# DETECTION AND DIAGNOSIS OF PLANT-WIDE OSCILLATIONS

BY

**MOHAMMED NADEEMULLAH SHAREEF**

A Thesis Presented to the

DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

## MASTER OF SCIENCE

In

## Systems Engineering

**January 2010**

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN 31261, SAUDI ARABIA

# DEANSHIP OF GRADUATE STUDIES

This thesis, written by MOHAMMED NADEEMULLAH SHAREEF under the direction of his thesis advisor and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN SYSTEMS ENGINEERING

Thesis Committee

Dr. Sami El-Ferik (Advisor)

Dr. Fouad M. Al-Sunni (Member)

Dr. Moustafa Elshafei (Member)

Dr. Fouad M. Al-Sunni
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

17/3/10
Date

# ACKNOWLEDGMENTS

**IN THE NAME OF ALLAH, THE MOST BENEFICIENT, THE MOST MERCIFUL**

All praises are due to Allah (SWT), the Cherisher and Sustainer of the worlds, none is worthy of worship but Him. I am sincerely thankful to Him for His kindest blessings on me and all the members of my family. I ask for His blessings, mercy and forgiveness all the time. May the peace and blessings of Allah be upon his dearest prophet, Muhammad (Peace Be upon Him).

I am grateful to King Fahd University of Petroleum & Minerals, for providing a great environment for research. I am also grateful to the chairman of Systems Engineering department Dr. Fouad Al-Sunni for his kind help and cooperation in every matter.

I would like to express my profound gratitude and appreciation to my advisor Dr. Sami El-Ferik, for his consistent help, guidance and attention that he devoted throughout the course of this work. He is always kind, understanding and sympathetic to me. His valuable suggestions and useful discussions made this work interesting to me. Special and sincere thanks go to my thesis committee members Dr. Fouad Al-Sunni and Dr. Moustafa Elshafei for their interest, cooperation and constructive advice. Special appreciation goes to Prof. L. Ettaleb, for providing the data which was used for this study.

I would like to thank my friends Mujahid, Sabih, Ali, Zahid Ilyas, Aves, Abdul Ghani, Minhaj, Hussain, Ahsan, Abdul Hafeez, Nazeer, Naseer, Muneer, Hyder, Imran and Mudassir for their concern and help.

Last but not least, I humbly offer my sincere thanks to my parents for their incessant inspiration, blessings and prayers. I owe a lot to my brothers and sisters for their unrequited support, encouragement and prayers. For all my studies, credit goes to my eldest brother Inayat who is always with me in my thoughts.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# THESIS ABSTRACT

**NAME:** Mohammed Nadeemullah Shareef

**TITLE:** Detection and Diagnosis of Plant-wide Oscillations

**MAJOR:** Systems Engineering

**DATE:** January 2010

Plant-wide oscillation assessment is one of the key research areas in process control plants. Performance monitoring and control can impact huge economical benefits to the industry as the competition and automation advances more. This work presents a methodology for detection and diagnosis of plant-wide oscillations. The proposed methodology is based on routine operating data and detects the problematic control loops by improved spectral factorization based on Genetic Algorithm (GA) search technique. The advantage of proposed GA based factorization lies in its ability to search the solution space globally. The proposed technique (GA-based factorization), along with existing techniques (Independent Component Analysis and Non-negative Matrix Factorization), are tested on two industrial case studies. Performance comparison of GA based factorization with that of ICA and NMF verifies the improved and efficient results of the proposed method for plant-wide oscillation detection. Higher order statistics (HOS) is used to identify the reasons behind the poor performance of the loops detected as root causes of plant-wide oscillations.

# ABSTRACT (ARABIC)

**الاسم:** محمد نديم الله الشريف

**العنوان:** فحص و تشخيص التذبذب المنتشر خلال المصنع

**التخصص:** هندسة النظم

**التاريخ:** يناير 2010

إن دراسة التذبذب المنتشر في الأنحاء المختلفة للمصنع من المجالات المهمة في أبحاث نظم التحكم الصناعية. ذلك أن مراقبة الأداء و التحكم به يمكن أن يقدم فوائد اقتصادية ضخمة للصناعة خصوصا مع اشتداد التنافس و التقدم المتزايد في نظم الأتمتة. هذا البحث يقدم منهجية لفحص و تشخيص التذبذب المنتشر في أنحاء المصنع. الطريقة المقترحة تنبني على استخدام بيانات التشغيل المعتادة حيث تقوم الطريقة بفحص حلقات التحكم المتسببة في التذبذب مستخدمة طريقة محسنة لتحليل المكونات الطيفية للإشارات مستعينة في ذلك تقنية بحث خوارزمية الجينات Genetic Algorithm (GA). الميزة من طريقة التصنيف المبني على خوارزمية الجينات يكمن في مقدرتها على الاستيعاب الكلي خلال البحث لمجال الحلول الممكنة. الطريقة المقترحة (التصنيف بناءً على خوارزمية الجينات) مع الطرق المطروحة في الابحاث السابقة (مثل طريقة تحليل المكونات المستقلة و طريقة التصنيف للمصفوفات غير السالبة) تم اختبارها على حالتين دراسيتين من الصناعة. أثبتت المقارنة بين أداء طريقة التحليل بناءً على خوارزمية الجينات المقترحة و أداء كل من طريقة تحليل المكونات المستقلة و طريقة التصنيف للمصفوفات غير السالبة كفاءة و أفضلية الطريقة المقترحة لفحص سبب التذبذب المنتشر. عقب ذلك يمكن إستخدام الإحصائيات ذات الرتب الأعلى للتعرف على الأسباب وراء ألاداء المتدني للحلقات التي عرف انها المصدر الاساسي للتذبب في أجزاء المصنع.

# CHAPTER 1

# INTRODUCTION

## 1.1 Overview

Automation in modern process industries is becoming increasingly important to handle the plant's operation efficiently. The main objective behind the automation of a plant is to reduce the manpower thereby reducing the operating and maintenance cost, maintaining good product quality, and ensuring operators safety. A typical process control plant consists of numerous control loops varying from few hundreds to thousands, and a majority of these control loops have the conventional PID controllers. Only around one third of these controllers provide acceptable performance (see for example [1]). The remaining two third of controllers perform below acceptable limits, due to the control valve nonlinearities, oscillatory load disturbances, or improper tuning of the controllers. Recently, a survey conducted by Desborough and Miller [11], showed that only 16 % of the controllers are operating at their best to give the optimal performance and about 36 % of the loops operate in manual mode. The Figure 1.1.1 below clearly shows the performance of controllers according to the survey which shows the necessity of monitoring the controllers' performance in a process plant.

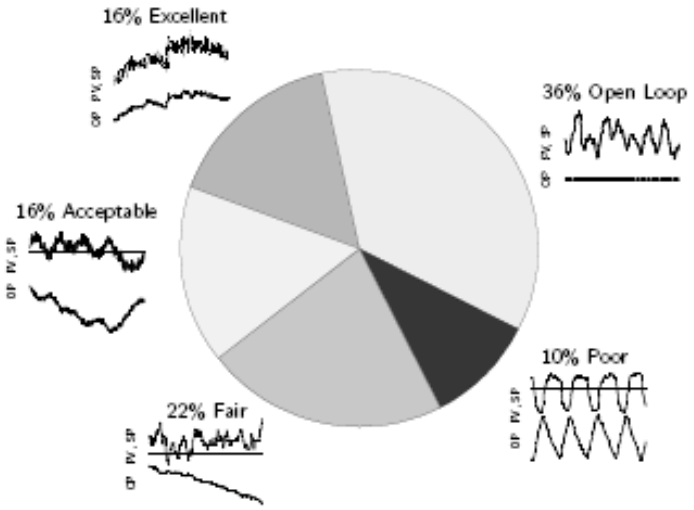



Figure 1.1.1: Controller Performance Demographics

All these problems cause the control loops to oscillate. The presence of these oscillations indicates degradation in the controller performance through increased variability in the process variables. In turn, increase in variability leads to the production of inferior quality products, increased energy consumption and larger rejection rate which ultimately leads to higher production costs and losses of a company. Thus, detection and diagnosis of root cause of such oscillations are crucial for optimal performance of the plant and economic survival in this time of globalization.

Many research studies have been carried out in the area of oscillation detection for the Single Input Single Output (SISO) case with success. However, the detection and diagnosis of plant-wide oscillations, i.e. the multivariable case, is still an open area for research. Indeed, in a process plant, control loops are highly integrated so as to utilize the energy through recycle streams. This causes the presence of disturbance/oscillations in one part of a plant to propagate to other loops through loop interactions thereby causing plant-wide oscillations. The detection and diagnosis of the source loop, which

represents the root cause of plant-wide oscillations is very important. The benefits of proper detection and diagnosis of the cause of oscillations is twofold.

1) It helps in eliminating the problem causing oscillations by suitable well guided maintenance actions.

2) Make the plant operate closer to the operating limits of the product quality.

Thus, more efficient methods in control loop performance monitoring that can identify the root cause of oscillations are needed.

## 1.2 Control Loop Performance Monitoring

The focus on the product quality and variability is increasing in order to remain competitive in the global market. The reason behind the increased focus is due to the recent advancements being made in the control technology as applied to control complex processes. The product variability limits are being reported to be as tight as 1% in some applications such as paper and pulp industries. The control engineers normally focus more on the performance of the plant during the design and implementation stage. But, the performance of the plant while running is not given much attention due to increased number of control loops and the limited manpower. The performance of a plant starts to deteriorate due to the process changes, wearing and deteriorating machinery and process drifts. Maintaining these controllers at near optimal performance can provide significant economic benefits.

The main criterion behind the rise of control loop performance monitoring (CLPM) is to obtain needed information about the plant's performance without disturbing

its normal operation. The normal closed loop operating data provides both the information about any degradation in the plant's performance with time and its relation with the operating conditions. This helps in improving the normal operating conditions leading to achieve the ultimate objective of running the plant close to its operating limits.

## 1.3 Problem Statement

The presence of plant-wide oscillations leads to reduced product quality, increased wear and tear of equipment, and high losses which ultimately deteriorate the performance of the whole plant. In plant-wide control, the presence of oscillations in one control loop gets transferred to the other loops due to interactions between the loops. The common causes of plant-wide oscillations could be improper controller tuning, nonlinearities in loops and external disturbances. The aim of this work is to detect those particular loops (also known as the root causes) which cause the whole plant to oscillate using only routine operating data i.e. *Set Point* (SP), *Process Variable* (PV) and *Controller Output* (OP). A spectral decomposition of the given time trends is done to detect the root cause. In addition to that, diagnosis of the root cause is carried out to find the reason behind its ill performance. Based on the results of diagnosis, appropriate actions could be taken to improve the plant's performance.

## 1.4 Thesis Objective

The objective of this thesis is to develop a non-invasive framework for the detection and diagnosis of plant-wide oscillations using routine operating data of a

process plant. During this work, the above mentioned objective is achieved using the following steps:

- Extraction of useful information from routine operating data using signal processing methods and statistical techniques.
- Study and implementation of existing methods in the literature for finding the root cause of plant-wide oscillations.
- Propose a new approach based on a meta-heuristic technique, namely Evolutionary Algorithm, to find the root cause of plant-wide oscillations.
- Study and analysis of an existing diagnostic method to find the reason behind the ill performance of the root cause.

## 1.5 Thesis Organization

The outline of this thesis is as follows. Chapter 1 provides an overview of the main areas of focus by outlining the objectives. Chapter 2 provides a detailed literature survey of the work done in the area of detection and diagnosis of plant-wide oscillations. In Chapter 3, study of the existing methods for the root cause detection of plant-wide oscillations and their implementation will be presented with case studies. In Chapter 4, the proposed method based on a meta-heuristic solution technique is presented. The solution technique used in this work is Evolutionary Algorithm for the root cause detection of plant-wide oscillations. Chapter 5 presents a method for the diagnosis of plant-wide oscillations so as to gain information about the particular problem in the loop

detected as the root cause. Lastly, in Chapter 6, conclusion and recommendations for future work are provided.

# CHAPTER 2

# LITERATURE SURVEY

The research in the area of Control Loop Performance Monitoring (CLPM) started about two decades ago and since then, this area gained a widespread interest from both the academic community and the industry. This chapter presents a detailed literature survey about the work published in the area of control performance monitoring with a particular interest in the field of detection and diagnosis of plant-wide oscillations.

The first research in the direction of control performance monitoring was carried out by Harris in the year 1989, who developed a remarkable Minimum Variance Controller (MVC), to evaluate control loop performance from routine operating data [19]. It has been found that, theoretically best achievable feedback control performance can be obtained by measuring the output mean square error. This mean square error can be taken as a benchmark. The current control performance can be compared with this benchmark so as to improve the controller's performance.

The research was further carried out by many authors (for instance, see, [20], [35], and [46]), who made extensions to the control performance assessment based on the Minimum Variance Controller. Huang and Shah [25] worked upon the estimation of interactor matrices consisting of process time delays, in order to determine

the performance in multivariable case. He developed an efficient algorithm known as FCOR (Filtering and Correlation Analysis) for the assessment of controller performance.

A review of the recent advancements made in the field of control performance monitoring is authored by Jelali [30]. In [30], Jelali made an extensive survey about all the developments made over the years since the development of MVC, and also about the software packages available in the market for controller performance monitoring. In [1], [11] and [12], authors have made an extensive survey, so as to gain information about the percentage of controllers that perform optimally. It was found that only one third of the controllers in a plant operate at an acceptable level and the remaining two-thirds of them are either controlled manually or suffer from various problems. There are various reasons due to which the performance of the controller degrades. Among these reasons, one can find improper controller tuning, control valve stiction and external oscillatory disturbances coming either from upstream or downstream process loops. All of the mentioned reasons for the controller performance degradation results in oscillations in the time dynamics of process variables. Desborough and Miller, in [11], states that with the improvement in controller performance, nearly 300 Million Dollars could be saved annually. Model Predictive Control is one of the latest control technologies on rise in recent times. Such technologies are applicable on the supervisory level in a process plant which in-turn is dependent on the conventional PID controllers at the lower levels.

## 2.1 Review for the detection of plant-wide oscillations

The research carried out for the assessment of controller performance involve the use of only routine operating data i.e. set point (SP), process variable (PV) and controller output (OP). The first work in the direction of oscillation detection was performed by Hägglund [18]. He developed a method to detect the oscillations caused by high static friction in control valves. He suggested the calculation of integral of absolute error (*IAE*) between successive zero crossings of the control error (i.e. SP - PV). When the oscillations are present, the *IAE* is large. If the value of *IAE* exceeds the limiting value $IAE_{\text{lim}}$, then it is concluded that the oscillations are present. The method is able to detect oscillations online without disturbing the routine operation.

Thornhill and Hägglund [56] inspected the regularity of the zero crossings of a time trend. Oscillations are declared present when this regularity was high. An estimate of signal to noise ratio of the oscillation was also given. Forsman and Statin [16] combined both the integral of absolute error and the zero crossing for the detection of oscillations. Their method was able to find asymmetrical oscillations based on the positive and negative deviations in the signal.

Miao and Seborg [41] proposed an offline approach for the detection of oscillations. Their method is based on the calculation of sample autocorrelation function of either the controlled variable or the control error signal. The method performs well even in presence of noise which is an advantage over earlier developed methods. They also proposed a measure which could differentiate a decaying oscillation from sustained oscillation. Their method resulted in a U.S. patent (Miao and Seborg, 1998).

The research mentioned above addresses only single loop cases. The first move in the direction of multi-loop case was made by Ettaleb *et al.* [14] and then Ettaleb [13] who proposed an approach in order to localize the oscillations in a multiloop system. The method assumed the knowledge of different frequencies of the periodic component of the output signal. An offline method based on Extended Least Squares utilizing parameter estimation for the calculation of time delay has been proposed in [13]. Also, a performance index based on the time delays between different input-output pairs has been developed which doesn't require the knowledge of interactor matrices. This performance index defines an absolute lower bound on the achievable output variance for each output which when added altogether gives the overall control loop performance. Although, the research is based on multi-loop case, loops are still treated on a loop by loop basis which is equivalent to SISO work.

The knowledge of process indicates that interaction within loops is the most complex phenomenon. Control loops in a process plant are interconnected so as to utilize the energy optimally which obviously saves millions of dollars. One major drawback of these interactions among the control loops is that, when one single loop suffers from a problem all other control loops connected to this loop also suffers from the same problem. As has been discussed earlier the problems in a control loop result in oscillations in the time trends which gets transferred to other loops due to control loop interactions. Although, there exists many *decoupling* methodologies which negates the interactions within loops. However, most of the time the interactions are necessary to optimize the plants performance. This shows that the solution to tackle the problem of

oscillations is not to use decoupling concept, rather there should be a methodology which could detect readily the problems within the loops. The focus should be on detection of the all the loops in a plant which suffer from various problems so as to resolve the persisting problems and prevent the oscillations to get transferred to other interacting loops. Therefore, research efforts are focusing on finding out the source of oscillations with only routine operating data. Thornhill and Horch in [57], presented an excellent review of all the advances made in plant-wide oscillation detection using only routine operating data. The research done until now can be mainly classified according to the tree diagram shown in Figure 2.1 [54].

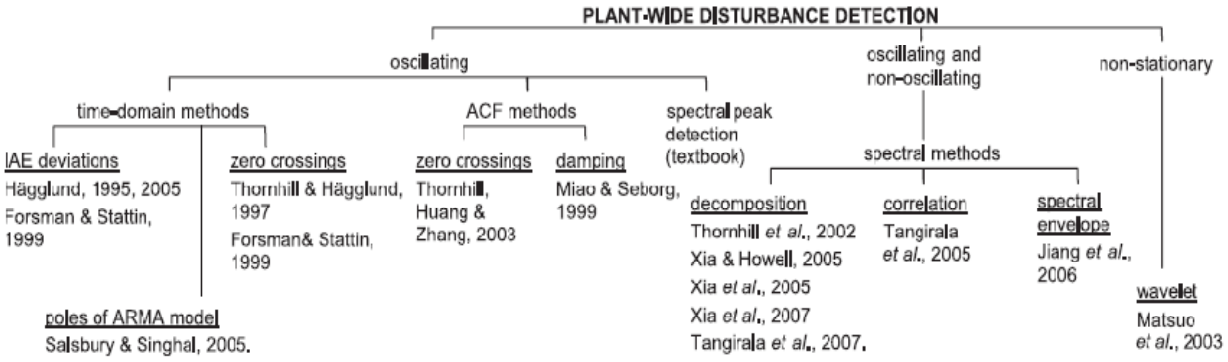



Figure 2.1: Tree diagram showing the data driven methods for finding plant-wide oscillations

The main classification of disturbance detection is based on oscillating time domain methods, oscillating and non-oscillating spectral methods and a non-stationary wavelet based method. The time domain methods are mostly based on SISO. The autocovariance based method by Thornhill, Huang and Zhang (2003) considers the multivariate nature of the process plant. The time domain methods and the autocovariance based methods suffer

from the time delays and lags in the measurements. Hence, power spectra with its property of invariance to time delays and lags, is widely used for the analysis of plant-wide oscillations. Methods such as spectral decomposition, spectral correlation and spectral envelope are based on the power spectra of measurement signals. Matsuo *et al.* [38] proposed a method based on wavelet analysis for the detection of plant-wide oscillations. They diagnosed the reason of poor performance by analyzing valve input output characteristics.

The work done by Thornhill and Hägglund [56], Forsman and Stattin [16] and Miao and Seborg [41] could detect the presence of oscillations on a loop-by-loop basis which is very time consuming. Thornhill *et al.* [59] used a well known statistical technique known as Principal Component Analysis (PCA) for the detection of plant-wide oscillations. PCA is the most common method for dimensionality reduction of high order matrices into two or more factors depending upon the application such as microarray data analysis, computer networks, etc. In plant-wide oscillation detection, the method was applied on the decomposition of power spectra matrix and hence named as Spectral Principal Component Analysis (SPCA). Full PCA decomposition reconstructs the power spectra matrix over orthonormal basis functions which are spectrum like functions arranged in the descending order capturing spectral variation. The main idea used in PCA is based on the fact that when a number of variables have same spectral features, then a single variable (principal component) can be used to describe it. There could be more than one principal component based on the application. The remaining variation which couldn't be captured by principal components is stored in an error matrix. The authors also performed PCA on the autocovariance function and found the results to be same as

spectral PCA. Inspite of being the most common method for dimensionality reduction, it suffers from the drawback of visualization in space, when number of principal components is more than three. Even with this drawback, PCA is the most popular technique for dimensionality reduction so as to gain information about the number of frequencies leading to plant-wide oscillations. The other drawback of PCA is that it provides the basis shapes which may have both positive and negative orientations i.e. there is no control over the sign of the weights of these basis vectors. More details on SPCA can be found in chapter 3.

Thornhill, Huang and Zhang [58] proposed a method based on the calculation of zero crossings of filtered Auto-Covariance Function (ACF) which could detect multiple plant-wide oscillations. Calculation of period of oscillations, regularity of oscillations and power of oscillations forms the basis for analysis of multiple loops. The benefit of using ACF is the reduction of spurious zero crossings caused by noise as the ACF is not affected by the presence of white noise. The presence of multiple oscillatory frequencies in a single measurement could also be detected with this method. The ACF method has a limitation in its application when it comes to the presence of colored noise and also for the detection of low frequency oscillations as it requires large amount of data. Thornhill, Cox and Paulonis [55] suggested the use of information about the structure of the process plant in addition to detection technique based on non-linearity index for improved analysis. The plant structure information considered is the connectivity and direction of flow of materials and heat to include process understanding.

Xia and Howell [61] proposed a procedure which could monitor the status of a loop and conclude the presence of abnormalities such as oscillations and drifts in multi-

loop configuration. They developed some statistical measures such as signal to noise ratio of the time series to calculate the proposed index known as Overall Loop Performance Index (OLPI) that could localize the type of fault.

Later on, Xia and Howell [62] used a technique known as Independent Component Analysis (ICA) to find out the root cause of plant-wide oscillations. ICA is the statistical and computational technique for revealing hidden factors that underlie a set of random variables. It works on the routine operating data and isolates the independent measurements which are the mixtures of the unknown sources. It works on the components or loops which are both statistically independent and non-Gaussian. The technique is applied on power spectral matrix which is non-Gaussian and the statistical measures such as kurtosis or negentropy are used for the isolation of oscillating sources. More details regarding ICA can be found in chapter 3. Xia *et al.* [63] also proposed the use of ICA to find out the high frequencies of small amplitude efficiently and named the technique as multi-resolution ICA. In this technique different frequencies are isolated by means of dividing the power spectra in different frequency ranges and then performing spectral ICA. Though the ICA technique works well in the detection of root cause analysis but it has one major assumption of statistical independence of sources which is not always the case.

Tangirala *et al.* [52] proposed the use of Non-negative Matrix Factorization (NMF) for the detection of root cause of plant-wide oscillations. NMF provides parts based representation that retains the causal basis spectral shapes. They proposed a new measure known as Pseudo Singular Value (PSV) which could determine the number of dominant frequency components. They also proposed total power plot and its

decomposition by NMF which could give correct information about the number of oscillatory frequencies. The method could also quantify the amount of oscillation present. In detail explanation of NMF method is presented in chapter 3. Xia *et al.*, [64], also worked upon using NMF but they used the concept of sparseness constraints using which they could control the amount of sparseness of basis spectral shapes. Specific level of sparseness could be assigned according to the application. An issue with the NMF method is that it provides local minima as only one of the decomposed matrices can be convex and not both of them simultaneously.

Tangirala *et al.* [53] proposed a new visualization tool known as PSCMAP which analyzes the variables with similar power spectral shapes. It is based on the proposed measure Power Spectral Correlation Index (PSCI). PSCI calculates the correlation between spectra of two signals as shown in Eq. (2.1).

$$\text{PSCI} = \text{Correlation}\left(\left|X_i(\omega)\right|^2, \left|X_j(\omega)\right|^2\right) = \frac{\sum_{\omega_k}\left|X_i(\omega_k)\right|^2\left|X_j(\omega_k)\right|^2}{\sqrt{\sum_{\omega_k}\left|X_i(\omega_k)\right|^4 \sum_{\omega_k}\left|X_j(\omega_k)\right|^4}} \quad \ldots (2.1)$$

where $\left|X_i(\omega_k)\right|^2$ and $\left|X_j(\omega_k)\right|^2$ represent power spectra's of measurements $i$ and $j$. The correlation measure PSCI is always bounded between 0 and 1. The color map (PSCMAP) is a useful visualization tool which could differentiate loops affected by different frequencies in a multivariate system. The effectiveness of this method decreases with the increase in the number of dominant frequencies and with the size of the plant.

Jiang *et al.* [32] proposed Spectral Envelope method for the detection of plant-wide oscillations. A statistical hypothesis test is used for categorizing the loops

oscillating at same frequencies. A new measure, Oscillation Contribution Index (OCI) in Eq. (2.2) is proposed for the detection of potential root cause candidate leading to whole plant oscillations.

$$\mathrm{OCI}_j(\omega) = \frac{\hat{\beta}_{1,j}(\omega)}{2\sigma_{\hat{\beta}}(\omega)} \qquad \ldots (2.2)$$

where $\hat{\beta}_{1,j}(\omega)$ is $j^{\text{th}}$ element of optimal scaling vector and $\sigma_{\hat{\beta}}(\omega)$ is the standard deviation of the optimal scaling of all the identified variables that have oscillations. The loops with $\mathrm{OCI}(\omega)>1$ are the most likely root cause candidates at frequency $\omega$ due to their strong contribution to spectral envelope.

Matsuo *et al.* in [38] and later in [39] presented the use of Wavelets to locate the root cause of plant-wide oscillations. The method could also detect the presence of multiple frequencies in an oscillatory loop based on Thornhill *et al.* [58]. The diagnosis was carried out by pattern matching the time resolved frequency spectrum using Wavelets along with input output response of each valve i.e. (*MV-PV* characteristics). The drawback of this method is the use of manipulated variable (*MV*) i.e. the output of the valve which is not usually available in routine operating data.

