as

$$a(i,m) = \frac{\|y_i(m+1) - y_{n(i,m)}(m+1)\|}{\|y_i(m) - y_{n(i,m)}(m)\|} \qquad i = 1, 2, \ldots, N - m\tau \tag{5}$$

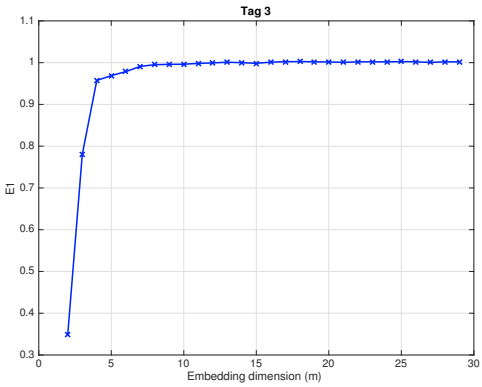

Fig. 3. Determining the embedding dimensions

where $\|.\|$ represents vector norm operation; $y_i(m+1)$ is the i^{th} delay vector with $m+1$ dimensions and $n(i,m)$ is integer in set $(1 \geq n \geq N - m\tau)$ such that $y_n(m)$ is the nearest neighbor of $y_i(m)$ in the m dimensional space.

The mean $E(m)$ of all $a(i,m)$s is defined as

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a(i,m) \tag{6}$$

The variation in E for the increase in embedding dimension to $m+1$ is given by the quantity E_1 defined as

$$E_1(m) = \frac{E(m+1)}{E(m)} \tag{7}$$

Therefore when E_1 stops changing for some $m \geq m_0$; $m_0 + 1$ is taken as the minimum embedding dimension for the reliable phase space reconstruction [Cao (1997)]. The E_1 measure for one of the data (Tag 3) is plotted in Figure 3. It can be seen that the E_1 approaches the unity value after $m = 8$ and stays there for all subsequent values of m. Therefore embedding dimension $m > 8$ can be used for the correct analysis.

5. PROPOSED METHOD

The proposed method for the identification of the non-linearity induced oscillations and isolation of the source using DVV is listed below.

(1) Mean center and normalize the time series from the control loop to unit standard deviation.
(2) End match the data to avoid spurious modes in the surrogate. End matching requires that the difference between both initial and final values d_0 and difference between initial and final gradient d_1 is minimized. d_0 and d_1 are calculated using (8) [Thornhill (2005)].

$$d_0 = \frac{(x_i - x_{i+n-1})^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2}$$
$$d_1 = \frac{[(x_{i+1} - x_i) - (x_{i+n-1} - x_{i+n-2})]^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2} \tag{8}$$

where x_i is the i^{th} element and $\bar{x}$ is the mean of time series $x_i \ldots x_{i+n-1}$.

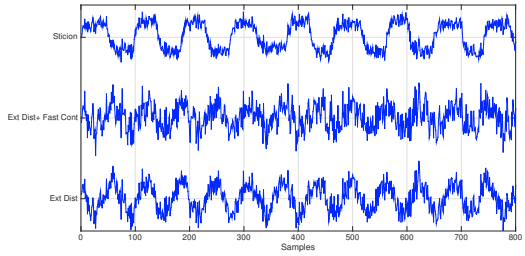

Fig. 4. Time trends from simulation example

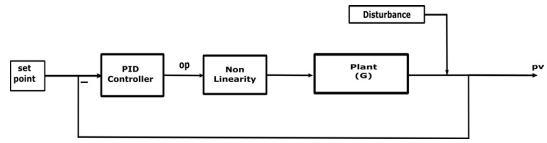

Fig. 5. Closed loop model

(3) Determine the minimum embedding dimension for the phase space reconstruction.
(4) Perform the DVV analysis on normalized, end matched data.
(5) Generate B surrogates of the loop data.
(6) Perform DVV analysis on all the surrogates.
(7) Compute the average target variance of all the surrogates.
(8) Compute the RMSE for input data and surrogates.
(9) Perform the rank based statistical test.
(10) Loops with RMSE rank $r > (B+1)(1-\alpha)$ are classified as non-linear at confidence level α.
(11) Perform steps 1-10 for all the loops under analysis.
(12) Once all the non-linear loops are identified look for the maximum RMSE value.
(13) The loop with maximum RMSE value is designated as the source of the non-linearity.

6. SIMULATION EXAMPLE

The simulation example used to distinguish non-linearity induced oscillations from the linear one is taken from our recent work [Aftab et al. (2016)]. The block diagram of the closed loop system is shown in Figure 6. Three scenarios, oscillations due to valve stiction(LuGre friction model), external sinusoidal disturbance and external dist added to a loop with a poorly tuned controller, are tested using the proposed method. Noise of variance 0.1 is also added to test the robustness of the proposed method. The time trends for these three test cases are shown in Figure 4.