## 2.2 Review for the diagnosis of plant-wide oscillations

Diagnosis of the root cause of plant-wide oscillations is also equally important as that of detection of plant-wide oscillations. This is due to the fact that, a suitable action could only be taken based on the knowledge of problem present in a particular control

loop. Taha *et al.* [51] proposed a friction index which could measure the valve's deterioration in normal operating conditions by comparing with valve's static characteristics given by manufacturer. This index was developed based on the distance between the curvatures of valve characteristics. Their method could differentiate oscillations due to friction in valves, improper controller tuning and external perturbation. If the proposed friction index had a high value, presence of friction in valves was concluded. For improper controller tuning, the closed loop poles of the process were analyzed. If the poles were close to zero, improper controller tuning was concluded as the possible cause and when the closed loop poles were far away from zero then the possible case of external perturbation was concluded.

Horch in [23] proposed a method which could automatically detect stiction in oscillating control loops based on cross-correlation between control input and process output. The method could also differentiate the presence of oscillations due to stiction from that of oscillations due to other external causes. If the result of cross-correlation between control input and process output is an odd function then the cause for the presence of oscillations would be stiction and conversely the presence of oscillations due to improper tuning or external disturbances results in an even function.

Choudhury *et al.* [6] used Higher Order Statistics (HOS) for the diagnosis of nonlinearities present in a plant. They proposed two new indices Non-Gaussianity Index (NGI) and Nonlinearity Index (NLI) based on the estimation of squared bispectrum and squared bicoherence. The method uses routine operating data and could clearly identify the presence of nonlinearities associated with a loop. The method basically focused on the detection of stiction in control valves based on NGI & NLI indices and analysis of

*PV-OP* graphical plots. If the values for NGI and NLI are greater than their critical values proposed, then non-linearity is confirmed. In addition to this, if the *PV-OP* plot results in an elliptical pattern, stiction is confirmed in that particular loop.

Choudhury *et al.* [4] proposed the idea of HOS for confirming stiction in control valve. They also proposed an idea based on change in controller gain so as to categorize an internally generated oscillation from an external oscillatory disturbance. The detection of oscillations was done using autocorrelation function based method. The proposed method was not based on the analysis using only routine operating data as it required plant model details, therefore the method couldn't be used when working with only routine operating data.

Zang and Howell [66] proposed a method based on Higher Order Statistics (HOS) used by Choudhury *et al.* [4] for isolation of root cause of plant-wide oscillations. They proposed a measure known as Bi-amplitude ratio which is the ratio between the bispectra between the first and the third harmonic. The number of oscillatory loops is assumed to be predetermined in this method and the other assumption made is the inherent low pass character of the plant which is not always the case. One more limitation of the bispectrum analysis is its inability to monitor the limit cycle oscillations due to symmetrical waveforms (for example square waves and triangular waves) because the bispectrum of a symmetrical waveform is zero.

# CHAPTER 3

# DETECTION OF PLANT-WIDE OSCILLATIONS

## 3.1 Introduction

Oscillations are common in almost all of the processes in an industry. Oscillations in the time trends of a measurement indicate the variability of the quantity being measured. Excessive variability takes the form of oscillations which in turn represents the deterioration in the performance of a loop. Nowadays, all of the process plants are based on the requirement of high heat integration for saving energy to a greater extent. This is the main reason for a problem occurring in one loop getting transferred to other connected loops thereby leading to plant-wide oscillations. Oscillations in a process results in the production of inferior quality products, excessive energy consumption, larger rejection rates which ultimately leads to loss of a company. Therefore, it becomes essential to resolve the problem of oscillations in a plant. The use of only routine operating data consisting of Set Point (*SP*), Process Variable (*PV*) and Controller Output (*OP*) is on a rise to tackle the problem of plant-wide oscillations.

In this chapter, we will discuss in detail four of the methods developed earlier to deal with the problem of plant-wide oscillations. The focus will be on the methods based on spectral decomposition principle due to the remarkable property of invariance of

power spectra to phase lags and time delays in a process. Also, a non-data based method, based on process understanding is presented. In Section 3.2, the phenomenon of oscillations is presented along with the probable causes that lead to oscillations. In Sections 3.3, 3.4 & 3.5, the spectral decomposition based methods Spectral Principal Component Analysis (SPCA), Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF) are presented respectively with case studies. Section 3.6 explains a non-data based approach based on control loop digraphs (*di*rectional-*graphs*) to detect the presence of plant-wide oscillations.

## 3.2 Oscillations and their causes

### 3.2.1 Oscillations

If a signal is periodic with well defined amplitude and frequency, it is called an oscillatory signal [5]. Sine wave, saw-tooth wave and rectangular waves are few examples of oscillatory signals. A signal $x(t) = A\sin(\omega t)$ is an oscillatory signal with amplitude $A$ and frequency $\omega$. Figure 3.1 shows time trends of some oscillatory signals. The visualization of time trends of an oscillatory signal doesn't give an idea of the frequency at which a particular response is oscillating. The determination of the oscillating frequency is a very important part in the oscillation detection and diagnosis. In order to determine the oscillating frequency, a very simple frequency domain tool i.e. the power spectrum is used. The power spectrum reveals the oscillatory frequencies in the signals as shown in Figure 3.2. The advantage of using power spectra for oscillation detection lies in its ability of invariance to phase lags and time delays.

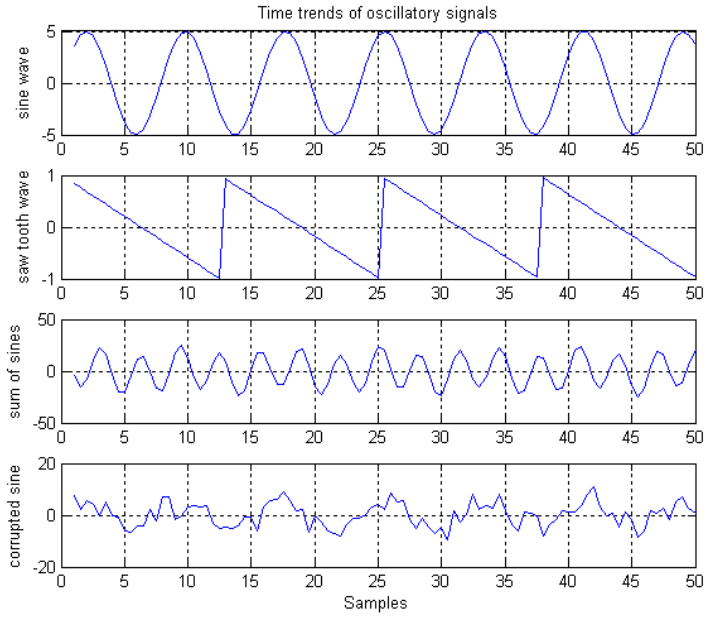



Figure 3.1: Plot showing various oscillatory signals

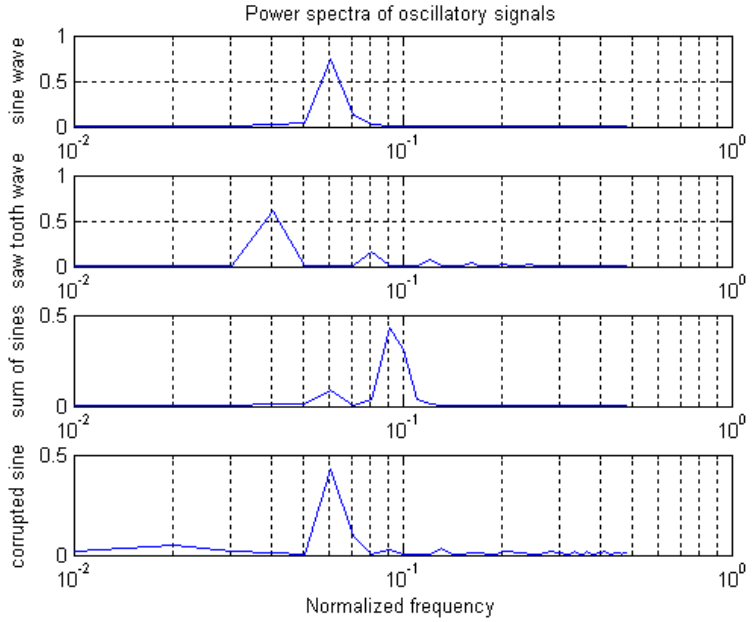



Figure 3.2: Power spectra used as a tool for detection of oscillatory frequencies

### 3.2.2 Plant-wide Oscillations

A lot of research has been done in the SISO case but complexities arise in the case of MIMO. As described earlier, in a plant there are interactions in the loops due to the recycle streams present in order to effectively utilize the energy. In process control, the variability is to be transferred from key process variables to other loops which can accommodate the variability. The reason for this being reduced inventory and high heat integration in modern plants. Thus if an oscillation occurs in any part of a plant, there is a possibility of it getting propagated to other loops due to the presence of interactions within loops thereby leading to oscillations throughout the plant.

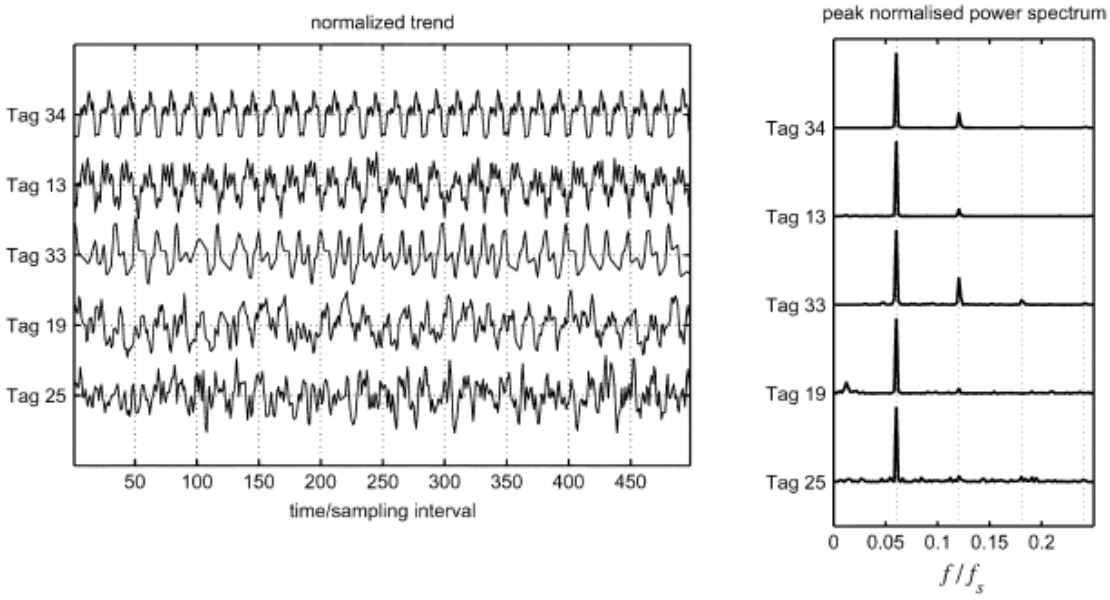



Figure 3.3: Time trends & Power Spectra of a plant suffering from plant-wide oscillations

In order to detect the plant-wide oscillations, the presence of every oscillatory frequency is to be explored in all of the loops to find the root cause. The mere visualization of power spectra of all loops doesn't give an idea about which loop is the

root cause or the primary candidate which is causing all the other loops to oscillate at the same frequency. An example of a plant-wide oscillation case from [54] is as shown in Figure 3.3 which shows the time trends and power spectra of five loops oscillating at same frequency in a plant. Thus, we get only a general idea about the frequencies at which the plant is oscillating from the power spectra plot of all the loops.

### 3.2.3 Causes of Plant-wide Oscillations

Faults in a process, either generated internally or as a result of external perturbation leads to oscillations in the measurement signals. It is important to note that fault detection and performance monitoring areas are different but are interlinked. The degradation in the performance of a plant may be due to a fault occurring in any of the physical elements in a plant. Figure 3.4 shows the block diagram of a typical process control loop wherein the signals *r(t)*, *u(t)* and *y(t)* represent set point, controller output and process variable respectively. The signals *d(t)* and *n(t)* represent the external disturbance signal and sensor noise respectively.

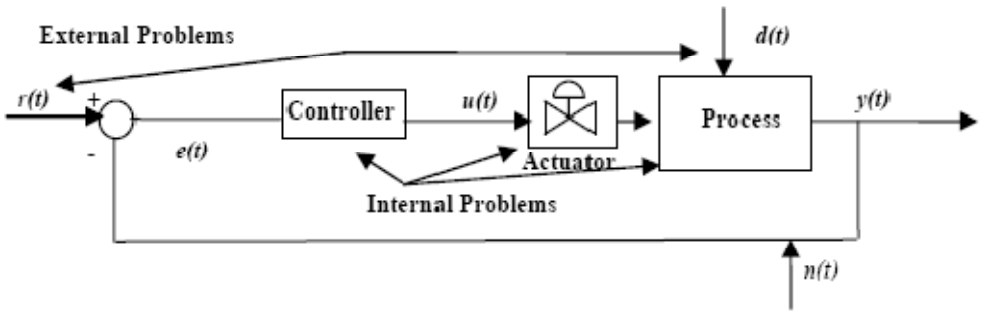



Figure 3.4: Block diagram of a control system with internal and external faults

The causes leading to plant-wide oscillations can be categorized under internal and external faults:

**Internal faults**

- Excessive corrosion and wear and tear of body parts of a control valve which leads to rough behavior of a valve known as *Stiction.*
- Inappropriate valve size that may cause either saturation or sustained oscillations
- High controller gain which leads to sustained oscillations

**External faults**

- Oscillatory disturbances coming from either upstream or downstream processes
- Inherent problems in master controller getting transferred to the slave controller incase of cascade loops
- Environmental noise or the noise generated by sensors

Some of the key requirements for the detection and isolation of plant-wide oscillations as described in [45] are:

- Detection of the presence of one or more periodic oscillations
- Detection of non periodic disturbances and plant upsets
- Determination of the location of the various oscillations/disturbances in the plant and their most likely root causes.

In modern process control plants, the routine operating data i.e. *SP, PV* and *OP* can be obtained from a Distributed Control System (DCS) very easily. Signal processing on the routine operating data is also being utilized currently as an efficient tool for the diagnosis of the root cause. Thornhill & Horch [57] presented a good review of the advances made in the direction of plant-wide oscillation detection and diagnosis. Multivariate spectral analysis is used more widely as compared to analysis in time domain as the power spectra are invariant to time delays and phase shifts caused by the process dynamics. The use of signal processing in combination with statistical techniques is becoming very popular for detection of root cause of plant-wide oscillations.

## 3.3 Spectral Principal Component Analysis (SPCA) [59]

Principal Component Analysis (PCA) is widely used for real time multivariate statistical process control. PCA is basically derived from the singular value decomposition of the data matrix. Details of PCA can be found in Chatfield and Collins [2] and Wold *et al.* [60]. The use of PCA in the area of plant-wide disturbance detection was done by Thornhill *et al.* [59]. As discussed earlier, the disturbance in control loops will cause excessive variations in their response which are represented as oscillations. In plant-wide performance analysis, it is very important to find all the measurements or control loops suffering from a disturbance, as the root cause will be among that group. Therefore, the main aim here is to find out all the control loops oscillating at the same frequency. Thornhill *et al.* [59] performed PCA on the power spectral matrix of measurements and thereby named their method as Spectral Principal Component Analysis (SPCA).

The data matrix $X$ consists of the measurements from $m$ sources, wherein each measurement consists of $N$ samples ( $N > m$ ). The rows of the data matrix $X$ consists of $N$ samples from a measurement and columns represent individual measurements as shown in Eq. (3.1).

$$X = \begin{matrix} \\ \begin{pmatrix} x_1(t_1) & \cdots & x_1(t_N) \\ \vdots & \vdots & \vdots \\ x_m(t_1) & \cdots & x_m(t_N) \end{pmatrix} \end{matrix} \begin{matrix} N \text{ time samples} \rightarrow \\ m \text{ measurements} \downarrow \end{matrix} \tag{3.1}$$

The rows of the data matrix $X$ are mean centered and scaled to unit standard deviation. In the detection of plant-wide oscillations, control error signal (i.e. $SP - PV$) is used for the analysis. This is done just to prevent the false detection of oscillations for the loops whose $SP$ and $PV$ are oscillating in phase. The power spectra of each measurement in the data matrix are obtained and stored in $X$ as shown in Eq. (3.2). SPCA is performed on the power spectra matrix $X$ in Eq. (3.2) wherein, the rows of the power spectra matrix, $X$, are the single sided power spectra $P(f)$ consisting of $N$ frequency channels up to the Nyquist frequency.

$$X = \begin{matrix} \\ \begin{pmatrix} P_1(f_0) & \cdots & P_1(f_N) \\ \vdots & \vdots & \vdots \\ P_m(f_0) & \cdots & P_m(f_N) \end{pmatrix} \end{matrix} \begin{matrix} N \text{ frequency channels} \rightarrow \\ \text{spectra of} \\ m \text{ measurements} \downarrow \end{matrix} \tag{3.2}$$

Decomposition of the spectral data matrix, $X$, by PCA can be represented as:

$$X = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} w'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} w'_2 + \dots + \begin{pmatrix} t_{1,m} \\ \dots \\ t_{m,m} \end{pmatrix} w'_m \quad (3.3)$$

Full PCA decomposition reconstructs the data matrix, *X*, as a sum of weighted basis vectors $w'_1$ to $w'_m$, which are spectrum like basis functions spread over *N* frequency channels, and t-vectors are the corresponding weights of the basis functions. In short, PCA decomposition can be represented as $X = TW^T$. This PCA decomposition has been developed from singular value decomposition $X = UDV^T$ such that $T = UD$ and $W^T = V^T$, where matrix *D* is the diagonal matrix consisting of the eigen values of the product $X'X$.

The main idea behind the use of SPCA decomposition is, if all the measurements have similar spectral features then, one single *principal component* (PC) can describe all of the variability present in the spectra. This decomposition of spectral data matrix is truncated after obtaining specific number of principal components which capture majority of the spectral variation. The decision to truncate the SPCA decomposition is based on eigen value percentage captured by next principal component. If the eigen value percentage of next principal component is less than 5 % of sum of all eigen values, then the SPCA decomposition is terminated. A simple 3 PC model can be represented as:

$$X = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{N,1} \end{pmatrix} w'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{N,2} \end{pmatrix} w'_2 + \begin{pmatrix} t_{1,3} \\ \dots \\ t_{N,3} \end{pmatrix} w'_3 + E \quad (3.4)$$

As described earlier, $w_i'$ vectors are the orthonormal basis functions, which are also known as *loadings*. These functions are arranged in a descending trend according to the eigen values of the matrix $X'X$ . The t – vectors indicate the amplitude of the basis functions in reconstruction of power spectral data matrix and their elements are known as *scores*. The variance of the decomposition which could not be captured by three principal components is stored in an error matrix, *E*, as shown in Eq. (3.4). The visualization of scores in a space gives an idea about the number of principal components needed to reconstruct the data matrix. Mostly, two or three principal components capture maximum variability within spectra, but, if *n* principal components are needed for reconstruction, where $n > 3$, then it becomes difficult to visualize scores in an *n* dimensional space.

### 3.3.1 Industrial Case Study 1 – *Entech Challenge Problem*

The challenge problem is of a paper manufacturing process from Entech Control Inc. The plant consists of paper pulp process wherein the hardwood and softwood pulps are mixed to obtain desired composition. The data consists of 1934 samples from 12 process measurements (tags). The schematic of the plant is shown in Figure 3.5. The time trends used here are control error signals (*SP-PV*) associated with each loop. Some detailed features about the process are given in Table 3.1.

| Tag | Special Features |
|---|---|
| 3 | Slave in cascade from tag 7; in ratio with tag 6 |
| 6 | Slave in cascade from tag 7; in ratio with tag 3 |
| 7 | Master in cascade onto the ratio controller tag 3/tag 6 |
| 8 | Influenced by tags 6 & 9 |
| 9 | Slave in cascade from tag 10 |
| 12 | Has feed-forward from tag 11 |

Table 3.1: Features of the control system from Entech Challenge Problem

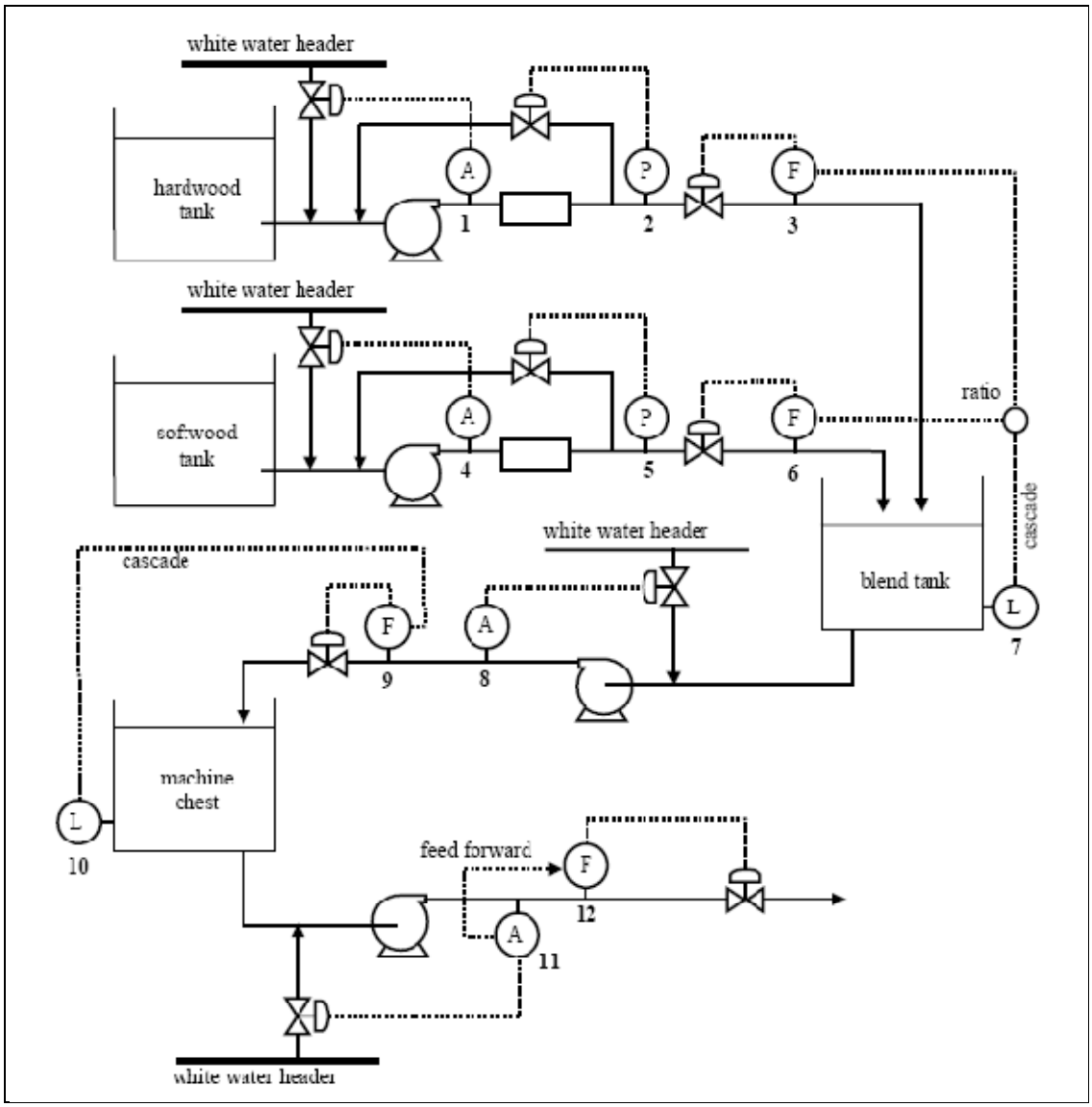



Figure 3.5: Schematic of Entech Challenge Problem

In order to locate the faulty loops in the process, firstly we need to find the number of oscillatory frequencies leading to plant-wide oscillations, and then shortlist the loops affected by individual frequencies. The spectral PCA is the most widely used technique for the detection of number of frequencies affecting a plant. To have a general idea about the oscillatory frequencies, the most common frequency domain tool is the power spectra of the control error signals. Figure 3.6 shows the time trends and power spectra of the 12 control error signals.

Some of the important points to be noted about data preprocessing are [58]:

(i) Mean centre the time trend and remove linear trends before calculation of power spectra.

(ii) Filter the spectra if required.

(iii) Scale the spectra to the same total power such that

$$\sum_{i=1}^{N} P(f_i) = \text{constant} \tag{3.5}$$

where $P(f_i)$ is the spectral power in the $i^{th}$ channel.

From Figure 3.6, the presence of two dominant frequencies is clear which causes the whole plant to oscillate. A low oscillatory frequency of 0.002 $min^{-1}$ (with a period of 500 min) is dominant in loops 3 – 10 & 12 whereas the faster oscillatory frequency of 0.02 $min^{-1}$ (with a period of 50 min) is dominant in the loops 1 & 11. The power spectrum for tag 2 doesn't contain any clear peak as it consists of noise over all of its frequencies.

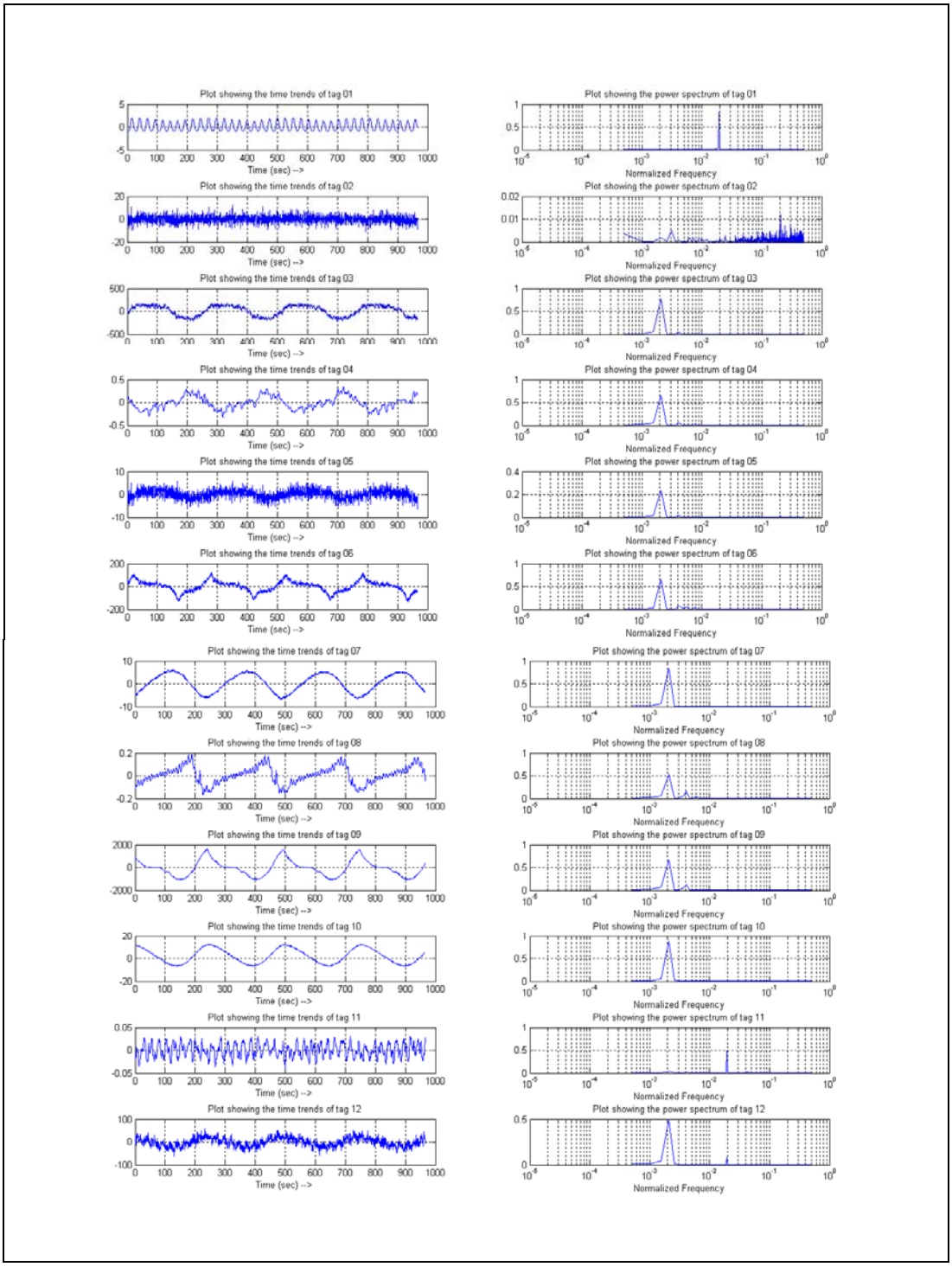



Figure 3.6: Time trends and power spectra of Entech Control Inc. data

The implementation of Spectral Principal Component Analysis on Entech data set resulted in 2 principal components. Table 3.2 shows the percentage of variance captured by each principal component to reconstruct the power spectrum data matrix ($X$). The total variance captured by these two principal components is approximately 99 %.

| Principal Component | Eigen Value of COV($X$) | % Variance Captured |
|---|---|---|
| 1 | 0.369 | 80.40 |
| 2 | 0.0858 | 18.69 |

Table 3.2: Percentage variance captured by individual principal components for Entech data set

The scores or weightings of the two principal components (PC's) are plotted against each other and such a plot is known as Score's plot as shown in Figure 3.7. The plot shows the clustering of scores over a 2 dimensional plot which clearly says that total number of PC's is two. All of the tags were affected by either of the two frequencies and as such the scores in the scores plot lies in two directions. The tags affected by the frequency 0.002 $min^{-1}$ lies along the horizontal axis i.e. tags 3 – 10 & 12 (represented by diamond symbols), whereas the tags affected by 0.02 $min^{-1}$ lies along the vertical axis (as represented by circles). Tag 2 consists of only noise and as such the scores for tag 2 (represented by square) are located at the origin in the Scores plot. Tag 5 has a small content of 0.002 $min^{-1}$ frequency, but was also noisy and due to this, the scores for tag 5 (represented by triangle symbol) lie along the horizontal axis but are also closer to origin.

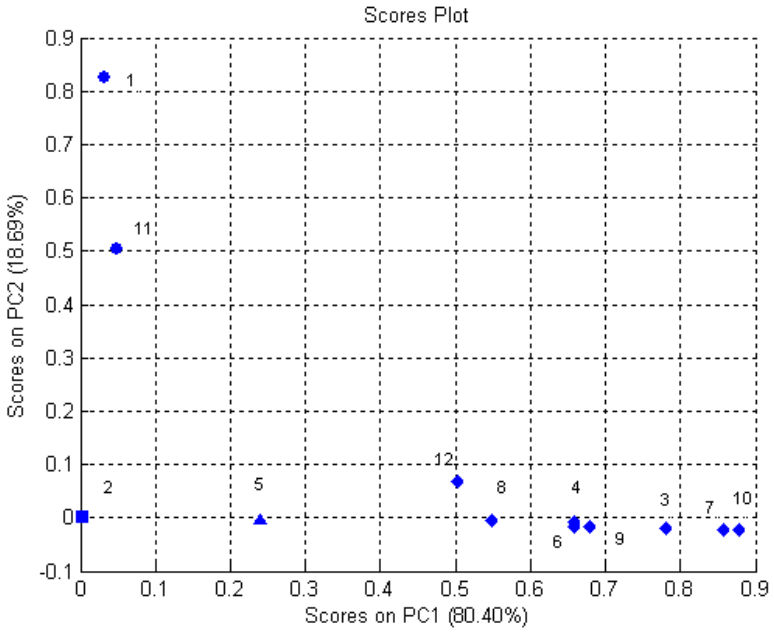



Figure 3.7: Scores plot of PC 1 and PC 2

The main aim of finding the oscillatory frequencies can be achieved by visualizing the loadings or the basis vectors which are spectrum like basis functions. The loadings for the Entech Challenge problem case are shown in Figure 3.8. It is clearly seen that the peak for PC1 appears at 0.002 $min^{-1}$ and that of PC2 appears at 0.02 $min^{-1}$. There is also a small peak at 0.004 $min^{-1}$ which is the harmonic of 0.002 $min^{-1}$ frequency component. Thus SPCA has detected correctly, the information about the number of frequencies and the frequencies themselves in case of Entech Challenge Problem which causes the whole plant to oscillate.

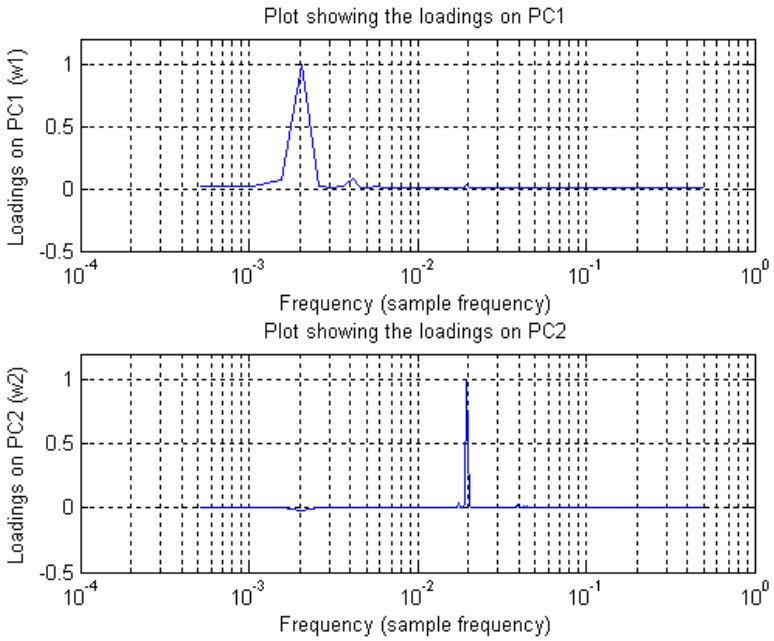



Figure 3.8: Loadings/basis functions of two principal components

### 3.3.2 Industrial Case Study 2 – *Paprican, Canada*

An industrial data set of a Boiler process has been provided by Paprican, Canada. The schematic of the process is as shown in Figure 3.9. The data set comprises of measurements from 11 control loops, each consisting of 2160 samples. There are four flow controllers, six pressure controllers and one pressure indicator in the process. The measurement signals used for analysis are control error signals. The power spectra are estimated for all the loops to infer by visualization, the frequencies affecting the plant's performance. The time trends and power spectra of 11 control error signals (*SP* – *PV*) are as shown in Figure 3.10. The control errors have been mean centered and detrended to remove any linear trends. It can be inferred from power spectra visualization, that the noise is present in almost all the loops.

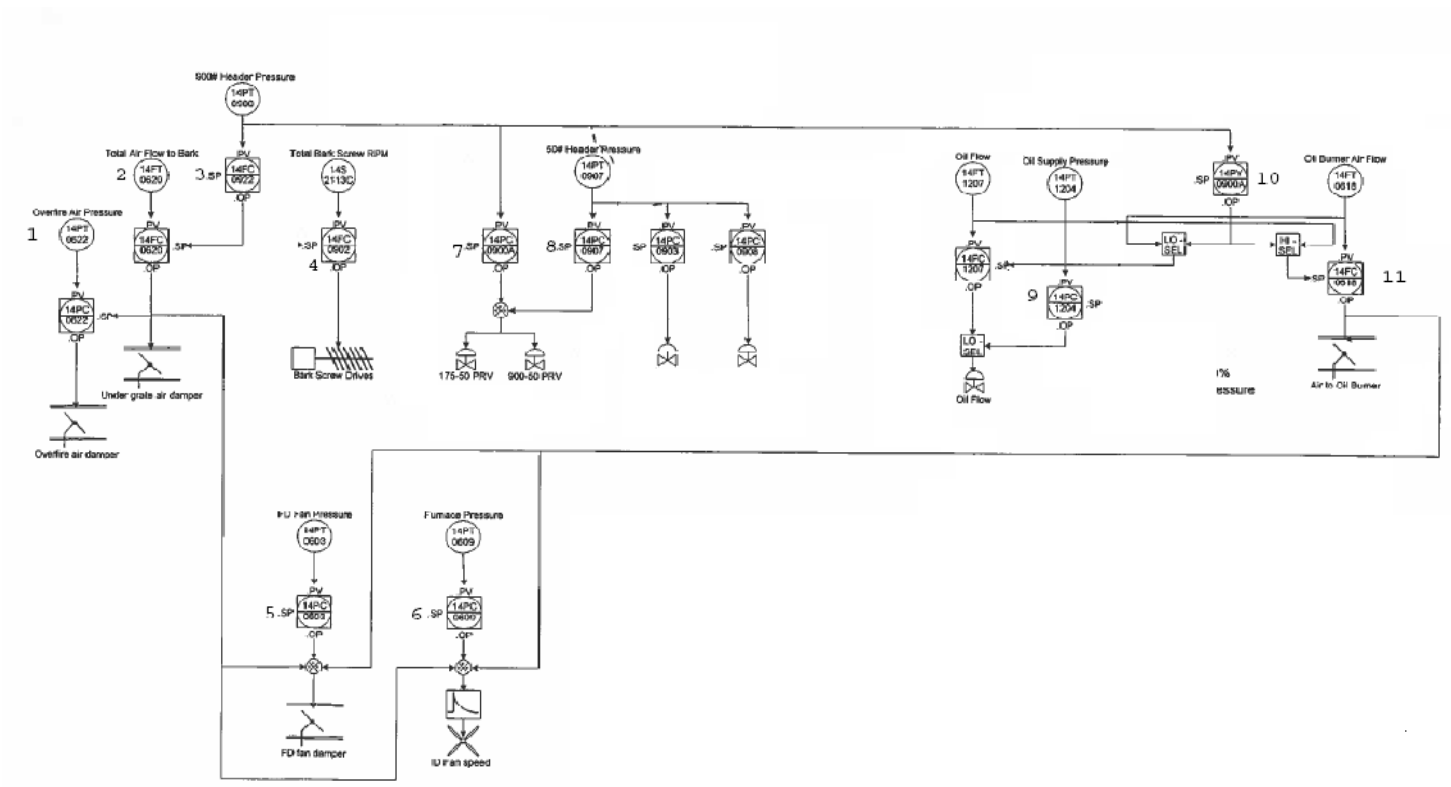



Figure 3.9: Schematic of a Power Boiler Process. Courtesy: Paprican, Canada

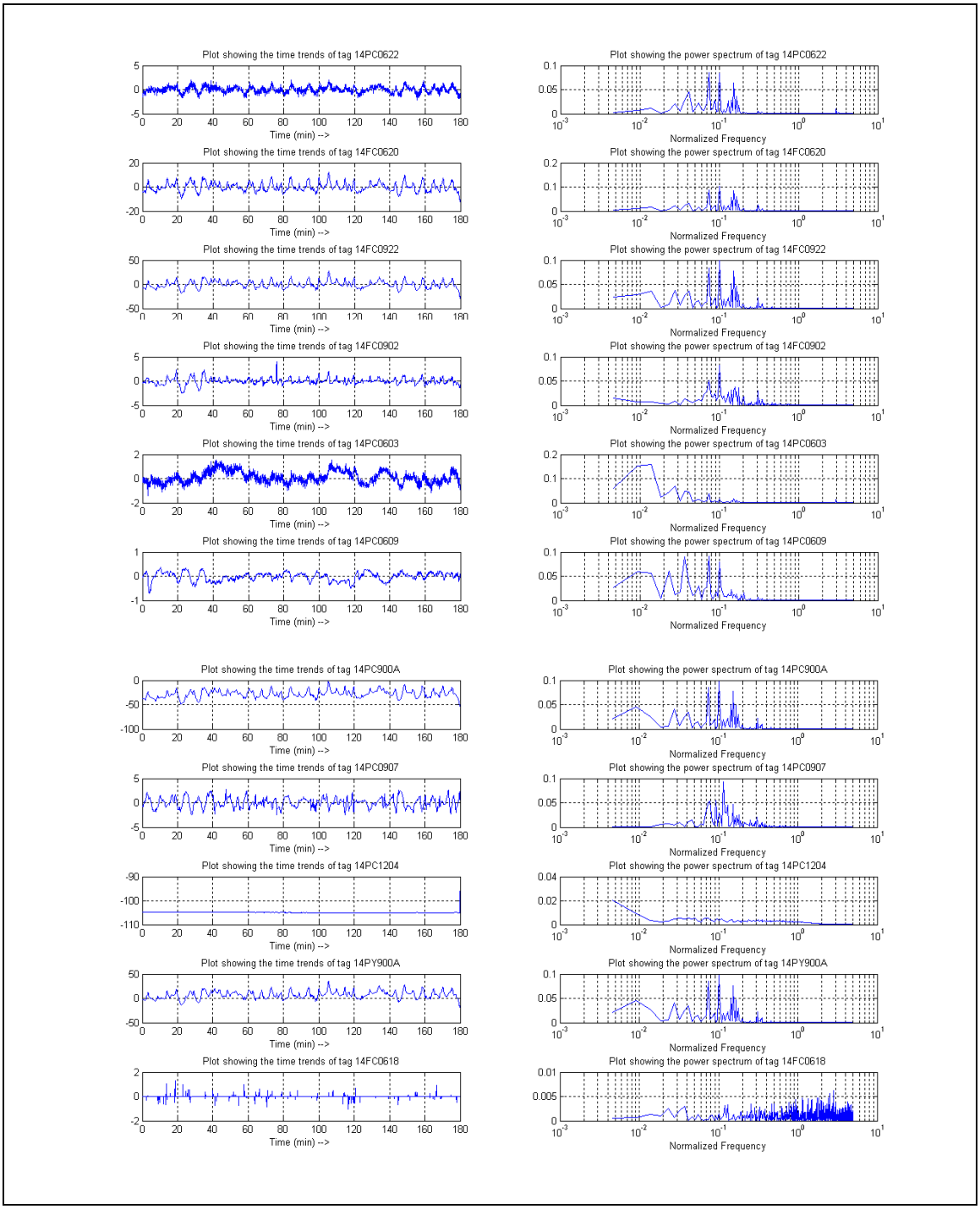



Figure 3.10: Time trends and power spectra of 11 CE signals of Industrial Boiler Process

The frequencies clearly visible in the power spectra causing the plant-wide oscillations are approximately 0.074 $min^{-1}$ (period of 13.33 min), 0.1 $min^{-1}$ (period of 10 min), and 0.16 $min^{-1}$ (period of 6.25 min). There is also a low frequency in the range of 0.009 $min^{-1}$ – 0.012 $min^{-1}$. The power spectrum for tag 11 is corrupted with noise along all the frequencies which is clear in Figure 3.10. The loops affected by the frequencies 0.074 $min^{-1}$, 0.1 $min^{-1}$ and 0.16 $min^{-1}$ are 1 – 4, 6 – 8, & 10, whereas the loops affected by a slower frequencies in the range of 0.009 $min^{-1}$ – 0.012 $min^{-1}$ are 3, 5, 6, 7, 9 & 10.

Implementation of SPCA on Industrial data set resulted in three principal components. Table 3.3 below shows the percentage variance captured by individual principal components. The decomposition of power spectra matrix using PCA has been truncated after three principal components. This is due to the fact that the percentage variance captured by fourth principal component is less than 5 % of total variance captured. The total variance captured by three principal components is approximately 94.21 %. The loadings plot for industrial data set is shown in Figure 3.11.

| Principal Component | Eigen Value of COV($X$) | % Variance Captured |
|---|---|---|
| 1 | 0.0262 | 73.07 |
| 2 | 0.0058 | 15.60 |
| 3 | 0.0019 | 5.53 |

Table 3.3: Percentage variance captured by individual principal components for Industrial data set

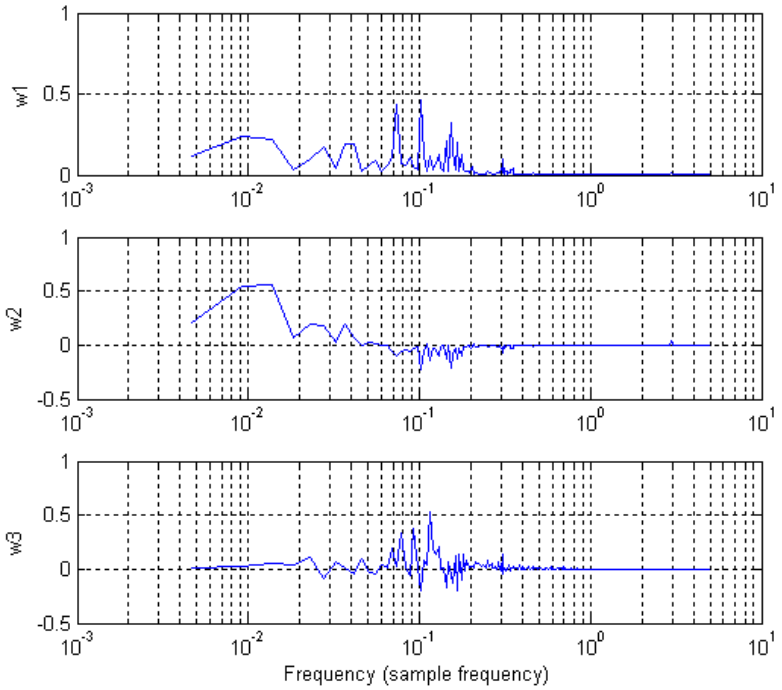



Figure 3.11: Loadings/basis functions of three principal components for industrial data set

The exact frequency values inferred from spectrum like basis functions are 0.074 $min^{-1}$, 0.1 $min^{-1}$ & 0.16 $min^{-1}$ for first principal component, 0.009 $min^{-1}$ – 0.012 $min^{-1}$ and 0.12 $min^{-1}$ for second and third principal components respectively. Some of the basis functions also possess negative values due to which their interpretation becomes difficult. To overcome this problem, methods like Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF) have been developed and will be described in next sections of this chapter. The scores plot for the industrial data set (Figure 3.12) is plotted in a three dimensional plane as the number of principal components are three. The clear distinction between the scores of three principal components is difficult. The tags affected by 0.074 $min^{-1}$, 0.1 $min^{-1}$ & 0.16 $min^{-1}$ are represented by diamond shaped

symbols, whereas the tags affected by a frequency range of 0.009 $min^{-1}$ – 0.012 $min^{-1}$ and 0.12 $min^{-1}$ are represented by squared shaped and round shaped symbols respectively.

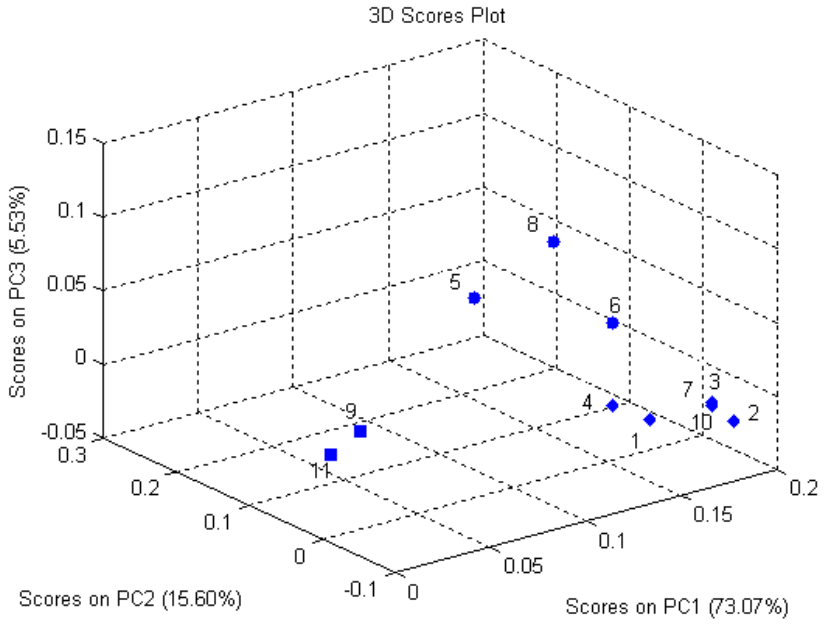



Figure 3.12: Three dimensional scores plot for industrial data set

## 3.4 Independent Component Analysis (ICA) [62]

Independent Component Analysis is a statistical and computational technique for revealing hidden factors that underlie sets of random variables [62]. Nowadays, ICA is becoming increasingly important to analyze large sets of multivariable data. Details pertaining to ICA can be found in Hyvärinen, Karhunen and Oja [28]. ICA is extensively used for blind source separation and feature extraction. ICA was used in the field of process monitoring and fault detection by Lee, Yoo & Lee [34] whereas Xia and Howell [62] were first to use ICA for the purpose of plant-wide oscillation detection.

ICA technique utilizes only the routine operating data and has the ability to locate faults. Moreover, it can isolate different independent sources responsible for the fault (even when more than one source is responsible for one particular fault). ICA uses statistically independent and non-Gaussian components and hence, the method is applied on the power spectral data matrix which is also non-Gaussian. The model used here is the instantaneous mixing model represented as:

$$X = A\,S \qquad \ldots (3.6)$$

where matrix $X = [x_1, x_2, ..., x_m]'$ represents the process variables of $m$ loops, $S = [s_1, s_2, ..., s_n]'$ represents all the independent non-Gaussian sources, and $A$ represents the mixing matrix. In ICA, the main aim is to find both $A$ & $S$. In Xia & Howell [62] it is achieved by calculating the separating matrix $W$ satisfying the condition:

$$\hat{S} = WX \qquad \ldots (3.7)$$

where $\hat{S} = [\hat{s}_1, \hat{s}_2, ..., \hat{s}_n]'$ is the estimate of source matrix $S$, consisting of spectrum like independent components (IC's) of each source, and each element in $W = [w_1, w_2, \ldots, w_n]$ are known as separating vectors constituting mixing matrix.

In the case of plant-wide oscillation detection, data matrix $X$ in Eq. (3.6) is replaced by the normalized power spectral matrix $P$ which is decomposed into mixing matrix $A_{m \times n}$, where $m \geq n$, and independent components matrix $Y$ given as:

$$\begin{bmatrix} p'_1 \\ p'_2 \\ . \\ . \\ p'_m \end{bmatrix} = A_{\mathrm{m}\times n} \begin{bmatrix} y'_1 \\ y'_2 \\ . \\ . \\ y'_n \end{bmatrix} \qquad \ldots (3.8)$$

Each of the IC's $y_i$, represent independent spectrum, and can be found by maximization of cumulant based kurtosis represented as:

$$kurt(\hat{y}_i) = \frac{E\left\{\hat{y}_i{}^4\right\} - 3\left(E\left\{\hat{y}_i{}^2\right\}\right)^2}{\left(E\left\{\hat{y}_i{}^2\right\}\right)^2} = E\left\{\hat{y}_i{}^4\right\} \qquad \ldots (3.9)$$

This kurtosis is very high for Power Spectral Density (PSD) of a pure sinusoid and is very low for PSD of white noise. The procedure described above for the decomposition of power spectral matrix, *P*, is a standard method for finding independent components developed by Hyvärinen, Karhunen and Oja [28]. The programs to perform spectral independent component analysis are readily available in the public domain [27], however, the independent components estimated are not unique in terms of sign and magnitude. Hence a post processing algorithm or constrained ICA has been developed by Xia & Howell [62] and can be described in the following steps:

**(i)** **Step 1:** The estimated independent components are scaled/ normalized as:

$$\bar{y}_j(f) = \frac{y_j(f)}{\sum_{f=1}^{r}\left|y_j(f)\right|}, \quad j = 1,2,3,\ldots,n \qquad \ldots (3.10)$$

where *r* is the total number of frequency channels for the discrete frequencies *f*.

**(ii)** **Step 2:** The mixing matrix, *A*, is replaced by a column scaled version matrix *B*:

$$P = B\bar{Y} \qquad \ldots (3.11)$$

where, $B = A^{*}\, diag(\Delta_1, \Delta_2, ..., \Delta_n)$ , and $\Delta_j = \sum_{f=1}^{r} \left| y_j(f) \right| \quad \forall j = 1, 2...., n$ . In order to make the independent components ($Y$) have spectrum like features, absolute values of $\Delta_j$ are considered.

**(iii)** **Step 3:** Find the sign of maximum absolute value in each normalized independent component $\bar{y}_j$, and represent them as $SN_j \ (j = 1, ..., n)$.

**(iv)** **Step 4:** Define the constrained independent components as:

$$c_j = SN_j \bar{y}_j \qquad \ldots (3.12)$$

thereby forming constrained IC matrix $C = [c_1, c_2, ..., c_n]'$.