Table 3. DVV analysis (Simulation example)

Case	rank	type	RMSE
Stiction	100	Non-Linear	0.042
External Dist+Fast Cont	49	Linear	–
Ext Dist	62	Linear	–

The results of these three scenarios are summarized in Table 3. The method is able to classify the test scenarios correctly. The stiction case is correctly identified as non-linear; the fact endorsed both by the rank statistics and

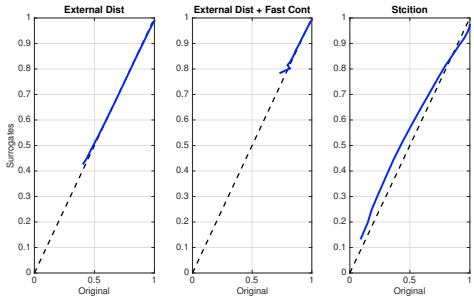

Fig. 6. DVV scatter plots for simulation example

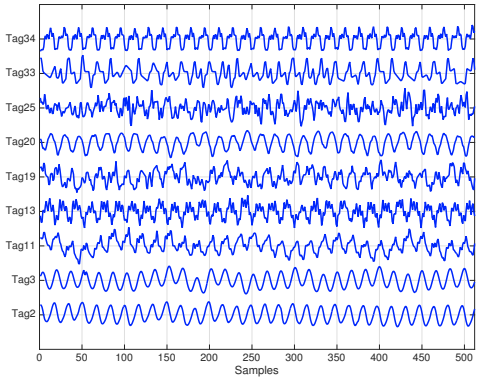

Fig. 7. Time trends of industrial data

the DVV scatter plot. The application of the proposed method to the industrial case study is presented next.

7. INDUSTRIAL CASE STUDY

The case study is taken from Thornhill (2005) and Zang and Howell (2005). The analysis of 37 different tags reveal that a plant wide oscillation with fundamental frequency of $0.06min^{-1}$ is found to be present in 9 Tags. The objective is to analyze these tags and to find the root cause of oscillation. The time trends of these 9 Tags are shown in Figure 7.

All these tags are analyzed using the DVV method proposed in this paper and results are summarized in Table 4. The analysis consists of two parts.

Table 4. DVV analysis (industrial case study)

Tag	rank	type	RMSE
Tag 2	79	Linear	–
Tag 3	38	Linear	–
Tag 11	95	Non-Linear	0.0378
Tag 13	99	Non-Linear	0.0415
Tag 19	94	Linear	0.0318
Tag 20	4	Linear	–
Tag 25	86	Linear	–
Tag 33	92	Non-Linear	0.0299
Tag 34	100	Non-Linear	0.0743

7.1 Detection of Non-Linearity

First of all the loops are analyzed for the presence of non-linearity. The DVV analysis points out that the Tags 11, 13, 19, 33, and 34 exhibit non-linearity induced oscillations; whereas no non-linearity is detected in other tags using the rank based statistics. The findings confirm the already reported findings that the oscillations are due to non-linearity in the control loops. The results are similar to the ones reported by Thornhill (2005) and Zang and Howell (2005) except that Tag19 which is classified here as non-linear.

7.2 Isolating the Source of Non-Linearity

Once it is concluded that the plant wide oscillation with frequency $0.06min^{-1}$ is due to non-linearity, the next step is to find the source of this non-linearity. As discussed in section 2.3 that different components of plant act as mechanical filter and tend to filter the higher order harmonics and the signature of non-linearity loses strength as we move away from the source. This fact dictates that the most non-linear element will be the source of the oscillation.

The RMSE values from the DVV analysis (Table 4) reveal that Tag34 has the largest non-linearity and is thus the root cause of the oscillation. The same finding is reported in other studies conducted by Thornhill (2005) for the same data set.

8. CONCLUSIONS

A novel approach based on delay vector variance (DVV) is presented to detect and isolate non-linearities in control loops. The proposed scheme is simple and effective; and doesn't require tuning of the parameters like prediction horizon, number of nearest neighbors etc as the case with existing methods like the one presented by Thornhill (2005). An automatic method to find the suitable embedding dimensions is also augmented with the DVV analysis to make it more robust than existing tools. Moreover, instead of assuming the Gaussian distribution of the test statistics, a non-parametric rank based statistic is used to test the Null hypothesis. The efficacy of the proposed scheme is established using both simulation and industrial case studies.

REFERENCES

Aftab, M.F., Hovd, M., Huang, N.E., and Sivalingam, S. (2016). An adaptive non-linearity detection algorithm for process control loops. *IFAC-PapersOnLine*, 49(7), 1020 – 1025.