**(v)** **Step 5:** Define the constrained mixing matrix:

$$\underline{A} = B\, diag\left(SN_1, SN_2, ..., SN_n\right) \qquad \ldots (3.13)$$

**(vi)** **Step 6:** Substitute Eqs. (3.12) and (3.13) in (3.8) to get the final decomposition as:

$$P = \underline{A}\, C \qquad \ldots (3.14)$$

Matrix $\underline{A}$ is known as Component Related Index (CRI) matrix. A high value of CRI for an independent component indicates as to which loop is oscillating severely thereby leading to the deterioration in control performance. An index known as Component Related Ratio (CRR) has been proposed to measure the percentage by which an individual independent component affects the whole plant and is given as:

$$CRR_{1\times n} = 100\left\{\frac{\sum_{i=1}^{m} CRI_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} CRI_{ij}}\right\}\% \qquad \ldots (3.15)$$

A loop with less than 5% of CRR is assumed to be insignificant in the contribution of performance deterioration. However, in most cases, SPCA is used prior to ICA so as to get information about the number of dominant frequencies affecting the plant.

### 3.4.1 Industrial Case Study 1 – *Entech Challenge Problem*

In section 3.3.1, we described a challenge problem from Entech Control Inc., wherein the hardwood and softwood pulps are mixed to get the desired composition. The schematic of the process is shown in Figure 3.5. The data matrix i.e. the power spectra matrix, $P$, consisting of power spectra for 12 control error signals, has been mean centered and normalized for analysis using independent component analysis (ICA). The time trends and power spectra for each of the control error signal is plotted in Figure 3.6. In section 3.3.1, we found two principal components and hence it can be inferred, that there would be two independent components.

The spectral ICA has been performed on the data set and the resulting two independent components are shown in Figure 3.13 associated with the frequencies 0.002 $min^{-1}$ and 0.02 $min^{-1}$. In the case of challenge problem, SPCA was only able to find out the number of oscillatory frequencies affecting the plant, whereas the ICA also has the ability to locate each oscillatory frequency to their loop of origin. The detection of the root cause is achieved using the Component Related Index (CRI) matrix. A tag with high

CRI value for an independent component gives us the source of oscillation. The CRI matrix for the Entech challenge problem is given in Table 3.4.

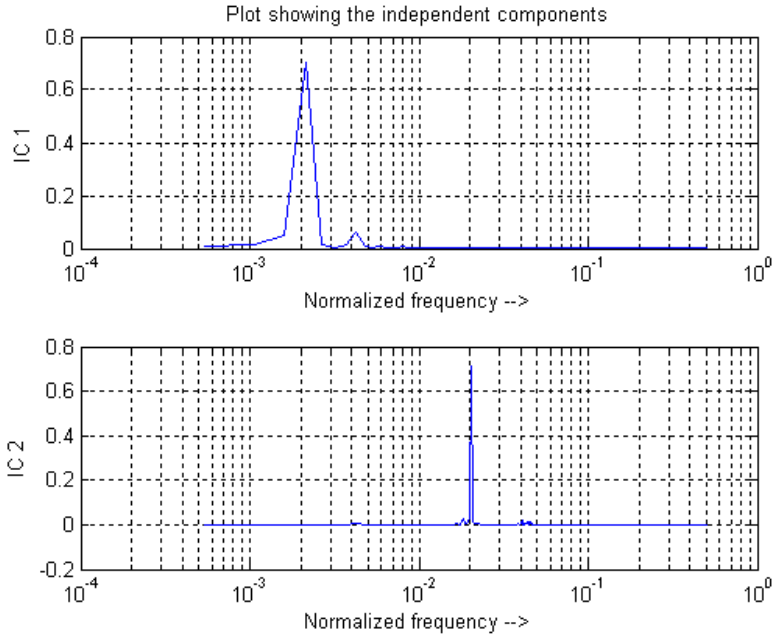



Figure 3.13: Independent Components associated with Entech data set

| CRI | IC1 | IC2 |
|---|---|---|
| Loop 1 | 0.0147 | **1.1591** |
| Loop 2 | 0.0012 | -0.0013 |
| Loop 3 | 1.1062 | 0.0001 |
| Loop 4 | 0.9323 | 0.0107 |
| Loop 5 | 0.3381 | -0.0001 |
| Loop 6 | 0.9309 | 0.0011 |
| Loop 7 | 1.2132 | -0.0002 |
| Loop 8 | 0.7757 | 0.0142 |
| Loop 9 | 0.9619 | 0.0008 |
| Loop 10 | **1.2421** | -0.0001 |
| Loop 11 | 0.0477 | 0.7074 |
| Loop 12 | 0.7070 | 0.1138 |
| **CRR %** | 80.48 % | 19.51 % |

Table 3.4: Component Related Indices (CRI) for each IC relative to all 12 tags of Entech data set

The CRI indices are plotted in form of stacked bar charts to visualize the contribution of each independent component in individual tags as shown in Figure 3.14. The length of each bar represents the sum of the CRI values for a particular loop. Figure 3.14 & Table 3.5, illustrates the dominance of IC1 in $10^{th}$ loop, with a CRI value of 1.2421 and that of IC2 in $1^{st}$ loop, with a CRI value of 1.1591. IC1 needs to be further investigated due to the presence of small peak in the IC1 plot. Hence in conclusion it can is clear that the reason for plant-wide oscillations are the loops 1 & 10. The Component Related Ratio (CRR) is used to show the contribution of each independent component in plant-wide oscillations as is shown in Figure 3.15.

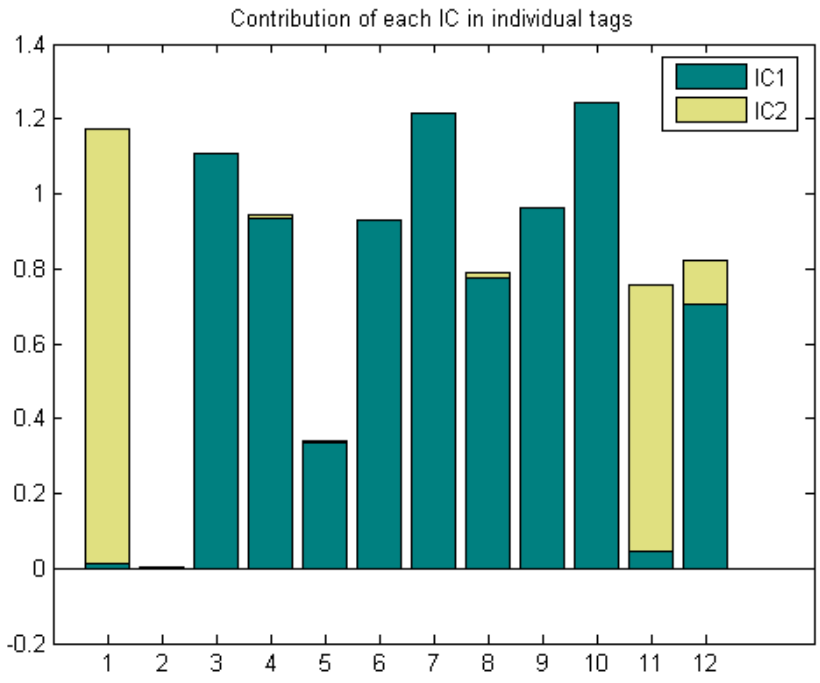



Figure 3.14: Stacked bars showing the contribution of each IC in individual tags

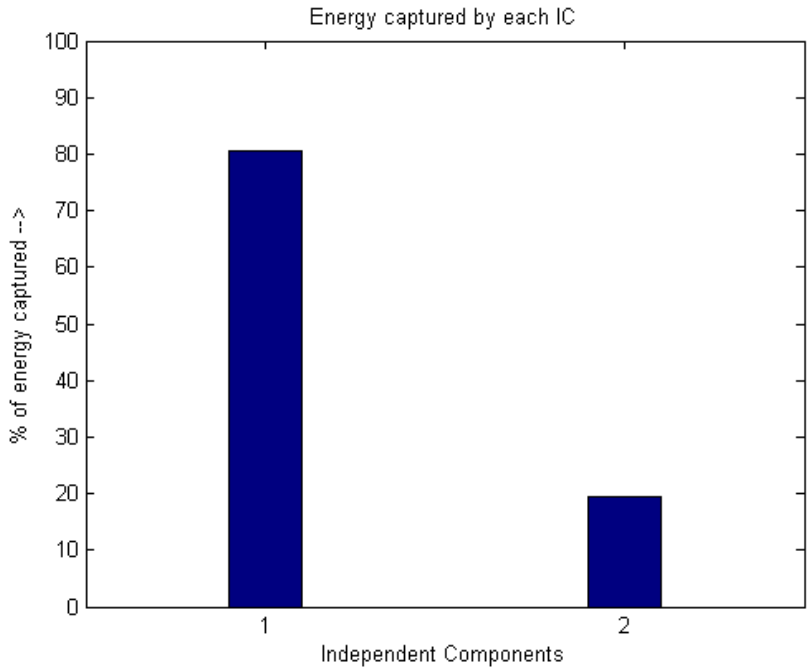



Figure 3.15: Percentage contribution of IC's in plant-wide oscillations for Entech data set

### 3.4.2 Industrial Case Study 2 – *Paprican, Canada*

The industrial case study of a Boiler process consisting of eleven loops, described in section 3.3.2, is used here for analysis using Independent Component Analysis (ICA). The schematic of the process and the time trends and power spectra pertaining to each of the control error signal are illustrated in Figure 3.9 and Figure 3.10 respectively. Analysis of this case study using Spectral Principal Component Analysis resulted in three principal components and hence same number of independent components could be found using ICA.

The power spectra matrix has been mean centered and normalized before decomposing it using ICA. The use of ICA resulted in three independent components as

shown in Figure 3.16. The decomposition of data matrix in ICA results in IC's which are only positive, unlike SPCA (Figure 3.11) wherein the loadings also take negative values.

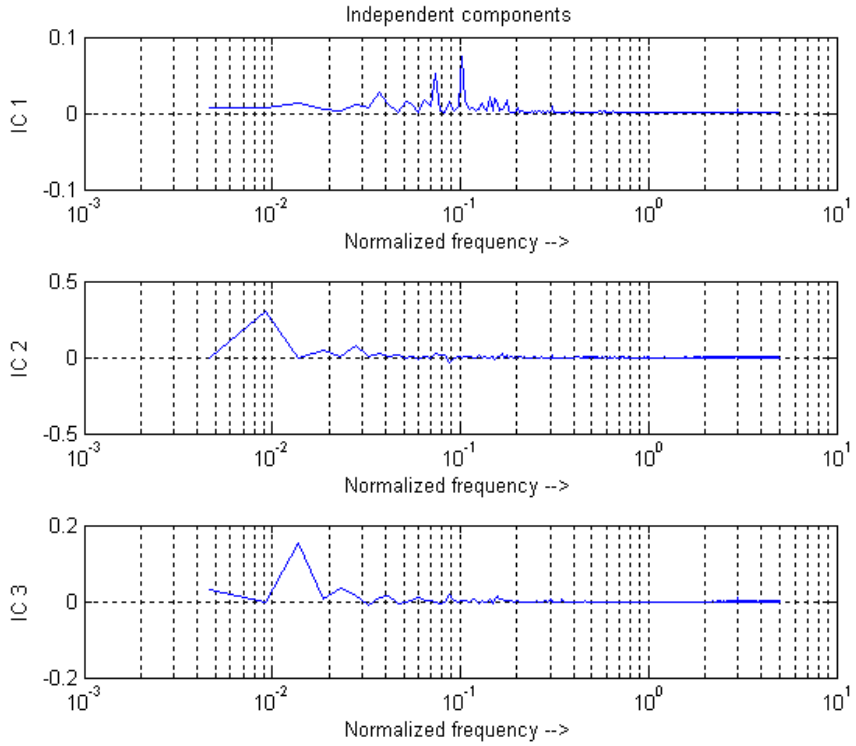



Figure 3.16: Independent Components associated with Industrial Boiler data

The frequencies associated with three independent components are 0.1 $min^{-1}$, 0.009 $min^{-1}$ and 0.0145 $min^{-1}$. The root causes behind the presence of three oscillatory frequencies are found using Component Related Index values as shown in Table 3.5. Figure 3.17 shows the plot of stacked bars representing the contribution of each IC for oscillations in individual loops. Table 3.5 and Figure 3.17 illustrates the dominance of IC1, with a CRI of 1.4068 in loop 14FC0620 and that of IC2 and IC3, with CRI values of 0.5160 and 1.0751 respectively in 14PC0603. The dominance of IC2 and IC3 in a single loop is suspicious in this case, relative to the assumption of statistical independence of individual sources, in ICA.

| CRI | Tag Name | IC1 | IC2 | IC3 |
|---|---|---|---|---|
| Loop 1 | 14PC0622 | 1.1909 | 0.0461 | 0.1849 |
| Loop 2 | 14FC0620 | **1.4068** | 0.0643 | 0.2296 |
| Loop 3 | 14FC0922 | 1.3558 | 0.1250 | 0.3588 |
| Loop 4 | 14FC0902 | 1.0615 | 0.0448 | 0.1395 |
| Loop 5 | 14PC0603 | 0.3410 | **0.5160** | **1.0751** |
| Loop 6 | 14PC0609 | 1.1900 | 0.2215 | 0.4861 |
| Loop 7 | 14PC900A | 1.3463 | 0.1766 | 0.2930 |
| Loop 8 | 14PC0907 | 0.4775 | 0.0118 | 0.0463 |
| Loop 9 | 14PC1204 | 0.0487 | 0.0288 | 0.0241 |
| Loop 10 | 14PY900A | 1.3425 | 0.1775 | 0.2942 |
| Loop 11 | 14FC0618 | 0.0031 | -0.0009 | 0.0031 |
| **CRR %** | | 68.23 % | 9.86 % | 21.91 % |

Table 3.5: Component Related Indices (CRI) for each IC relative to all the 11 tags for Industrial Boiler Process

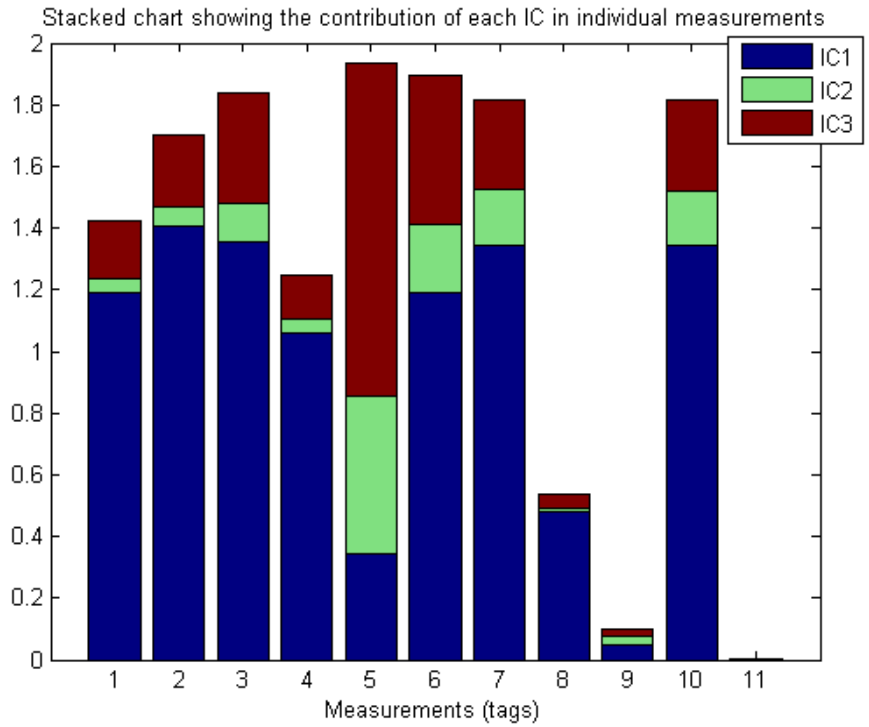



Figure 3.17: Contribution of each IC in individual tags for Industrial Boiler data set

CRR is used to detect the influence of each IC, which altogether causes oscillations in the whole plant. The higher the value of CRR associated with a loop, the stronger is its influence resulting in plant-wide oscillations. Therefore, a loop with the highest CRR value should be given the top priority as it is the main root cause for plant-wide oscillations. Incase of industrial case study of a Boiler process, the highest value of CRR is associated with 14FC0620 (Figure 3.18), and hence it is the prime target to remove oscillations in the plant.

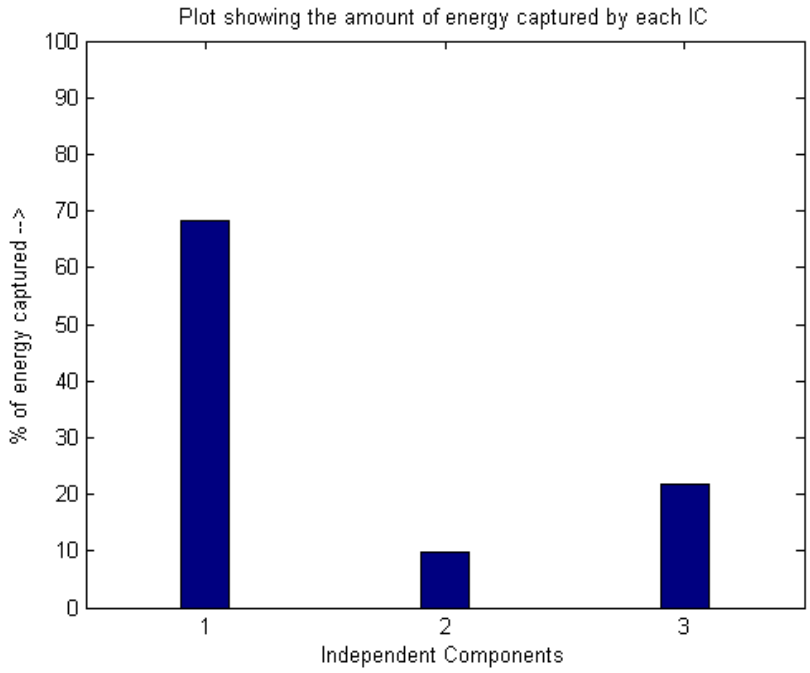



Figure 3.18: Percentage variance captured by each IC for Industrial Boiler data

## 3.5 Non-negative Matrix Factorization (NMF) [52]

Non-negative Matrix Factorization provides parts based representation of multivariate data such that causal basis spectral shapes are retained. NMF has been

extensively used in the area of image analysis, chemometrics and pattern recognition. In NMF, the spectral source doesn't need to be orthogonal, as incase of PCA to find principal components. It is due to this orthogonality property of PCA that, it consists of both positive and negative values of loadings. This problem is overcome with the use of NMF, which considers only non-negative entries for decomposition and also it doesn't pose any statistical assumption on sources as in case of ICA. Details pertaining to NMF can be found in [9] and [10]. Consideration of sparseness concept with NMF has been provided by Hoyer [24]. The remarkable use of NMF for detection of plant-wide oscillations was made by Tangirala *et al.* [52].

In Tangirala *et al.* [52], the multivariate data matrix is decomposed into product of two non-negative matrices using NMF, which represent basis vectors and basis weights.

$$X_{N\times M} = B_{N\times r} W_{r\times M} + E \qquad \ldots (3.16)$$

where $X$ is $N\times M$ ($M \ll N$) data matrix, which here is represented by the power spectra matrix. Each column of $X$ represents power spectra for individual tags. The matrix $B$ of size $N\times r$ represents the basis vectors or basis spectral shapes, where $r$ is the reduced dimension. The matrix $W$ of size $r\times M$ represents the basis weights. The matrix $E$ represents the residual error of reconstruction of original spectral matrix. The residual error in matrix $E$ is zero if $r = M$. Thus matrix $X$ can be represented in rank reduced form as:

$$X_{i\mu} \approx \sum_{k=1}^{r} B_{ik} W_{k\mu} \qquad \ldots (3.17)$$

The power spectra matrix is mean centered and normalized such that the total spectral content over all frequencies is unity. Iterative optimization of the cost function

representing the nearness of $X$ to $BW$ provides solution to NMF. The Euclidean distance based cost function is the most commonly used cost function represented as:

$$\left\| X - BW \right\|^2 = e \qquad \ldots (3.18)$$

The objective here is to minimize error $e$ in equation (3.18). The error is zero when the product $BW$ exactly reconstructs the data matrix $X$. The condition imposed here by NMF is that the elements $\{b_{ik}\}, \{w_{kj}\} \geq 0$.

In order to reach the objective of minimizing the reconstruction error between $X$ and $BW$, two update rules exist which can be stated as:

$$B_{ka} \leftarrow B_{ka} \frac{\left[ XW^T \right]_{ka}}{\left[ B^T WW^T \right]_{ka}}$$

$$W_{aj} \leftarrow W_{aj} \frac{\left[ B^T X \right]_{aj}}{\left[ B^T BW \right]_{aj}} \qquad \ldots (3.19)$$

where $W_{aj}$ refers to a$^{th}$ row of $W$ and $B_{ka}$ refers to a$^{th}$ column of $B$. The proof of convergence for the above stated multiplicative update rules can be found in Lee and Seung [9]. The initialization of matrices $B$ & $W$ is an open ended issue, but most of the researchers have used random non-negative matrices. Tangirala *et al.* [52] proposed the use of SVD for the initialization of matrices $B$ & $W$, as it gives better solution when compared to random initialization.

*Note: The cost function in Eq. (3.18) is convex either in B or W, but not in both simultaneously. This shows that only a locally optimum solution exists for the given cost function using NMF method.*

A new concept of total power plot has been introduced in Tangirala *et al.* [52]. The power spectra's associated with each tag in a data matrix are added up at their respective frequencies as given in Eq. (3.20). It provides a good overview of all the existing oscillatory frequencies in a plant, thereby leading to plant-wide oscillations.

$$X_T(f_l) = \sum_{j=1}^{M} X_j(f_l) \qquad \ldots (3.20)$$

where $X_j$ is the power spectrum associated with measurement $j$.

The Non-negative Matrix Factorization of the power spectra matrix $X$ into $B$ & $W$ can be interpreted as the decomposition of the total power $X_T$ into spectra like individual power components provided by each basis shape which can be formulated as:

$$C_k = \sum_{j=1}^{M} B_k W_{kj} \qquad (\text{for } k = 1,...,r) \qquad \ldots (3.21)$$

where $C_k$ is the spectrum like power component due to $k^{\text{th}}$ basis shape, $B_k$ is the $k^{\text{th}}$ column of $B$, and $W_{kj}$ is the $k^{\text{th}}$ element of $j^{\text{th}}$ variable corresponding to $B_k$.

A novel measure termed as Pseudo Singular Value (PSV) has been proposed to determine the basis size. The PSV is based on the incremental variance captured by each power component relative to the total power. The PSV for $j$th basis shape can be represented as:

$$\rho_j = \begin{cases} \dfrac{\left\| \sum_{k=1}^{j} C_k \right\|_2^2 - \left\| \sum_{k=1}^{j-1} C_k \right\|_2^2}{\|X_T\|_2^2} & (\text{for } j = 2,...,M) \\ \dfrac{\|C_1\|_2^2}{\|X_T\|_2^2} & (\text{for } j = 1) \end{cases} \qquad \ldots (3.21)$$

where $C_k$ is the $k^{\text{th}}$ power component and $X_T$ is the total power. The PSV is in the range of [0 1] with a value closer to 1 indicating strong influence of the corresponding basis shape (i.e. the power captured by that particular power component completely describes the total power) and a value closer to 0 indicates negligible influence. The PSV's for all the variables are calculated at once in the order power components are obtained and then rearranged in the decreasing order. If the first *r* power components are able to represent total power then the remaining $M - r$ power components are set equal to zero.

The basis weights obtained are a good measure to indicate the strength of each power component but when noise overshadows the power content, it cannot represent efficiently the strength of each power component. Hence a novel measure that indicates the correct strength of a particular power component even in presence of noise is proposed and is termed as '*Strength Factor*' (SF). The Strength Factor can be interpreted as the decomposition of the individual spectra into its constituent components. The Strength Factors are calculated after reordering basis shapes and corresponding weights in the order PSV's are rearranged. The Strength Factor of $k^{\text{th}}$ power component is given as:

$$SF_{kj} = \begin{cases} \dfrac{\left\| \sum_{p=1}^{k} B_p W_{pj} \right\|_2^2 - \left\| \sum_{p=1}^{k-1} B_p W_{pj} \right\|_2^2}{\left\| X_j \right\|_2^2} & \left(\text{for } k = 2,...,M \; ; \forall j\right) \\ \dfrac{\left\| B_1 W_{1j} \right\|_2^2}{\left\| X_j \right\|_2^2} & \left(\text{for } k = 1 \; ; \forall j\right) \end{cases} \quad \ldots (3.22)$$

SF can be analogously represented as PSV. SF lies within the range of [0 1] with a value of 1 indicating that *k*th power component entirely captures the spectral behavior

of $j^{th}$ measurement and a value equal to zero indicating negligible presence of $k^{th}$ power component in $j^{th}$ measurement. The summation of SF's of $r$ reduced components is less than or equal to unity i.e. $\sum_{k=1}^{r} SF_{lj} \leq 1 \ \forall \ j$ .

### 3.5.1 Industrial Case Study 1 – *Entech Challenge Problem*

The example described in section 3.3.1 is considered here to illustrate the effectiveness of NMF over SPCA and ICA. The process is a challenge problem from Entech Control Inc consisting of a pulp manufacturing process. Hardwood and softwood are mixed to get the resulting pulp in desired proportion. The schematic of the setup is shown in Figure 3.5. The data consists of control error signals from 12 process measurement tags. The power spectra of 12 control error signals are mean centered and normalized before the analysis using Non-Negative Matrix Factorization. The time trends and power spectra of all the 12 tags are presented in Figure 3.6.

The analysis of this data set using SPCA and ICA resulted in two principal and independent components. The frequencies for these two components were found to be 0.002 $min^{-1}$ and 0.02 $min^{-1}$. The number of principal components was found based on the percentage variance captured by them which shouldn't be less than 5 %. Two novel measures known as total power plot and Pseudo Singular Value (PSV) has been presented in [52]. The visualization of total power plot could illustrate the total number of oscillatory frequencies affecting the whole plant and same is the case with PSV, which could determine the number of dominant components capturing maximum variance. The total power plot and PSV's are shown in Figure 3.19 and Figure 3.20 respectively. If we

see the total power plot in Figure 3.19, we could clearly see strong peaks at frequencies 0.002 $min^{-1}$ (period of 500 min) and 0.02 $min^{-1}$ (period of 50 min). There is also a small peak at 0.004 $min^{-1}$ with a period of 250 min which was not discovered with SPCA and ICA.

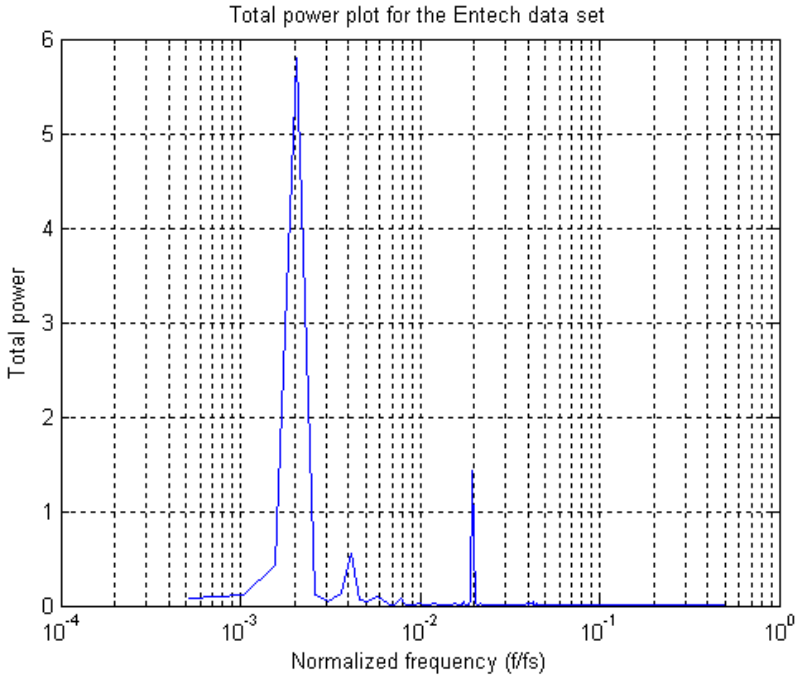



Figure 3.19: Total power plot for Entech data set

It is inferred from Figure 3.20 that PSV's beyond three are very small and hence can be neglected. The use of NMF for Entech data set results in three power components which provides basis shapes as shown in Figure 3.21. The strength factors associated with each power component provides the quantitative information about the influence of individual power component in each of the measurements as shown in Figure 3.22. Table 3.6 shows the ranking of loops in accordance with the effected frequencies.

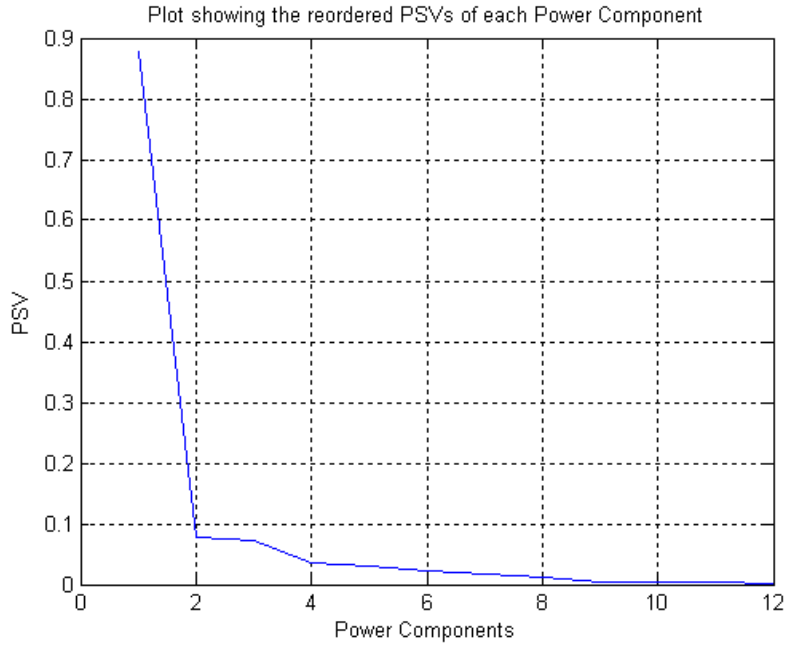



Figure 3.20: Pseudo Singular Values obtained in Entech data set

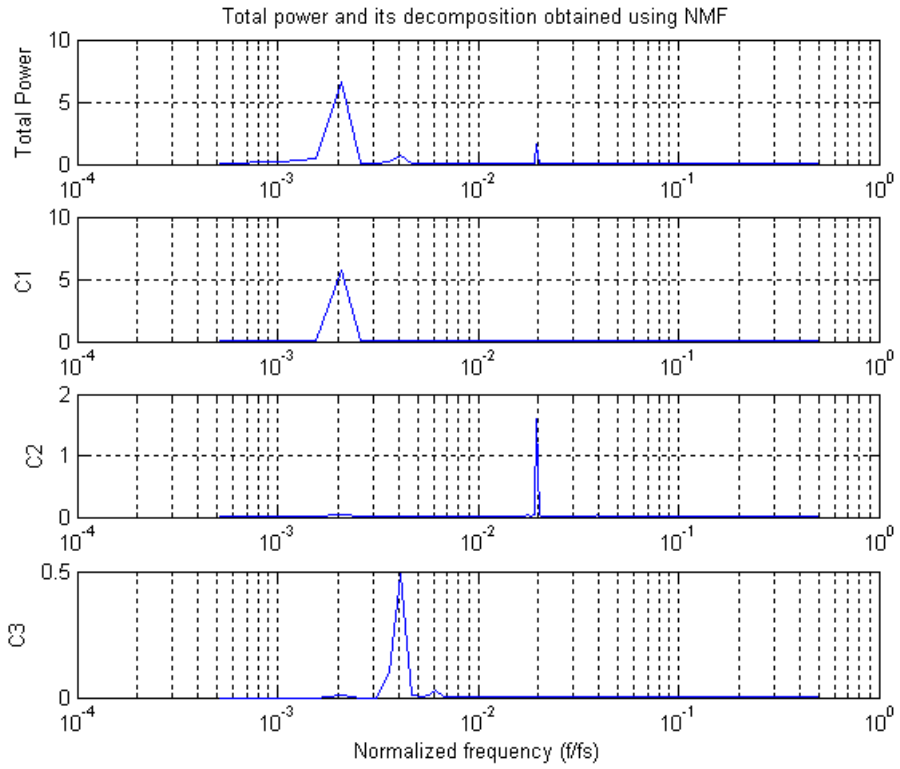



Figure 3.21: Total power plot and its decomposition using NMF for Entech data set

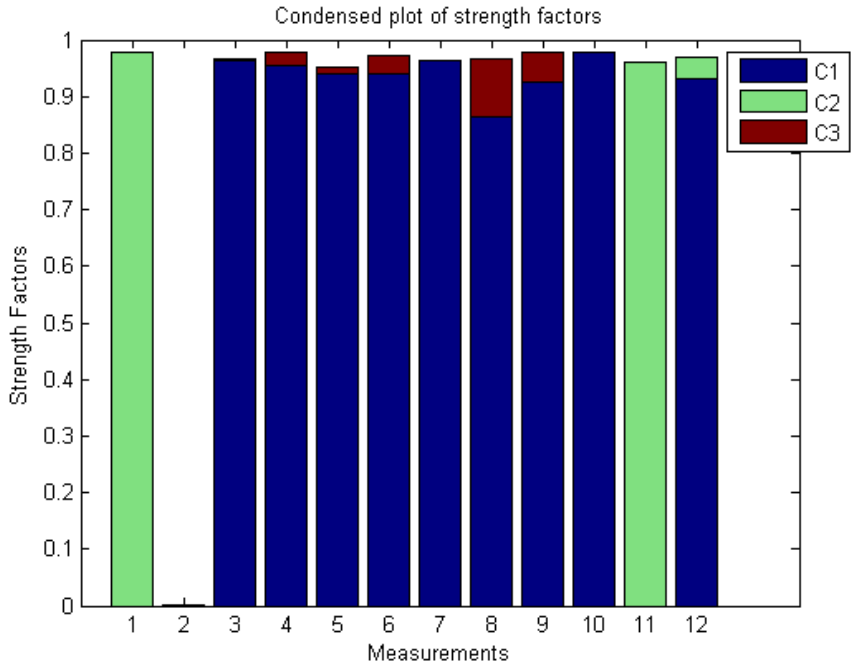



Figure 3.22: Condensed plot showing the stacked strength factors in 12 tags for Entech data set

| Component | Frequencies | Measurements (ranked in order) |
|---|---|---|
| 1 | 0.002 | 10, 3, 7, 4, 5, 6, 12, 9, 8, 1, 2, 11 |
| 2 | 0.02 | 1, 11, 12 |
| 3 | 0.004 | 8, 9, 6, 4, 5, 3, 2, 1, 7, 10, 11, 12 |

Table 3.6: Ranking of measurements/tags for Entech data set

### 3.5.2 Industrial Case Study 2 – *Paprican, Canada*

Non negative Matrix Factorization is applied on a data set from an industrial boiler process described in section 3.3.2. The schematic for the process is shown in Figure 3.9. The data set consists of eleven control error signals for analysis. The control

error signals have been mean centered before estimation of their respective power spectra. The power spectrum of each measurement is normalized such that the area under each curve is unity. The time trends and power spectrum pertaining to each measurement is shown in Figure 3.10. The total power plot gives the visualization of all the existing frequencies in the plant which here is illustrated in Figure 3.23. We could interpret three clear peaks at the frequencies 0.074 $min^{-1}$, 0.1 $min^{-1}$ and 0.16 $min^{-1}$. There are also low frequency peaks in the range of 0.009 $min^{-1}$ – 0.028 $min^{-1}$.

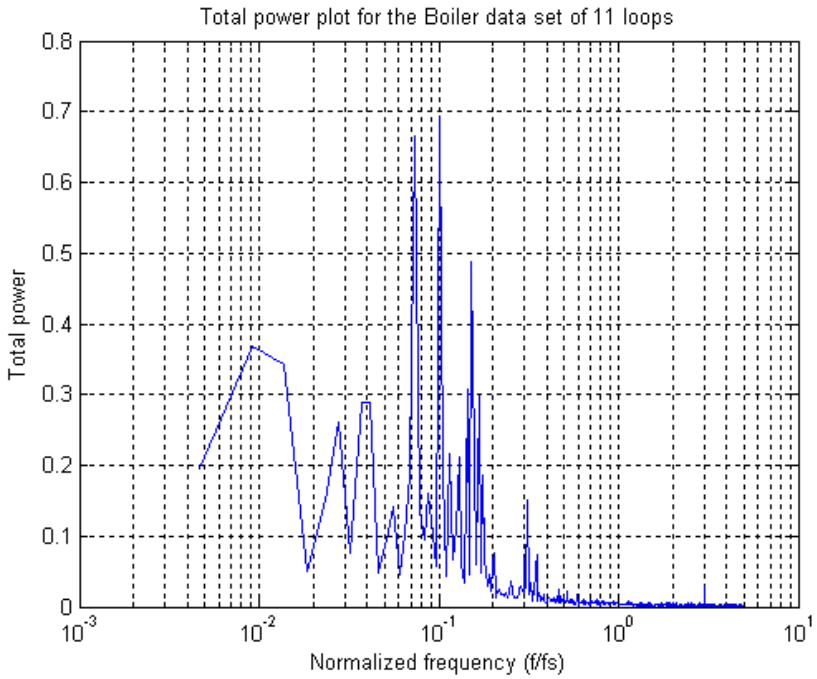



Figure 3.23: Total power plot for Industrial Boiler process

In case of such an industrial data, wherein the noise and external affects has corrupted the measured data, keen observation is required with the use of total power plot and PSV's to visualize power components. The PSV's for industrial data resulted in three power components that could capture most of the variance in the power spectra as shown

in Figure 3.24. The decomposition of total power spectra into power components is not that efficient as incase of Entech data set. This is due to the presence of harmonics of the frequencies which are further corrupted with high amount of noise. The highest peaks incase of first and third power components are obtained at 0.074 $min^{-1}$ and 0.12 $min^{-1}$ respectively, whereas incase of second power component, the highest peak is not visible clearly and a band of peaks is visible. One more point to be noted here is the presence of highest peak at 0.1 $min^{-1}$ in total power plot whereas incase of first power component the highest peak occur at a frequency of 0.074 $min^{-1}$. This may also occur due to the fact that NMF method could only find a local minimal solution of the data set.

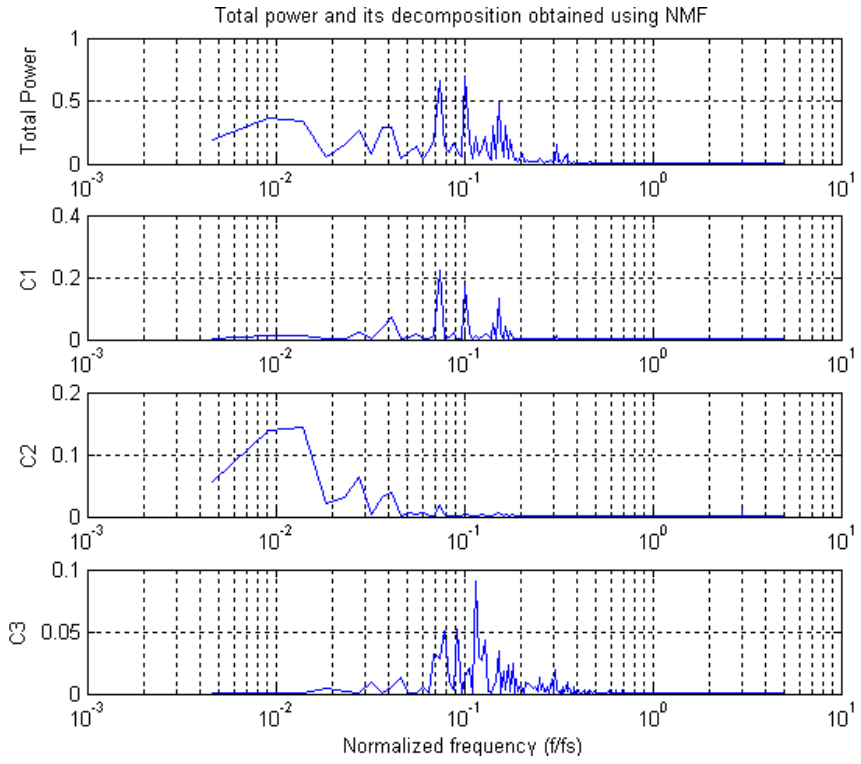



Figure 3.24: Total power plot and its decomposition using NMF for Industrial Boiler data set

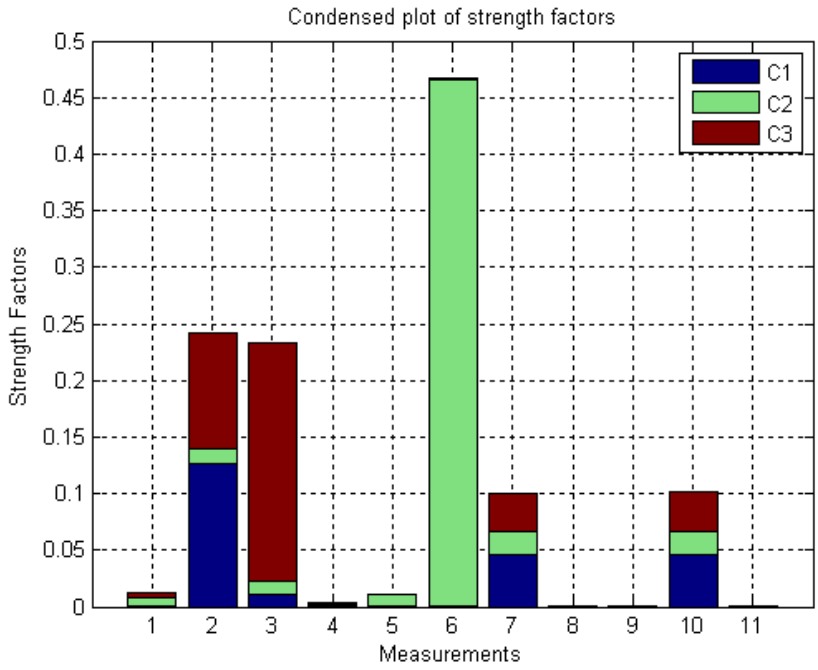



Figure 3.25: Condensed plot of strength factors associated with three power components

| Component | Frequencies | Measurements (ranked in order) |
|---|---|---|
| 1 | 0.074 | 2, 10, 7, 3, 1, 4, 5, 9, 6, 8, 11 |
| 2 | 0.01 – 0.015 | 6, 10, 7, 3, 2, 5, 1, 4, 8, 9,.11 |
| 3 | 0.12 | 3, 2, 10, 7, 1, 6, 5, 4, 9, 8, 11 |

Table 3.7: Ranking of measurements/tags for Industrial Boiler data set

In conclusion, we can say that NMF method can perform very well incase of low noisy data whereas its performance deteriorates in the presence of high amount of noise and external effects. Also, it provides only a locally minimal solution as a result of decomposition of the power spectral data matrix into basis shapes and their respective weights. To overcome this problem of providing a local minimal solution, a method

which can find and reach global minima of the power spectral decomposition is presented in next chapter.

## 3.6 Digraphs based Oscillation detection

A process plant consists of loops varying from few to thousands in number. It is rather a very difficult task to detect all the loops which leads to plant-wide oscillations. But, research has progressed well and many techniques have been developed to detect the root cause of oscillations. Most of the methods developed are based on routine operating data i.e. *SP, PV* and *OP*. These methods aim to find the root cause without any knowledge about the process. It is always important to use process knowledge when dealing with the root cause detection of plant-wide oscillations.

Thornhill *et al.* [58] proposed the use of process understanding when dealing with plant-wide oscillations as it enhances the possibility of root cause detection accurately. Maurya *et al.* [40] proposed the use of graph based approach known as *signed digraphs* to deal with fault detection and safety analysis. More recently, the signed digraph approach was used by Jiang [31] to deal with the problem of plant-wide oscillations. He used the method to analyze all the control loops in a plant, and hence named the approach as *Control loop digraphs*. In addition to a data based method, the use of control loop digraphs could drastically enhance the probability of being right in the root cause detection for plant-wide oscillations.

### 3.6.1 Digraphs and Adjacency Matrix

A graph is a representation which shows the relationship among different variables being analyzed. Set of variables are represented by nodes or points and the relationship between them is represented by lines or edges. A process consists of a flow of material, energy, etc. This flow could be represented in a meaningful manner by directional signs between different variables. The process flow represented by directional arrows forms a graph between the variables, and is known as control loop digraphs (directional-graphs).

A simple example from Jiang [31] explains the meaning of digraphs. Consider five nodes which are connected together as shown in Figure 3.26. The most common approach to represent a digraph is by Adjacency Matrix. The rows and columns of Adjacency Matrix are equal to the number of nodes. If there is an edge or directed line from node *i* to node *j*, then the entry for this is supposed to be one in adjacency matrix. On the contrary, if there is no directed line between node *i* and *j*, then the entry is supposed to be zero.

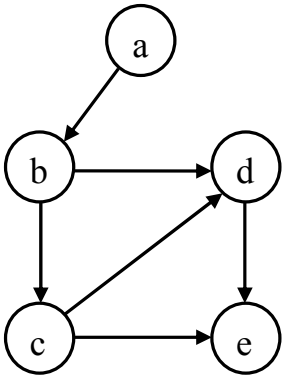



Figure 3.26: A simple digraph

Adjacency matrix for the example shown in Figure 3.26 is as shown below:

$$X = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Figure 3.27: Adjacency matrix for a simple digraph

One of the most useful properties of adjacency matrix which could be used for plant-wide oscillations detection is its ability to show the number of *k*-step edges in the $(i, j)^{th}$ entry of matrix $X^k$. If the entry (2,5) of $X^2$ matrix is considered as shown in Figure 3.28, we can see the value of that entry as 2. This represents the number of steps from node *b* to node *e* are 2, *i.e.* the two paths $\{b \to c \to e\}$ and $\{b \to d \to e\}$. The Boolean equivalent of the matrix *A*, can be given as [36]:

$$A^{\#}(i, j) = \begin{cases} 0 & \text{if } A^{\#}(i, j) = 0 \\ 1 & \text{if } A^{\#}(i, j) \neq 0 \end{cases} \quad \ldots (3.22)$$

Reachability matrix shows the inter-relation of each of the variables or the connectivity in a digraph. The reachability matrix for a digraph with *N* nodes can be represented as:

$$R = \left(X + X^2 + X^3 \ldots + X^N\right)^{\#} \quad \ldots (3.33)$$

Reachability matrix for the five node example considered is also shown in Figure 3.28. From matrix *R,* it can be inferred that node *a* can reach all the other nodes, but no other node can reach node *a,* and same is the case with other nodes in matrix *R*.

$$X^2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad X^3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$X^4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad X^5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 3.28: Higher powers of Adjacency matrix and Reachability matrix

The methodology of digraphs and adjacency matrix was used by Jiang [31], but he proposed some modifications in the digraph strategies. For plant-wide oscillation detection, each of the controllers in a plant is taken as a node. If the controller output of node *i* directly affects the process variable of node *j*, then such a connection is taken as an edge. One more point to be considered here is, the controller output of $i^{th}$ node, firstly affects its own process variable then affect other loops. With these modifications, digraphs were termed as *control loop digraphs* in [31]. Two of the case studies considered before are used in the next section to analyze plant-wide oscillations using control loop digraphs.

### 3.6.2 Industrial Case Study 1 – *Entech Challenge Problem*

The control loop digraph for the Entech challenge problem is shown in Figure 3.29. The red lines marked in the figure represent the direct interactions in the process. The process consists of twelve control loops wherein hardwood and softwood are mixed together to get the desired composition of pulp. Each of the controllers in the process is taken as a node. Tag 7 and tag 3 are master and slave controllers respectively in the cascade configuration. Tag 3 and tag 6 are ratio controllers onto the master controller tag 7. Any change in level of the blend tank affects the flow rate of hardwood and softwood from tags 3 & 6 respectively. Change in control output of tag 1 affects the process variable of tag 2 and vice versa. Tag 4 & tag 5 are also in same configuration as that of tag 1 and tag 2 wherein the change in control output of one affects the process variable of other. Also, tag 9 and tag 10 are in cascade configuration. Change of level in tag 10 directly affects the flow rate of tag 9. Tag 12, which is the control for final flow rate of the desired pulp combination, is having a feed-forward connection from tag 11, the control for white water header.

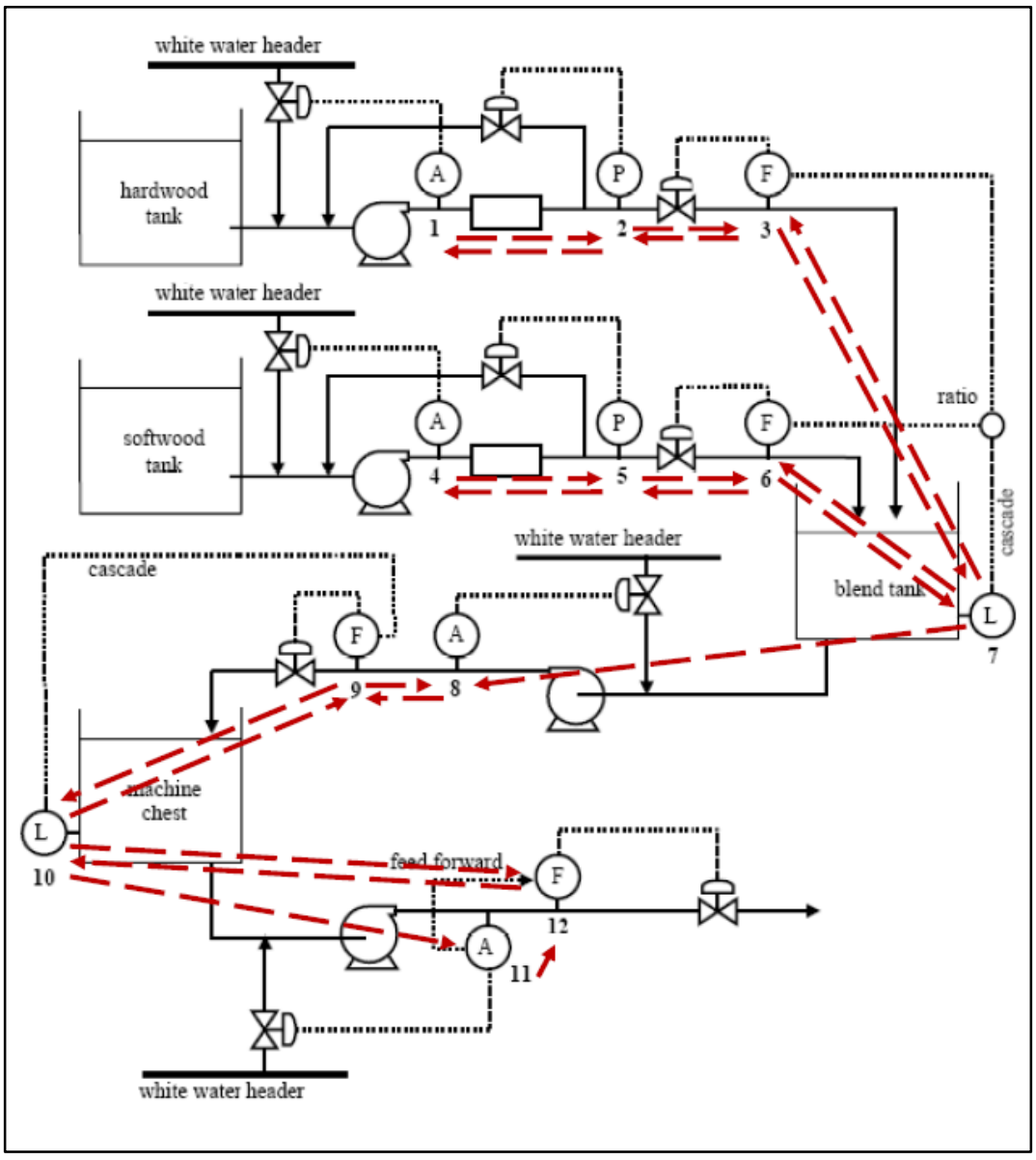



Figure 3.29: Control Loop Digraph for Entech Challenge problem

The adjacency matrix involving all the directions in the control loop digraph of Entech Challenge process is shown in Figure 3.30. As has been described earlier, a controller firstly affects it own process variables and then others. Due to this, all the diagonal entries are unity in the Adjacency matrix. The reachability matrix for Entech Challenge problem is shown in Figure 3.31. It shows the connectivity of all the control loops within a plant. The control loop which has the maximum connectivity in the plant, it is termed as the most probable element for diagnosis of plant-wide oscillations. In Figure 3.31, the loops which have the maximum connectivity are found to be $4^{th}$, $5^{th}$ and $6^{th}$ loops. This solution is in contrast to all the oscillation detection methods used in earlier sections, wherein loops 1 & 10 are found to be more likely causes for plant-wide oscillations. Therefore, a further diagnostic method is needed to figure out the most probable root cause. The most efficient method developed till now for diagnosis of control loops is based on Higher Order Statistics, developed by Choudhury *et al.* [6] described in chapter 5.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |

Figure 3.30: Adjacency matrix for Entech Challenge problem

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 3 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 7 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |

Figure 3.31: Reachability matrix for Entech Challenge problem

### 3.6.3 Industrial Case Study 2 – *Paprican, Canada*

The industrial case study of a Power Boiler process, courtesy of *Paprican, Canada* is analyzed using control loop digraphs. The process consists of 12 PID control loops. Each of the controllers is taken as a node and the direct interactions between them are taken as edges in control loop digraph shown in Figure 3.32. The red lines in the figure show the direct interaction between the controllers. Tags 5 and 6 are in cascaded configuration. If the controller output of one of them changes, then the process variable of other gets affected. Tags 5, 6 and 12 are in ratio control mode, wherein 12 is the master and the other two are slaves. Tag 2 also has direct interaction with tags 5 and 6. Tag 3 has direct interactions with tags 2, 4, 7 and 11. The adjacency matrix showing all of the connections in the control loop digraph is presented in Figure 3.33. As described earlier, each of the diagonal is taken as unity as the control output of any controller affects its own process variable at first and then affects others. The reachability matrix for

Power Boiler process is shown in Figure 3.34, which shows the overall connectivity in the plant. The reachability matrix indicates loops 2 (14FC0620), 3 (14FC0922) and 4 (14FC0902) to have maximum connections in the plant. Therefore, the analysis of the Power Boiler process using routine operating data also indicate loop 2 (14FC0620) as the most affected loop. The results are in agreement with that of spectral decomposition methods. The root cause diagnosis can be done for the suspected loops 2, 3 and 4 using Higher Order Statistics, which is presented in chapter 5. One more thing to be noted is, the oscillation analysis using digraphs only indicate the loops which are suspected to be the root cause and doesn't point out exactly a particular loop. Thus, the method is used only to include the process knowledge for plant-wide oscillations analysis.