Babji, S., Gorai, P., and Tangirala, A.K. (2009). Detection and quantification of control valve nonlinearities using hilbert–huang transform. *Advances in Adaptive Data Analysis*, 1(03), 425–446.

Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 110(1), 43–50.

Choudhury, A.A.S., Shah, S.L., and Thornhill, N.F. (2008). *Diagnosis of process nonlinearities and valve stiction: data driven approaches*. Springer Science & Business Media.

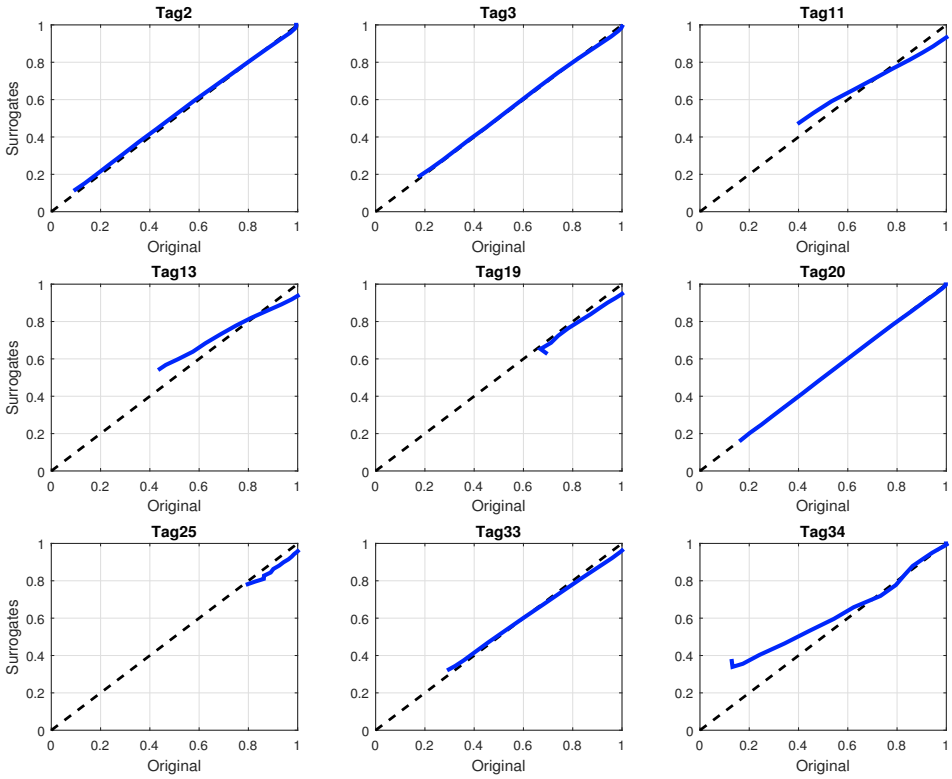

Fig. 8. DVV scatter plot (industrial case study)

di Capaci, R.B. and Scali, C. (2015). Review on valve stiction. part i: From modeling to smart diagnosis. *Processes*, 3(2), 422. URL `http://www.mdpi.com/2227-9717/3/2/422`.

Gautama, T., Mandic, D.P., and Van Hulle, M.M. (2004a). The delay vector variance method for detecting determinism and nonlinearity in time series. *Physica D: Nonlinear Phenomena*, 190(3), 167–176.

Gautama, T., Mandic, D.P., and Van Hulle, M.M. (2004b). A novel method for determining the nature of time series. *IEEE Transactions on Biomedical Engineering*, 51(5), 728–736.

Hägglund, T. (2011). A shape analysis approach for diagnosis of stiction in control valves. *Control Engineering Practice*, 40, 782–789.

Kantz, H. and Schreiber, T. (2004). *Nonlinear time series analysis*, volume 7. Cambridge university press.

Sauer, T., Yorke, J.A., and Casdagli, M. (1991). Embedology. *Journal of statistical Physics*, 65(3-4), 579–616.

Schreiber, T. and Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77(4), 635.

Srinivasan, R., Rangaswamy, R., and Miller, R. (2005). Control loop performance assessment. 1. a qualitative approach for stiction diagnosis. *Industrial Engineering Chemistry Research*, 44, 6708–6718.

Theiler, J. and Prichard, D. (1996). Constrained-realization monte-carlo method for hypothesis testing. *Physica D: Nonlinear Phenomena*, 94(4), 221–235.

Thornhill, N., Shah, S., and Huang, B. (2001). Detection of distributed oscillations and root-cause diagnosis.

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

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

Yamashita, Y. (2005). An automatic method for detection of valve stiction in process control loops. *Control Engineering Practice*, 14, 503–510.

Zang, X. and Howell, J. (2003). Discrimination between bad tuning and non-linearity induced oscillations through bispectral analysis. In *SICE 2003 Annual Conference*, volume 1, 896–900. IEEE.

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