## 3.7 Conclusion

In this chapter, firstly, the concept of oscillations and plant-wide oscillations are explained. This is followed by a complete analysis of plant-wide oscillations using four of the methods developed (SPCA, ICA, NMF and Control Loop Digraphs) earlier. Two industrial case studies have been used so as to know, which of the developed methods perform well. It has been found that although SPCA is the most commonly known method to gain information about the number of frequencies affecting a plant, but also has a drawback of having negativity in basis shapes. On the other hand, ICA and NMF methods overcome the problem of negativity in basis shapes, but they also have their own drawbacks. ICA assumes the statistical independence of sources which is difficult to verify, whereas NMF can provide only a local minimal solution. Control loop digraphs provide process understanding to deal with the problem of plant-wide oscillations. The

drawbacks associated with the existing techniques are overcome in the proposed method based on Evolutionary Algorithms, described in chapter 4.

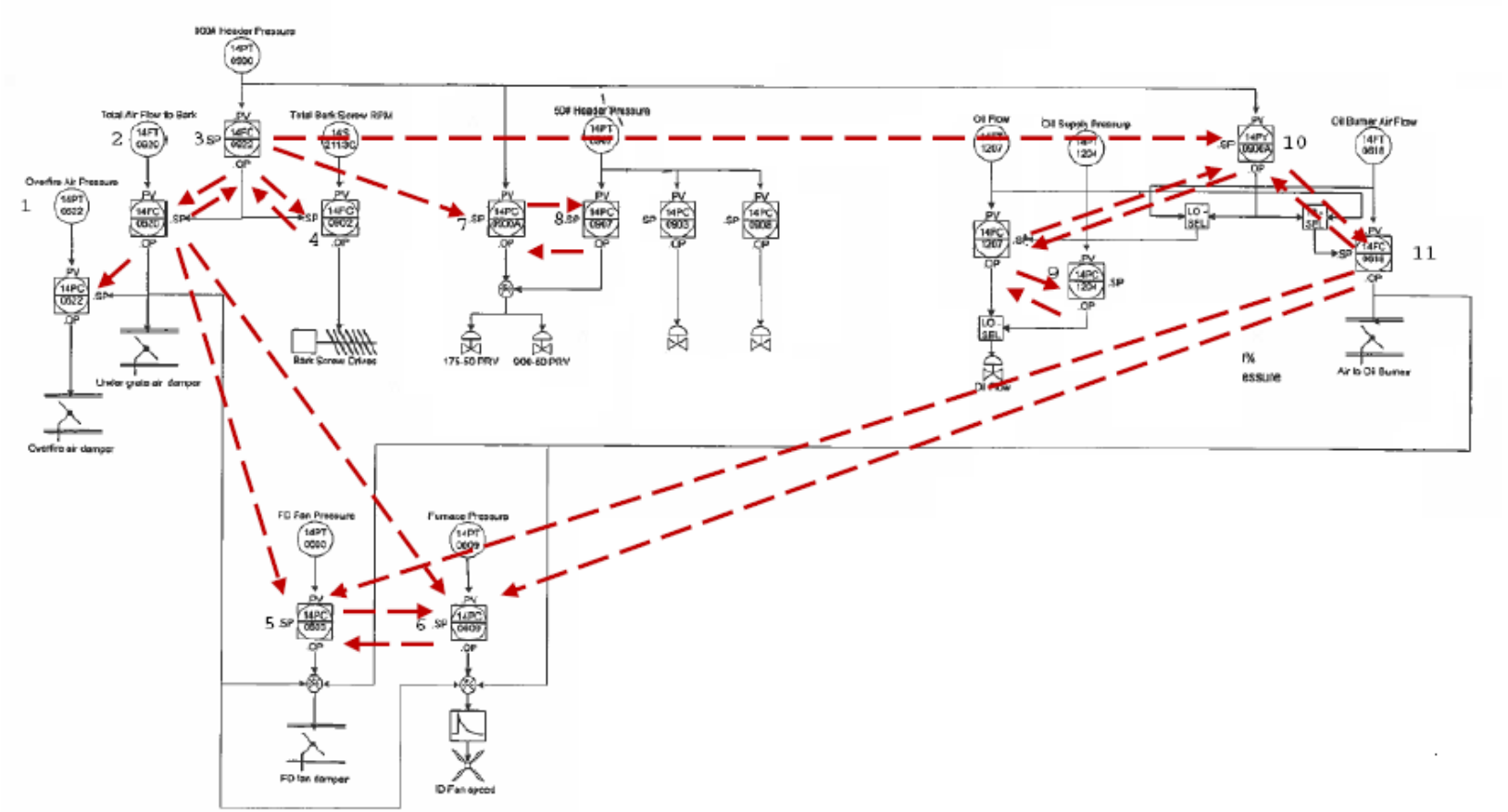



Figure 3.32: Control loop digraph for Industrial Boiler process

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 14PC0622 | 14FC0620 | 14FC0922 | 14FC0902 | 14FC0603 | 14FC0609 | 14PC900A | 14PC0907 | 14FC1207 | 14PC1204 | 14PY900A | 14FC0618 |
| 1 | 14PC0522 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 14FC0620 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 14FC0922 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 4 | 14FC0902 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 14PC0503 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 14PC0509 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 14PC900A | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | 14PC0907 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 9 | 14FC1207 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 10 | 14PC1204 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 11 | 14PY900A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 12 | 14FC0618 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |

Figure 3.33: Adjacency matrix for the Industrial Boiler process

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 14PC0622 | 14FC0520 | 14FC0922 | 14FC0902 | 14PC0603 | 14PC0609 | 14PC900A | 14PC0907 | 14FC1207 | 14PC1204 | 14PY900A | 14FC0618 |
| 1 | 14PC0522 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 14FC0620 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 14FC0922 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 14FC0902 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 14PC0503 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 14PC0509 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 14PC900A | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | 14PC0907 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 9 | 14FC1207 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 10 | 14PC1204 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 11 | 14PY900A | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 12 | 14FC0618 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |

Figure 3.34: Reachability matrix for Industrial Boiler process

# CHAPTER 4

# OPTIMIZATION BASED PLANT-WIDE OSCILLATION DETECTION

## 4.1 Introduction

Recall that the degradation in the performance of a plant can be observed in form of oscillations in the time trends of measurements. The causes behind the oscillations could be improper controller tuning, valve nonlinearities or external oscillatory disturbances. Accurate detection and diagnosis of the causes of such oscillations results in substantial profit of an industry. Each of the methods described in chapter 3 has a drawback associated with them. The drawbacks associated with SPCA are twofold. The first is the presence of negativity in the basis shapes whereas the second is its inability to analyze the results in space if number of principal components is greater than three. ICA poses a strong assumption of statistical independence of sources whereas NMF can provide only a local minimal solution to the problem in focus. This chapter proposes a new method based on Evolutionary Algorithms (more commonly known as Genetic Algorithms) to detect the root cause of plant-wide oscillations. The proposed method overcomes the problem of local minimal solution associated with NMF method. One

simulation example and two industrial case studies used in chapter 3 shows the efficiency of proposed approach over other spectral decomposition methods such as ICA and NMF.

## 4.2 Evolutionary Algorithms

Evolutionary Algorithms (EA's) are search methods that take their inspiration from Darwin's theory of natural selection and survival of the fittest in the biological evolution. EA's are a blend of Evolutionary Programming (EP), Evolutionary Strategies (ES) and Genetic Algorithms (GA). The common idea behind all these techniques is the same, *i.e.,* given a population of individuals (known as chromosomes), the environmental pressure causes natural selection (survival of the fittest) thereby causing an increase in fitness of the population. EA's have been originally developed to deal with the data in binary form and were known as Genetic Algorithms. The ideas of mutation, recombination or crossover and selection are used to reach the solution to the problem. These strategies altogether come under EP and ES. Later, these ideas were also used to deal with real world data. EA's are more commonly known as Genetic Algorithms as each of the individual is considered a gene or chromosome.

Genetic Algorithm was originally developed by Holland in 1975 [22] which was later extended by Goldberg [17]. Details pertaining to GA's can be found in [22], [17], [42] and [43]. Man *et al.* [37] presented the use of Genetic Algorithms in the area of Control, Computational Intelligence, Speech Recognition and Communications. Jarvis and Goodacre [29] proposed the use of GA's for variable selection of spectroscopic data. The use of Genetic Algorithms for matrix factorization was first made by Snášel *et al.* [49]. They implemented GA for factorizing high

dimensional matrices containing binary data so as to achieve dimensionality reduction for analysis. The next section presents some important terminology used in Genetic Algorithms.

### 4.2.1 Terminology of Genetic Algorithms

In this section, Genetic Algorithms are discussed in detail. The components and operators used in GA's are:

- **Representation:** The first and the foremost step when using GA's is to link GA with that of real world. The objects that form the possible solutions in the original problem context are known as *phenotypes* and their encoding or representation in GA context is known as *genotypes* or *chromosomes*. There are different encoding methods such as Binary, Permutation, Value and Simulated Binary. The type of encoding selected is based on the problem.
- **Population:** The role of population is to represent a set of possible solutions to a given problem. A complete set of population forms a unit of evolution. The best chromosomes are selected in each generation of GA using the selection criteria and seeded for the next generation. The initialization of population could be random or may be seeded with potentially good solutions [50]. The population size should be adequate as low population size causes the GA to converge prematurely and large population size increases the process time.
- **Fitness Function:** It is also called as *Evaluation function* or *Objective function*. It basically represents the requirements to reach the solution. It forms the basis for selection of chromosomes. Each of the chromosomes is assigned a fitness value with which it is identified. The chromosomes with higher fitness value are

selected, and can proceed to the next generation and form the initial population (parent population).

- **Crossover:** The crossover operator is also known as *recombination*. This is the most important operator which causes the variation in the population. As the name indicates, this operator merges information from two parent chromosomes to form two child chromosomes. It is a stochastic operator to which probability is assigned. Hence the chromosomes are selected randomly according to the assigned probability. The crossover may result in the offspring which may be better or worse than parent chromosomes. The worse solutions are discarded using selection criteria. The crossover used for binary data is One Point Crossover whereas for real valued data Simulated Binary Crossover (SBX) is used.
- **Mutation:** It is a unary variation operator which is applied on a chromosome to produce a slightly modified child or offspring. Mutation operator is also stochastic and it selects chromosomes randomly based on the assigned probability to produce offspring of different kind. Generating a child amounts to stepping to a new point in the solution space. It is due to this operator that variations in the chromosomes are created such that the whole of the solution space is searched so as to obtain global minima.
- **Termination Condition:** The generational cycle of GA is continued until the termination condition is met. Some of the termination conditions commonly used are:
  (i) A solution is found which satisfies minimum criteria.
  (ii) Fixed number of generations reached.

(iii) The highest ranking solution is obtained, wherein the further increase in generations no longer produce better results.

### 4.2.2 Selection methods in GA

The role of survival selection or environmental selection is to differentiate among individuals based on their fitness value or quality. The selection procedure is carried out in two stages in GA. One is for the selection of parents based on their fitness value to reproduce better offspring whereas the other is for the selection of individuals among the offspring based on their fitness value to make them proceed to next generation for mating. As the population size at any point of time is constant in GA, the selection of best individuals among the offspring replace other chromosomes having a lesser quality in the previous pool of parent population. It can be inferred that the selection of chromosomes is based on the survival of fittest phenomenon. The most common methods for selection are:

- **Roulette Wheel Selection:** Each of the chromosomes is assigned a fitness value and is represented in form of roulette wheel like pie charts. The chromosomes with higher fitness value are assigned larger space in the roulette wheel when compared to less fit chromosomes. The roulette wheel is spun and chances of chromosomes with higher fitness values occupying larger area on the wheel to get selected are more.
- **Tournament Selection:** The selection in GA is based on the selection pressure. In tournament selection, this selection pressure is created among individuals by holding a tournament among *s* competitors. The winner of this tournament will have higher fitness value which is then forwarded to mating pool. Since, the

mating pool comprises of all the winners from tournaments, all the succeeding generation chromosomes will have better fitness values. Increased selection pressure can be provided by increasing the tournament size, as the winner in a large sized tournament is better when compared to a winner from a small sized tournament.

- **Rank Selection:** Each chromosome in the population is assigned a rank according to their fitness value. This type of selection method has advantage over roulette wheel selection as in this the fittest chromosome is also assigned a rank rather than a major portion in roulette wheel. The allotment of major area to the fittest individual in roulette wheel selection hinders the possibility of selection of other less fit chromosomes thereby leading to reduced genetic diversity.
- **Elitism:** In this type of selection, the best chromosome (some of the best chromosomes) is copied in the mating pool for the next generation and same number of less fit chromosomes is deleted. In this way, the fit chromosomes are not lost when performing crossover and mutation. Elitism drastically increases the performance of GA as it prevents the loss of best found solutions.

### 4.2.3 Working of GA

A simple pseudo code illustrating the working of GA is shown on the next page.

Define Objective Function

Choose initial population

Evaluate the fitness of each individual according to the objective function

Repeat

    Select individuals with best fitness value to reproduce

    Apply Crossover operator (with crossover probability)

    Apply Mutation operator (with mutation probability)

    Evaluate the fitness of each individual offspring

    Replace the least fit individuals with best ones

until termination criteria is met

Figure 4.1: Pseudo-code for Genetic Algorithm Implementation

## 4.3 Proposed Plant-wide Oscillation Detection using GA

The proposed method of detecting plant-wide oscillations using GA utilizes only routine operating data i.e. *SP, PV* and *OP* for analysis. The method works on the decomposition of spectral data matrix into basis vectors and their corresponding weights. Matrix decomposition using GA was firstly done by Snášel *et al.* [49] for the approximate reconstruction of a matrix containing binary data. In our case, the data matrix $X$ consists of real valued entries (spectral data) that are non-negative, which is the inherent quality of power spectra. The decomposed matrices also possess non-negative

entries to make sense of the basis shapes and their corresponding weights. The data matrix $X$ of the order $N \times M$ where $(M \ll N)$, is decomposed using GA as:

$$X_{N \times M} = W_{N \times r} H_{r \times M} + E \quad \ldots (4.1)$$

Each column of $X$ represents power spectra for individual tags. The matrix $W$ of size $N \times r$ represents the basis vectors or basis spectral shapes, where $r$ $(r \ll N \ \& \ r < M)$ is the reduced dimension for the approximation of original power spectra matrix. The value for the reduced dimension ($r$) is obtained using pseudo singular values (PSV) described further in this section. The matrix $H$ of size $r \times M$ represents the basis weights. The matrix $E$ represents the residual error of reconstruction of original spectral matrix. Therefore the original power spectral matrix $X$ can be approximated as:

$$X_{i\mu} \approx \sum_{k=1}^{r} W_{ik} H_{k\mu} \quad \ldots (4.2)$$

This condition imposed on each of the elements of decomposed matrices $W$ and $H$ can be described as $\{w_{ik}\}, \{h_{kj}\} \geq 0$. The purpose of using GA for decomposition of power spectra matrix $X$ is to obtain matrices $W$ and $H$ of reduced dimension that approximately reconstruct the original matrix. Implementation of GA requires the initial population which can be random or may be seeded with potentially good solutions [49]. The initial population of matrices $W$ and $H$ are thus created using Singular value decomposition which is further made random by multiplying each of the basis shapes and weights with different random entries. The matrices $W$ and $H$ at initialization are of full dimension *i.e.* $r = M$, which is finally reduced based on PSV. The objective function used in GA for spectral decomposition is based on Euclidean distance minimization between the original matrix $X$ and the estimated product of matrices $W$ and $H$ represented as:

$$\left\|X - WH\right\|^2 = \sum_{ij}\left|X_{ij} - \left(WH\right)_{ij}\right|^2 \quad \ldots (4.3)$$

In other words, the objective function is to minimize the reconstruction error *e*, between the original matrix *X* and product of matrices *W* and *H* represented as:

$$\left\|X - WH\right\|^2 = e \quad \ldots (4.4)$$

Based on the initial population created from singular value decomposition of original matrix, and objective function presented in Eq. (4.4), the Genetic Algorithm is proposed as:

---

Step 1: Create initial population of $N$ $\left(WH\right)$ chromosomes

Step 2: Evaluate the fitness of each chromosome according to the objective function (4.4)

Step 3: Evolve population

- Select suitable chromosomes (parent chromosomes) for reproduction
- Apply crossover on selected parent chromosomes w.r.t. crossover probability
- Apply mutation on selected parent chromosomes w.r.t. mutation probability
- Migrate offspring chromosomes to the population (parents) for next generation

Step 4: Evaluate current population based on the objective function

Step 5: Stop if termination condition is achieved else return to Step 3

---

Figure 4.2: Pseudo-code for the GA used in plant-wide oscillation detection

The flowchart for the pseudo code in Figure 4.2 is illustrated in Figure 4.3. The selection procedure used in the proposed GA is based on tournament selection described in section 4.2.2.

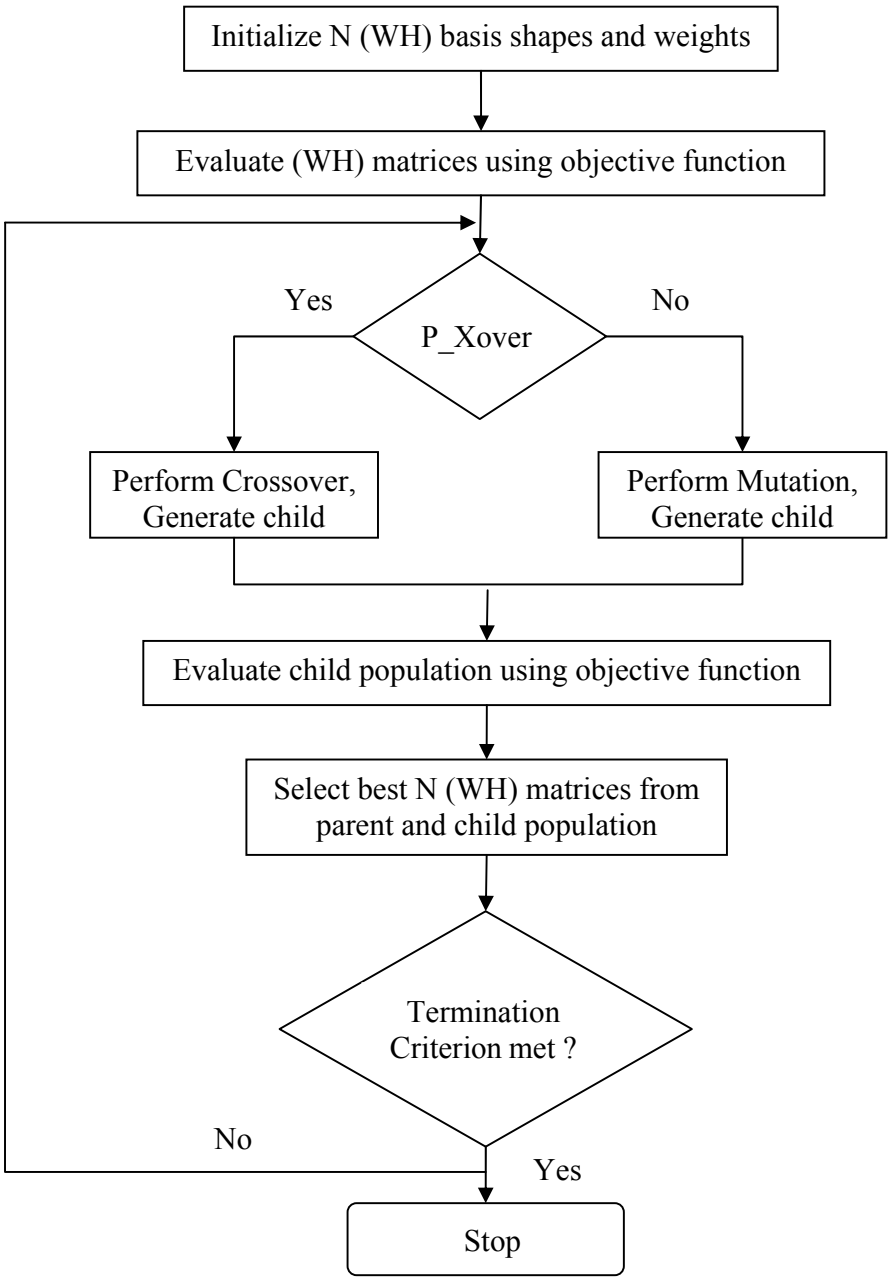



Figure 4.3: Flow chart for GA based factorization used in plant-wide oscillation detection

Simulated Binary Crossover (SBX) is used here which can handle real valued data entries with a crossover probability of 0.9. The mutation probability is set to 0.1. The concepts of total power plot, pseudo singular values and strength factors developed by Tangirala *et al.* [52] are used here in combination with GA based factorization of spectral data matrix for the detection of plant-wide oscillations.

The power spectra of individual tags are added at their respective frequencies to obtain total power as presented in Eq. (4.5). Total power plot shows all the oscillatory frequencies present in a plant.

$$X_T(f_l) = \sum_{j=1}^{M} X_j(f_l) \qquad \ldots (4.5)$$

where $X_j$ is the power spectrum associated with measurement $j$.

The decomposition of power spectral data matrix $X$ into the product of matrices $W$ and $H$ can be interpreted as decomposition of total power spectra $X_T$ into spectra like power components represented as:

$$C_k = \sum_{j=1}^{M} W_k H_{kj} \qquad (\text{for } k = 1,...,r) \qquad \ldots (4.6)$$

where $C_k$ is the spectrum like power component due to $k^{\text{th}}$ basis shape, $W_k$ is the $k^{\text{th}}$ column of $W$, and $H_{kj}$ is the $k^{\text{th}}$ element of $j^{\text{th}}$ variable corresponding to $W_k$.

The *pseudo singular values* (PSV) are used for the determination of basis size. The PSV is based on the incremental variance captured by each power component relative to the total power. The PSV for $j$th basis shape can be represented as:

$$\rho_j = \begin{cases} \dfrac{\left\| \sum_{k=1}^{j} C_k \right\|_2^2 - \left\| \sum_{k=1}^{j-1} C_k \right\|_2^2}{\left\| X_T \right\|_2^2} & \left(\text{for } j = 2,...,M\right) \\ \dfrac{\left\| C_1 \right\|_2^2}{\left\| X_T \right\|_2^2} & \left(\text{for } j = 1\right) \end{cases} \quad \ldots (4.7)$$

where $C_k$ is the $k^{\text{th}}$ power component and $X_T$ is the total power. The PSV is in the range of [0 1] with a value closer to 1 indicating strong influence of the corresponding basis shape. The PSV's for all the variables are calculated at once in the order power components are obtained and then rearranged in the decreasing order. If the first $r$ power components are able to represent total power then the remaining $M - r$ power components are set equal to zero.

The inability of obtained basis weights to indicate the true strength of each power component in presence of noise leads to the development of a novel measure known as '*Strength Factor*' (SF). The SF can be interpreted as the decomposition of the individual spectra into its constituent components as shown in Eq. (4.8). The Strength Factors are calculated after reordering basis shapes and corresponding weights in the order PSV's are rearranged. The Strength Factor of $k^{\text{th}}$ power component is given as:

$$SF_{kj} = \begin{cases} \dfrac{\left\| \sum_{p=1}^{k} W_p H_{pj} \right\|_2^2 - \left\| \sum_{p=1}^{k-1} W_p H_{pj} \right\|_2^2}{\left\| X_j \right\|_2^2} & \left(\text{for } k = 2,...,M \; ; \forall j\right) \\ \dfrac{\left\| W_1 H_{1j} \right\|_2^2}{\left\| X_j \right\|_2^2} & \left(\text{for } k = 1 \; ; \forall j\right) \end{cases} \quad \ldots (4.8)$$

SF can be analogously represented as PSV. SF lies within the range of [0 1] with a value of 1 indicating that $k$th power component entirely captures the spectral behavior of $j^{\text{th}}$

measurement. The summation of SF's of *r* reduced components is less than or equal to unity i.e. $\sum_{k=1}^{r} SF_{lj} \leq 1 \ \forall \ j$ .

## 4.4 Case Studies

This section presents a simulation example and two industrial case studies described earlier in chapter 3 to illustrate the efficiency of proposed method.

### 4.4.1 Simulation Example

The example consists of four signals $x_1, x_2, x_3$ and $x_4$ . Each of the signals is generated as follows:

$$x_1[n] = 0.6\sin(2\pi f_1 n) + d(n)$$

$$x_2[n] = 0.4\sin(2\pi f_2 n) + d(n)$$

$$x_3[n] = d(n)$$

$$x_4[n] = 0.3\sin(2\pi f_1 n) + 0.2\sin(2\pi f_2 n) + d(n)$$

where $f_1 = 0.004$ and $f_2 = 0.03$ . $d(n)$ is the white noise sequence with unit variance. The power spectra of each measurement is found and stored in a matrix as:

$$X = [X_1 \ \ X_2 \ \ X_3 \ \ X_4]$$

The time trends and power spectra of these four signals are shown in Figure 4.4. The data matrix *X* described above is to be decomposed to find the root cause of oscillations. The maximum number of generations and population size in GA are chosen as 300 and 200

respectively. The pseudo singular values suggest that there are only two power components leading to oscillations. The total power and its decomposition using GA based factorization are as shown in Figure 4.5.

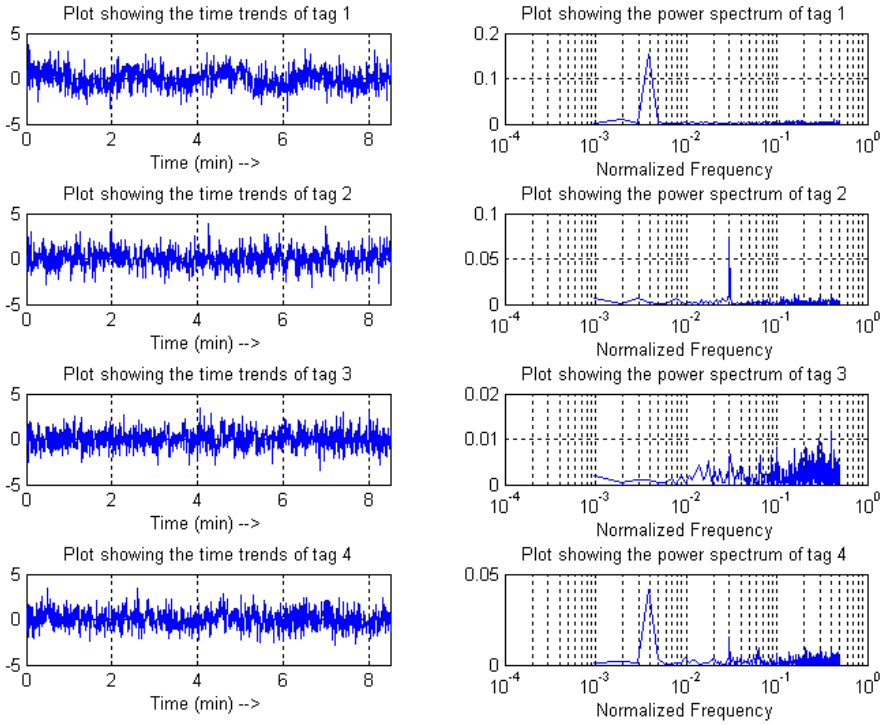



Figure 4.4: Time trends and power spectra of four measurement signals

The condensed plot of strength factors and error minimization between the original matrix and its reconstruction using GA are shown in Figure 4.6 and Figure 4.7 respectively. The condensed plot rightly shows the presence of oscillations in individual measurements. The first and second measurements completely capture first and second power components respectively. The third measurement consists of only noise and hence neither component 1 nor component 2 is present in it. The fourth measurement consists of both the power components and is clearly seen in Figure 4.6.

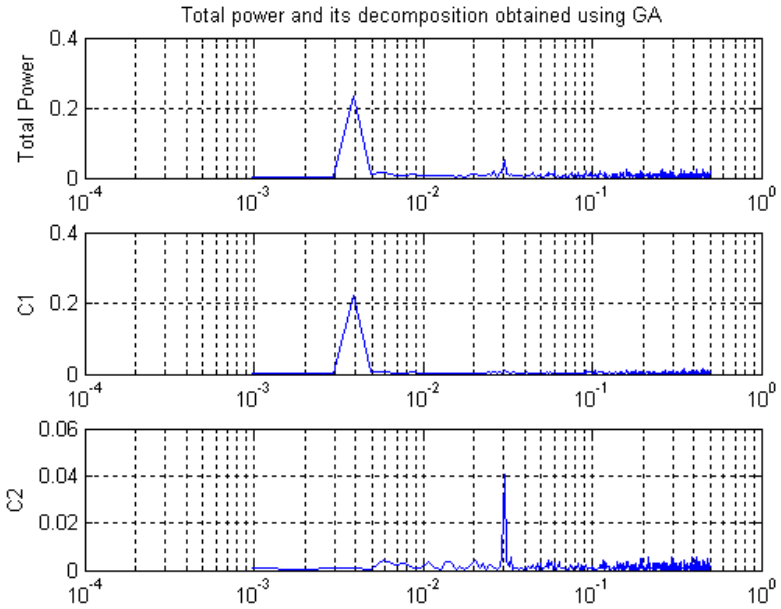



Figure 4.5: Total power and its decomposition using GA based factorization

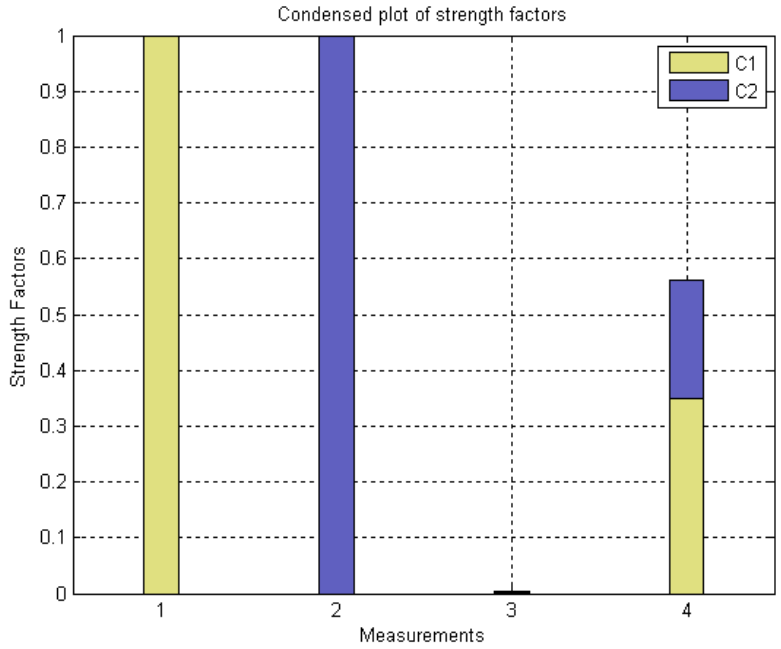



Figure 4.6: Condensed plot showing the root cause behind each of the power components for four measurement signal

The termination criterion for GA is selected as maximum number of generations. A steep decrease in reconstruction error with increase in number of generations is clearly seen in Figure 4.7. This proves the efficiency of GA to globally search the solution space in order to find the best solution.

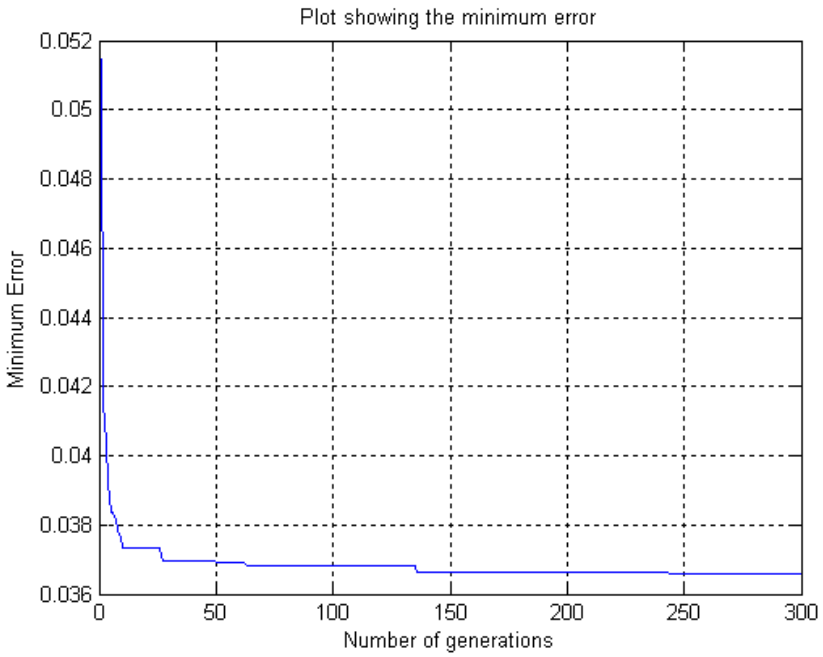



Figure 4.7: Reconstruction error minimization with increase in number of generations

### 4.4.2 Industrial Case Study 1 – *Entech Challenge Problem*

The case study considered has been described in detail in section 3.3. The process is a Challenge problem from Entech Control Inc. wherein the hardwood and softwood pulps are mixed to get the desired composition in the mixture. The process consists of twelve controllers associated with each control loop. The schematic of the process has been shown in section 3.3. Control error signals have been considered in this case study for analysis. Power spectra of each of the control error signal has been found and stored

in a matrix known as power spectra matrix which is used for the analysis using GA. The data pre-processing steps consists of mean centering the control error signals and normalizing each of the power spectra such that the area under each curve us unity. The time trends and power spectra for the twelve measurements are shown in Figure 3.6.

The power spectra matrix has been decomposed into basis shapes and their corresponding weights using GA based factorization. The maximum number of generations and population size were considered to be 100 and 75 respectively. The initial population used in GA was taken as the variants of SVD of the power spectra matrix. Such a kind of initialization was found to be better than the random initialization. The selection in the implemented GA was based on Tournament Selection and the termination criteria were set to be maximum number of generations.

Total number of frequencies leading to plant-wide oscillations was found to be two using SPCA and ICA, whereas NMF method could detect the presence of three oscillatory frequencies. GA based factorization also detected three frequencies as the cause of oscillations in the whole plant. This detection was based on the estimated pseudo singular values. The three oscillatory frequencies were found to be 0.002 $min^{-1}$ (period of 500 min), 0.02 $min^{-1}$ (period of 50 min) and 0.004 $min^{-1}$ (period of 250 min). The total power and its decomposition into constituent power components is as shown in Figure 4.8. These power components form the basis shapes of the decomposed matrices. The estimated strength factors represent the weights corresponding to the basis shapes. Only three power components could approximately reconstruct the original power spectra matrix.

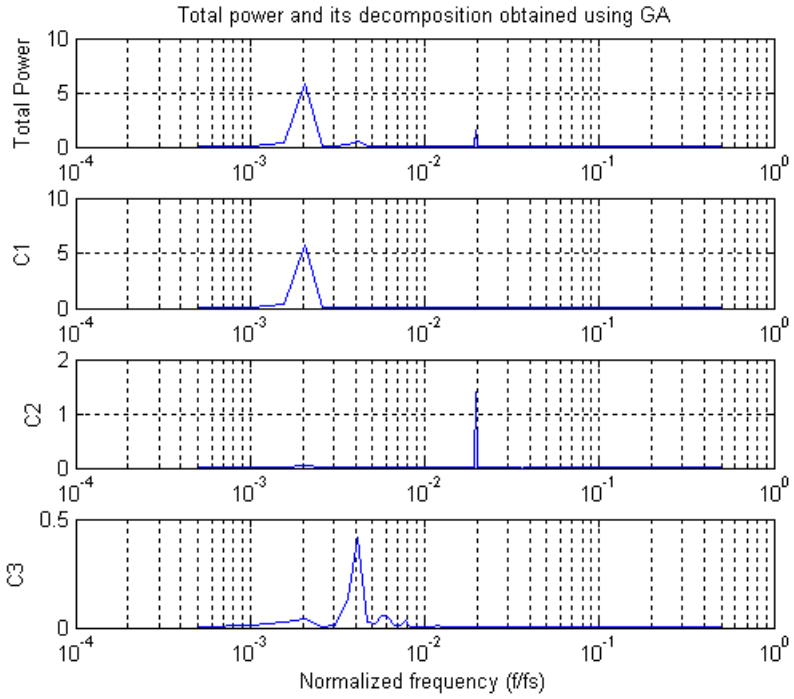



Figure 4.8: Total power plot and its decomposition using GA based factorization

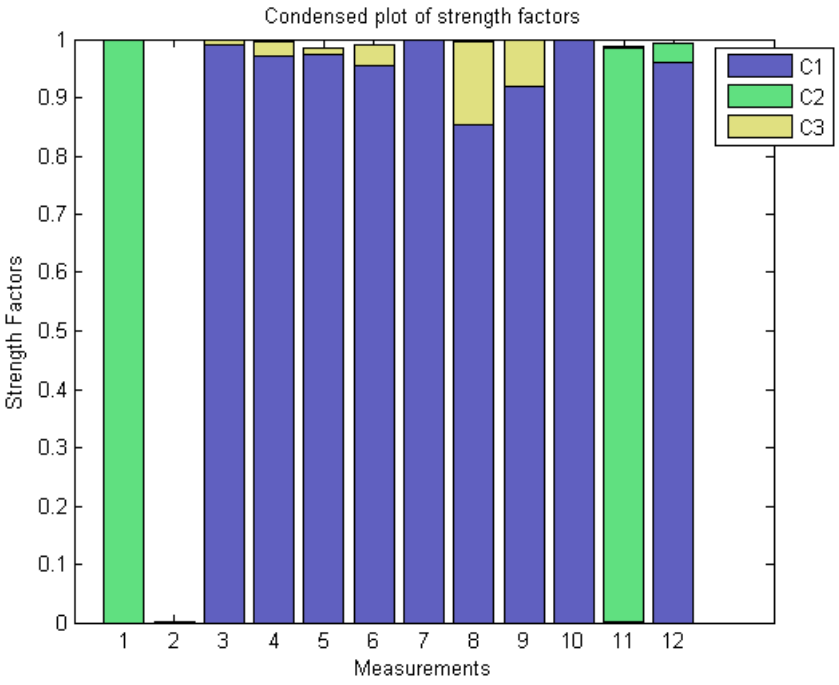



Figure 4.9: Condensed plot of strength factors for Entech data set

The strength factors relative to these power components are plotted in form of stacked bar chart as shown in Figure 4.9. This condensed plot of stacked bars illustrates the dominance of each of the three power components in individual measurements. The exact ranking of loops based on the presence of oscillatory frequencies is shown in Table 4.1. There is a difference in ranking of tags when compared to that of NMF method. This difference is due to the fact that NMF method can obtain only local minima whereas GA based factorization search for the global minima in the solution search space. The minimization of reconstruction error with increase in the number of maximum generations is illustrated in Figure 4.10.

| Component | Frequencies | Measurements (ranked in order) |
|---|---|---|
| 1 | 0.002 | 10, 7, 3, 5, 4, 12, 6, 9, 8, 2, 11, 1 |
| 2 | 0.02 | 1, 11, 12, 8, 4, 5, 6, 10, 9, 7, 3, 2 |
| 3 | 0.004 | 8, 9, 6, 4, 5, 3, 11, 12, 7, 10, 1, 2 |

Table 4.1: Ranking of measurements/tags of Entech data set using GA based factorization

The comparison of GA based factorization with ICA and NMF could only be made on the basis of reconstruction error *i.e.* how best the decomposed matrices $W$ and $H$ could reconstruct the original power spectra matrix $X$. The SPCA has not been considered for comparison due to the presence of negativity in their estimated basis shapes. The bar chart in Figure 4.11 shows norm 2 of error matrix $E$ obtained using each of the methods ICA, NMF and GA. The reconstruction error obtained using GA based factorization is drastically less when compared to that of NMF and ICA.

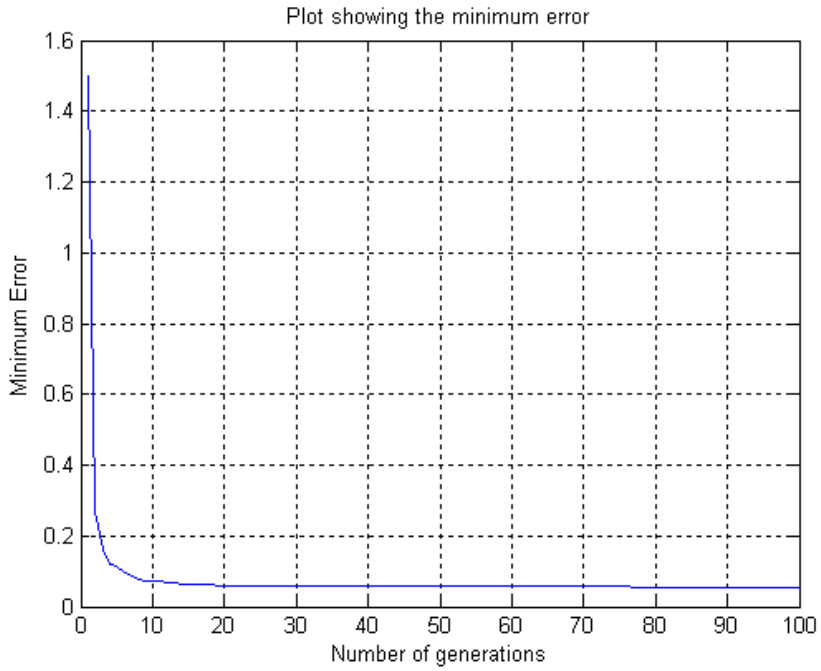



Figure 4.10: Reconstruction error minimization with increase in number of generations

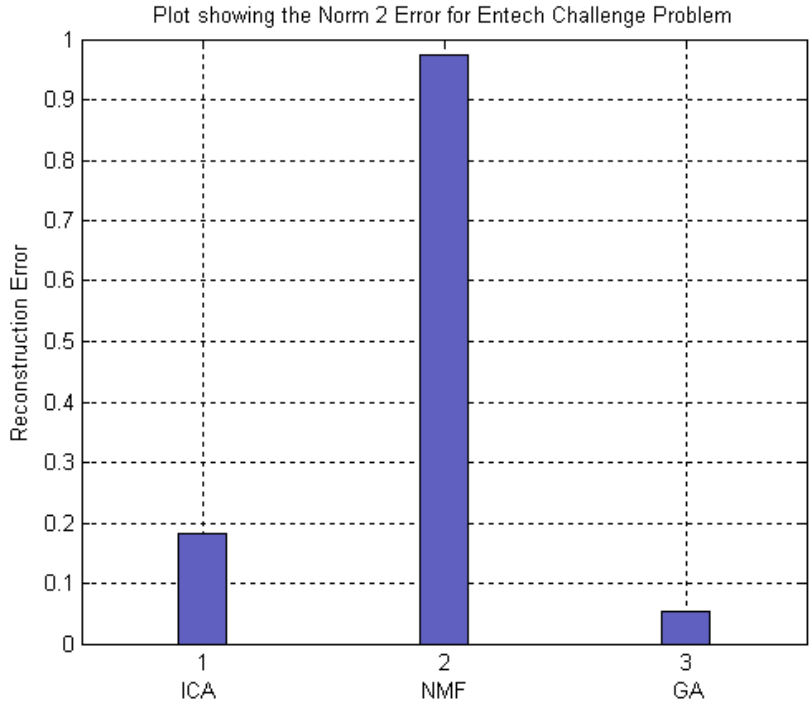



Figure 4.11: Norm 2 of reconstruction error matrix for Entech Challenge problem

The ICA method is also having considerably less reconstruction error compared to NMF due to its ability to reach global minima. This shows the efficiency of GA based factorization over ICA and NMF to deal with the problem of plant-wide oscillation detection using spectral decomposition technique.

### 4.4.3 Industrial Case Study 2 – *Paprican, Canada*

GA based factorization has been implemented on an industrial boiler process. The process has been described in detail in section 3.3.2 and the schematic of process is shown in Figure 3.9. There are eleven controllers associated with each of the control loop. The data used for the analysis of plant-wide oscillation detection is the control error signals. The data pre-processing steps consists of mean centering the control error signals to remove outliers and normalization of estimated power spectra such that the area under each curve is unity. Time trends and power spectra for these eleven control error signals has been illustrated in Figure 3.10.

GA based factorization is used for the decomposition of power spectral data matrix into its constituent basis shapes and their corresponding basis weights. The maximum number of generations and population size for the GA based factorization is set as 250 and 200 respectively. The initialization population of GA is taken as the random variants of SVD due to its improved performance. The selection of parents and healthy offsprings in GA is based on Tournament Selection. The termination criterion is set as the maximum number of generations.

Decomposition of spectral matrix using GA based factorization resulted in three power components. This selection is made, as the pseudo singular values beyond three

are insignificant and can be neglected. The total power and its decomposition using GA based factorization are as shown in Figure 4.12. There are three high peaks in the first power component which were also obtained in ICA and NMF. These peaks are present at the frequencies of 0.0074 $min^{-1}$, 0.1 $min^{-1}$ and 0.16 $min^{-1}$, wherein the peak at 0.1 $min^{-1}$ (period of 10 min) is highest. In second and third power components, peaks are obtained at the frequencies of 0.014 $min^{-1}$ (period of 71.42 min) and 0.12 $min^{-1}$ (period of 8.33 min) respectively. There is a change in frequency at which peaks occur using GA based factorization compared to that of NMF. This is due to the ability of GA to search globally, rather than getting stucked at local minima of solution search space as that of NMF.

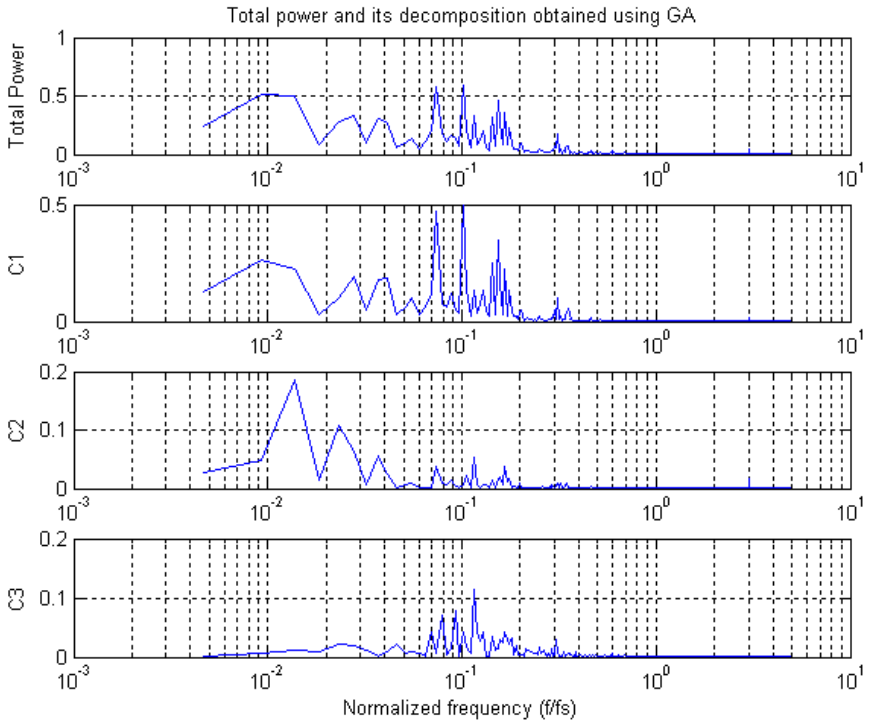



Figure 4.12: Total power and its decomposition using GA based factorization for Industrial Boiler data set

Thus three power components could capture maximum variance leading to plant-wide oscillations. The task of finding out the root cause loop is achieved using the strength factors. The strength factors represent the basis weights corresponding to three power components. The condensed plot of stacked bars illustrating the strength factors in each measurement is shown in Figure 4.13. It can be seen that three power components could capture appreciable amount of variance. This was not detected using NMF as can be seen in Figure 3.25. The dominance of first power component can be seen in tags 2 and 10. The exact ranking of loops according to the dominance of each power component is presented in Table 4.2.

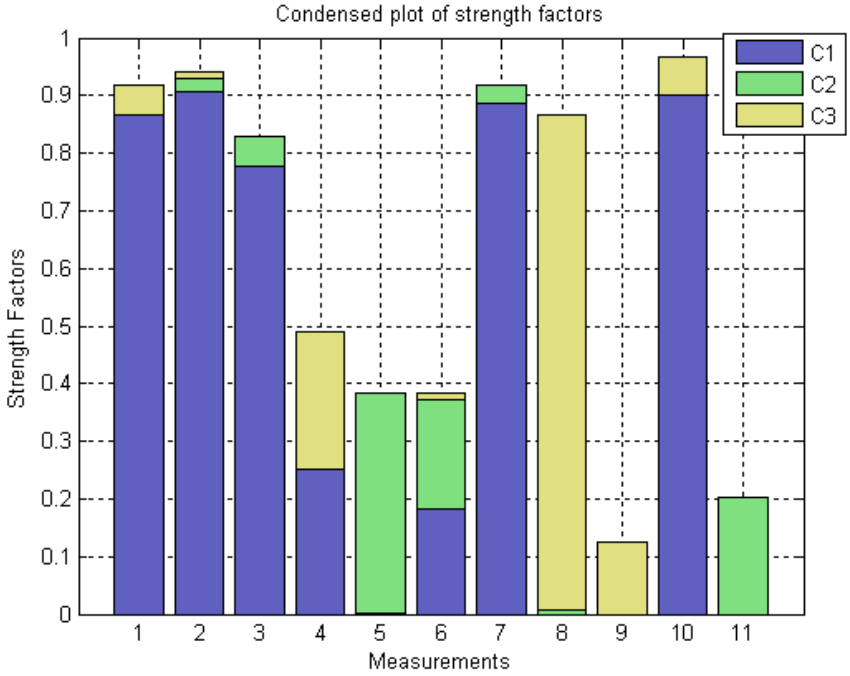



Figure 4.13: Condensed plot showing dominance of strength factors for Industrial Boiler data set

The frequency forming the second power component occurred as a band which was unclear using NMF, but GA based factorization rightly finds the exact frequency as 0.014

$min^{-1}$ and also finds the root loop affected by this frequency. The minimization of reconstruction error between the original power spectra matrix *X* and the decomposed matrices *W* & *H* is illustrated in Figure 4.14. This error is calculated using Norm 2 of the error matrix *E* which is the objective function in GA based factorization. There is always a decrease in error with increase in maximum number of generations. Thus choosing right number of initial population and maximum generations is important in GA based factorization.

| Component | Frequencies | Measurements (ranked in order) |
|---|---|---|
| 1 | 0.1 | 2, 10, 7, 1, 3, 4, 6, 5, 8, 9, 11 |
| 2 | 0.014 | 5, 11, 6, 3, 7, 2, 8, 1, 4, 9,.10 |
| 3 | 0.12 | 8, 4, 9, 10, 1, 6, 2, 3, 5, 7, 11 |

Table 4.2: Ranking of the loops for Industrial Boiler Process using GA based factorization

The comparison of GA based factorization with that of ICA and NMF for Industrial Boiler Process can be done on the basis of Norm 2 of error matrix *E*. This comparison is represented in form of a bar chart as shown in Figure 4.15. Similar to the previous case study, reconstruction error obtained using GA based factorization is extremely less when compared to that of ICA and NMF. The only reason behind increased reconstruction error using NMF is its inability to find global minima. Due to this reason, the error obtained using ICA is also less when compared to that of NMF. Proofs regarding ICA searching for global minima are present in the literature [65]. Despite the property of ICA reaching global minima, it still has a drawback of statistical independence of sources which is very difficult to achieve.

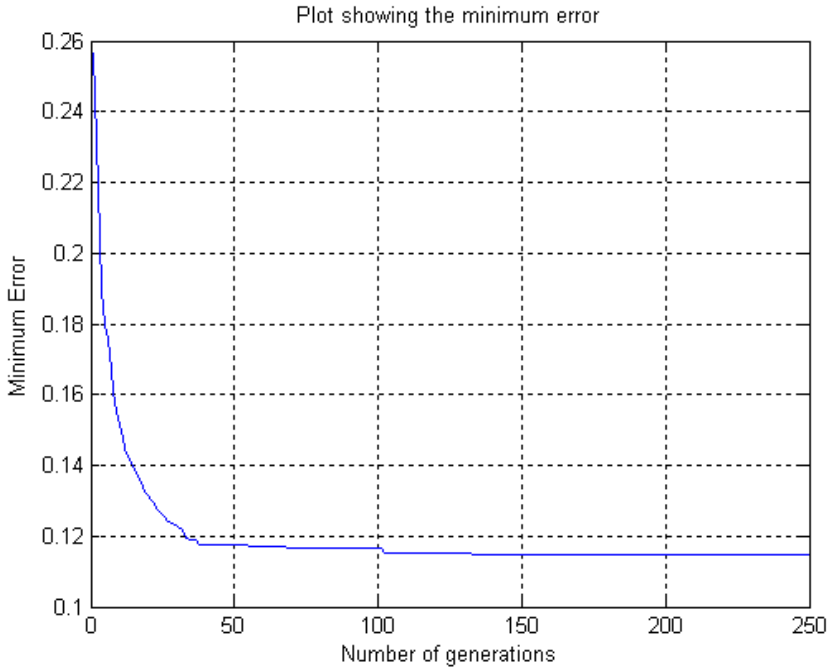



Figure 4.14: Reconstruction error minimization obtained for Industrial Boiler data set

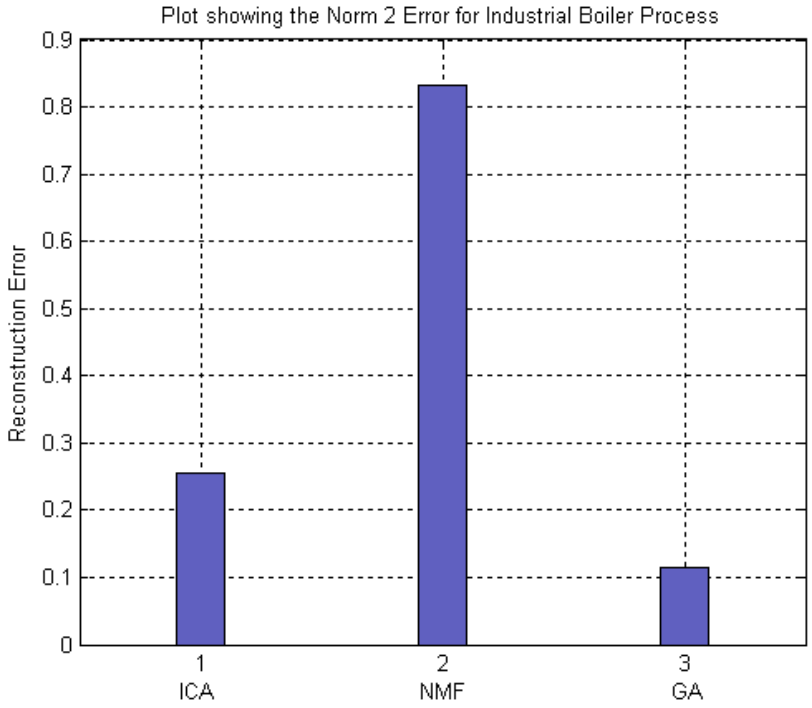



Figure 4.15: Norm 2 of reconstruction error matrix for Industrial Boiler data set

GA based factorization neither consider any statistical independence of sources as that of ICA nor gets stuck in the local minima as NMF. This proves the efficiency of GA based factorization over ICA and NMF.

## 4.5 Conclusion

To overcome the drawback associated with ICA and NMF, this chapter presents a new approach namely Genetic Algorithm (GA) based factorization to deal with the problem of plant-wide oscillations. Firstly, necessary background required for Evolutionary Algorithms is presented. This follows by the explanation of general working of GA's. Later, a novel GA algorithm is formulated to handle spectral decomposition task. GA based factorization has the ability to provide the global minimal solution. Indeed, a simulation example and two industrial case studies are presented to illustrate the efficiency of the proposed method. Performance comparison of GA based factorization with that of ICA and NMF is also presented highlighting the benefits of the new approach to provide better diagnosis results.

# CHAPTER 5

# DIAGNOSIS OF PLANT-WIDE OSCILLATIONS

## 5.1 Introduction

As stated earlier, oscillations in the measurements represent deterioration in the performance of a plant. The reason for the occurrence of oscillations could be either improper controller tuning, nonlinearities in control valve or external perturbations. The accurate detection of the root cause or the main problematic loops is achieved using GA based factorization method described in the previous chapter. Accurate root cause detection is achieved due to the ability of GA based factorization to search the solution space globally. Furthermore, the identification of the reason behind the poor performance of the loop detected as root cause is equally crucial. This chapter uses a method based on Higher Order Statistics (HOS) [6] to differentiate the oscillations occurring due to nonlinearities in valves from that of improper controller tuning and external disturbances. The efficiency of HOS approach to diagnose the root cause is illustrated using an example and two industrial case studies. The simulation example illustrates the efficiency of HOSA to deal with both strong and mild nonlinearities.

## 5.2 Higher Order Statistics (HOS)

Digital signal processing is an important area used for the analysis of data to identify hidden information in the measurements. It is rather important to convert the measurements from time domain to frequency domain to obtain complete hidden signal characteristics. This can be achieved by the use of a simple tool known as Discrete Fourier Transform (DFT). The linearity and Gaussianity of a signal can be conveniently studied using the first and second order statistics such as mean, variance, autocovariance, power spectrum, etc. In practice, there are many situations where the process deviates from linearity and Gaussianity [6]. The hidden knowledge of such processes which are nonlinear and non-Gaussian can be obtained using third order statistics and above. The statistics greater than second order is known as Higher Order Statistics.

The concept of Higher Order Statistics was originally developed by Rosenblatt and Van Ness [48]. Further advancements were made by Kim and Powers [33], Hinich [21], Rao and Gabr [47] and Nikias and Petropulu [44]. The most important of the higher order statistics consists of bispectrum and trispectrum which are frequency domain counterparts of third and fourth order moments or cumulants. The estimation of bispectrum is just an extension of power spectrum estimation. The values of bispectrum are not bounded, and hence, its normalization is termed as bicoherence. The estimation of bicoherence is mainly addressed in Collis *et al.* [8]. Choudhury *et al.* [6] used HOS for the analysis of non-Gaussianity and nonlinearity of a signal. They also developed two new indices termed as non-Gaussianity index (NGI) and nonlinearity index (NLI) for the diagnosis of plant-wide oscillations.

### 5.2.1 Bispectrum

The presence of nonlinearity in the measurements could be conveniently studied using bispectrum which is the frequency domain counterpart of third order moment or cumulant. The bispectrum is defined as

$$B(f_1,f_2) \triangleq DDFT\left[c_3(\tau_1,\tau_2)\right] \equiv E\left[X(f_1)X(f_2)X^*(f_1+f_2)\right] \quad \ldots (5.1)$$

where DDFT stands for double discrete Fourier transformation, $X(f)$ is DFT of the measurement signal and $^*$ represents complex conjugate. Eq. (5.1) illustrates that bispectrum is dependent on both frequencies $f_1$ and $f_2$ thereby having magnitude and phase. The bispectrum is visualized in a three dimensional figure against the frequencies $f_1$ and $f_2$. Similar to the power spectra which is redundant above Nyquist frequency, bispectrum is also redundant in twelve planes. So, in order to obtain the correct information conveyed by bispectrum, it is enough to visualize it in only a single plane of symmetry known as principal domain.

Figure 5.1 shows the triangular region (OAB) which is the non-redundant principal domain of bispectrum. This is sometimes divided into two triangles known as inner triangle (OAC) and outer triangle (ABC). The bispectrum $B(f_1,f_2)$ at point $(f_1,f_2)$ measures the interaction between the signal components at frequencies $f_1$ and $f_2$. This interaction can be related to nonlinearities present in the signal generating systems [15].

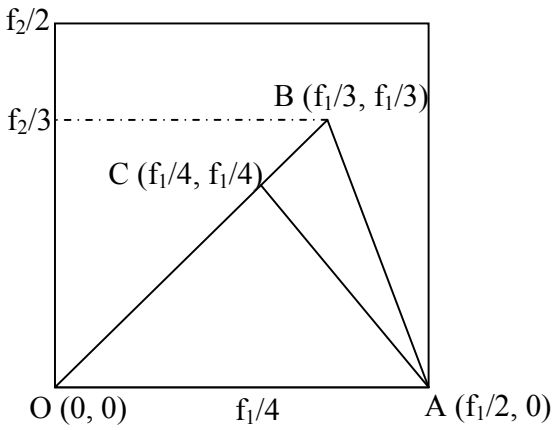



Figure 5.1: Principal domain of the bispectrum

### 5.2.2 Bicoherence

It has been shown in Eq. (5.1) that the estimation of bispectrum is dependent on first order spectral properties like DFT. This means that variance of the estimated bispectrum would be dependent on second order spectral property i.e. power spectrum and can be represented as

$$\mathrm{var}\left(\hat{B}\left(f_1,f_2\right)\right) \propto P\left(f_1\right)P\left(f_2\right)P\left(f_1+f_2\right) \qquad \ldots (5.2)$$

The variance of the estimated bispectrum will be uneven and higher where the spectral energy is high and low where the spectral energy is low. To overcome this issue, it has been suggested by Hinich [21] to normalize the bispectrum so as to make the bispectrum variance independent of the signal energies as shown in Eq. (5.3).

$$s^2\left(f_1,f_2\right) \triangleq \frac{\left|E\left[B\left(f_1,f_2\right)\right]\right|^2}{E\left[P\left(f_1\right)\right]E\left[P\left(f_2\right)\right]E\left[P\left(f_1+f_2\right)\right]} \qquad \ldots (5.3)$$

This normalization represented by $s^2(f_1,f_2)$ is known as skewness function. This function successfully removes the undesirable variance properties but also has a drawback associated with it. This drawback can be visualized as unboundedness of the magnitude of skewness function. This skewness function was modified by Choudhury *et al.* [6] and can be represented as

$$bic^2(f_1,f_2) \triangleq \frac{\left|B(f_1,f_2)\right|^2}{E\left[\left|X(f_1)X(f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]} \qquad \ldots (5.4)$$

where *bic* is the bicoherence function. The squared bicoherence represented in Eq. (5.4) is bounded in the range of [0, 1]. This has been proved in [6] using Cauchy – Schwartz inequality.

## 5.3 Estimation of Squared Bicoherence

The estimation of squared bicoherence in Eq. (5.4) involves the estimation of bispectrum and then normalizing it. The estimation can be done in two ways. The first approach involves the estimation of third order cumulants and then finding their DFT to convert them into frequency domain. This method is called as indirect method for bispectrum estimation. The second approach for estimation of bispectrum is based on segment averaging method, an extension for Welch's periodogram for spectral estimation, and is called as direct method for bispectrum estimation. Most commonly, direct method of bispectrum estimation is used.

The steps for the estimation of squared bicoherence are as follows:

(a) Divide the data series $x(k), k = 0,1,...,N-1$ into $K$ segments such that length of each segment is $M$. These segments can overlap to provide unbiased and consistent estimate. Thus $K \geq N/M$ and $i^{\text{th}}$ segment is $x_i(k)$, $k = 0,1,...,M-1$.

(b) Calculate the mean $\mu_i$ of each segment and subtract it from each sample of respective segments to make each segment mean centered at zero.

$$\mu_i = \frac{1}{M}\sum_{k=0}^{M-1} x_i(k) \qquad \ldots (5.5)$$

$$x_i'(k) = x_i(k) - \mu_i \qquad \ldots (5.6)$$

(c) Multiply each of the mean centered segment $x_i'$ with a window function $w(k)$. This is done to prevent any spectral leakage into adjacent frequency channels. It has been found that Hamming window gives the best results to resolve the spectral peak [15].

$$x_i''(k) = w(k)x_i'(k) \qquad \ldots (5.7)$$

(d) Compute the DFT for each segment $X_i(f)$

$$X_i(f) = \sum_{k=0}^{M-1} x_i''(k)e^{-j2\pi kf/M} \qquad \ldots (5.8)$$

where $f$ is the discrete frequency. The bispectrum for each segment is then estimated as

$$\hat{B}_i(f_1,f_2) = X_i(f_1)X_i(f_2)X_i^*(f_1+f_2) \qquad \ldots (5.9)$$

(e) The squared bicoherence function can be calculated as

$$bic^2(f_1,f_2) \triangleq \frac{\left|E\left[X(f_1)X(f_2)X^*(f_1+f_2)\right]\right|^2}{E\left[\left|X(f_1)X(f_2)\right|^2\right]E\left[\left|X(f_1+f_2)\right|^2\right]} \quad \ldots (5.10)$$

(f) The expectation operator is now removed and replaced with number of data segments

$$\hat{bic}^2(f_1,f_2) \triangleq \frac{\left|\frac{1}{M}\sum_{i=1}^{M}X_i(f_1)X_i(f_2)X_i^*(f_1+f_2)\right|^2}{\frac{1}{M}\sum_{i=1}^{M}\left|X_i(f_1)X_i(f_2)\right|^2\frac{1}{M}\sum_{i=1}^{M}\left|X_i(f_1+f_2)\right|^2} \quad \ldots (5.11)$$

Eq. (5.11) is the standard equation for the estimation of squared bicoherence but it has been found that the occurrence of small values in denominator leads to the presence of small spurious peaks in the bicoherence plot. To overcome this issue, it has been suggested in [5] to add a dynamic constant $\varepsilon$ to the denominator of Eq. (5.11).

$$\hat{bic}^2(f_1,f_2) \triangleq \frac{\left|\frac{1}{M}\sum_{i=1}^{M}X_i(f_1)X_i(f_2)X_i^*(f_1+f_2)\right|^2}{\frac{1}{M}\sum_{i=1}^{M}\left|X_i(f_1)X_i(f_2)\right|^2\frac{1}{M}\sum_{i=1}^{M}\left|X_i(f_1+f_2)\right|^2 + \varepsilon} \quad \ldots (5.12)$$

The value of $\varepsilon$ is obtained automatically by choosing it as the maximum of $P^{\text{th}}$ percentile of the calculated values of the denominator in Eq. (5.11). If it has been assumed that 25 % of the values in the denominator of Eq. (5.11) consists of peaks, then $P$ can be chosen as 75$^{\text{th}}$ percentile [5]. Thus, the squared bicoherence is estimated in this work by using Eq. (5.12). In the next section, two new indices non-Gaussianity Index (NGI) and nonlinearity index (NLI), developed by Choudhury *et al.* [5] are described to test the Gaussianity and linearity of a time series.

## 5.4 Test of Gaussianity and Linearity of a signal

The examination of non-Gaussianity and nonlinearity of a time series serves as a helpful diagnostic tool to study its performance. The earlier proposed methods used Sample interquartile range [21] and Hotelling-$T^2$ test [47] of the skewness function to test signal's linearity. Choudhury *et al.* (2004b) [6], proposed the use of bicoherence based chi-square test to study signal's nonlinearity. They proposed this statistical test in two indices known as Non-Gaussianity index (NGI) and nonlinearity index (NLI).

The squared bicoherence for a linear signal is independent of frequencies $f_1$ and $f_2$ and could be either zero or a non-zero constant. If the squared bicoherence is zero, then the signal is Gaussian and the process is linear whereas, if the squared bicoherence is a non-zero constant, then the signal is non-Gaussian and the process is linear [6]. The statistical chi-square based hypothesis test incorporates the use of average squared bicoherence over the principal domain to examine the importance of bicoherence magnitude at each individual bifrequency. The statistical test can be formulated as

$$P\left\{bic^2(f_1,f_2) > \frac{c_\alpha^{\chi^2}}{2K}\right\} = \alpha \qquad \ldots (5.13)$$

where $c_\alpha^{\chi^2}$ is the critical value obtained from central chi-square distribution table for a critical value of $\alpha = 0.05$ with two degrees of freedom and $K$ is the number of segments used for bicoherence estimation. Here the degrees of freedom are taken as two since the squared bicoherence also depends on two frequencies $f_1$ and $f_2$.

The proposed non-Gaussianity index (NGI) to test the signal's Gaussianity can be defined as [5]

$$\text{NGI} \triangleq \frac{\sum bic^2_{significant}}{L} - \frac{c_\alpha^{\chi^2}}{2KL} \qquad \ldots (5.14)$$

where $bic^2_{significant}$ are the squared bicoherence values which satisfy Eq. (5.13) and $L$ is the number of $bic^2_{significant}$ values. In a simplified way, the values that satisfy Eq. (5.13) can be evaluated using $bic^2(f_1, f_2) > \frac{c_\alpha^{\chi^2}}{2K}$. In this way, the signal's Gaussianity is tested at a confidence interval of $\alpha$. The values of NGI are in the range of [-1, 1]. If the values of NGI ≤ 0, then it is concluded that the signal is Gaussian whereas if NGI > 0, it is concluded that the signal is non-Gaussian.

If the signal is found to be non-Gaussian, then it is tested for linearity. As described earlier, if the signal is non-Gaussian and linear then the magnitude of squared bicoherence will be a non-zero constant. This means that the variance of the squared bicoherence will be zero. To test the signal's nonlinearity, the proposed index can be defined as

$$\text{NLI} \triangleq \hat{bic}^2_{\max} - \left(\overline{\hat{bic}^2_{robust}} + 2\sigma_{\hat{bic}^2,robust}\right) \qquad \ldots (5.15)$$

where $\hat{bic}^2_{\max}$ is the maximum value of estimated bicoherence, $\overline{\hat{bic}^2_{robust}}$ and $\sigma_{\hat{bic}^2,robust}$ are robust mean and robust standard deviation respectively. Here the robust values of squared bicoherence are extracted by removing the largest and smallest $Q$ % and the value of $Q$ is taken as 10. The robust mean and robust standard deviation values are calculated to get

rid of spurious peaks in the estimated bicoherence. Just as the NGI is in the range of $[-1, 1]$, NLI also lies in the same range. If the calculated NLI $\leq 0$, the signal is confirmed to be linear whereas if NLI > 0 the signal is confirmed to be nonlinear. In practice it is difficult to attain the critical value of zero, hence the critical values can be taken as low values depending upon the data length [7] as described in Table 5.1.

| Data Length | $NGI_{cric}$ | $NLI_{cric}$ |
|---|---|---|
| 4096 | 0.001 | 0.01 |
| 2048 | 0.002 | 0.02 |
| 1024 | 0.004 | 0.04 |

Table 5.1: Critical values for NGI and NLI

To compare the nonlinearity between different time series, a new index known as Total Nonlinearity Index (TNLI) is proposed in [5] and can be defined as

$$\text{TNLI} \triangleq \sum bic^2_{significant} \quad \ldots (5.16)$$

where $bic^2_{significant}$ are the squared bicoherence values which satisfy Eq. (5.13). The index TNLI is proposed, as the index NLI finds the nonlinearity only for the maximum bicoherence peak and not the overall nonlinear interactions among the frequencies. The value of TNLI is bounded between 0 and $L$ where $L$ is the number of $bic^2_{significant}$ values.

If the signal is found to be non-Gaussian and nonlinear using Eq's. (5.14) and (5.15), the probable cause of oscillations would be nonlinearity in the valve like stiction,

backlash, deadband, etc. If the signal is concluded to be linear, then the probable cause could be improper controller tuning or external disturbance. The method for testing the non-Gaussianity and nonlinearity of the signal can be summarized in the form of a flowchart shown in Figure 5.2, [7].

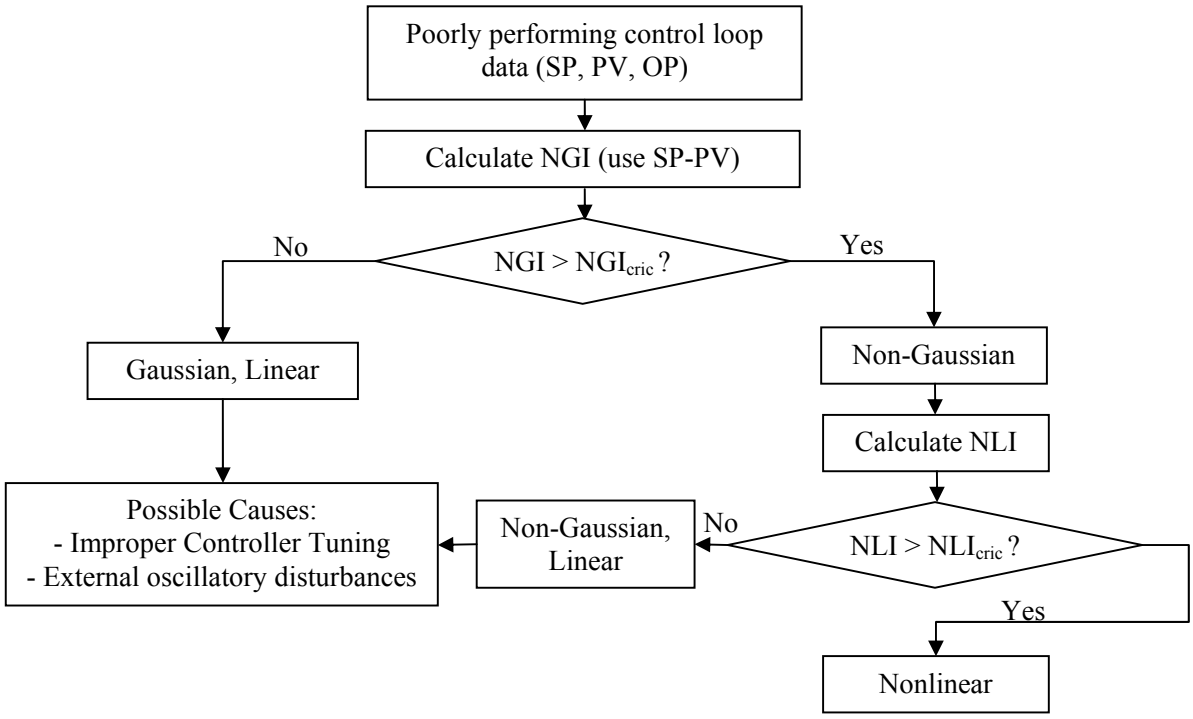



Figure 5.2: Flow diagram for the detection of loop nonlinearity

## 5.5 Case Studies

In this section, the efficiency of squared bicoherence to detect nonlinearities present in a signal is illustrated using an example and two industrial case studies.

### 5.5.1 Example: Bicoherence of a Nonlinear Sinusoid Signal with Noise [5]

The input signals $x$ & $y$ are created by adding two sinusoids having different frequency and phase. They can be represented as

$$
\begin{aligned}
x'(k) &= \sin(2\pi f_1 k + \phi_1) + \sin(2\pi f_2 k + \phi_2) \\
x(k) &= x'(k) + d(k) \\
y(k) &= x'(k) + n_l x'(k)^2 + d(k)
\end{aligned}
\qquad \ldots (5.17)
$$

where $f_1 = 0.12, f_2 = 0.30, \phi_1 = \pi/3, \phi_2 = \pi/8$, $n_l$ is factor that represents the amount of nonlinearity present and $d(k)$ is the white noise signal with variance 0.04. The frequencies are normalized such that $f_s = 1$. The presence of square term in the signal $y$ will introduce phase coupling. Six possible peaks at $(f_1, f_1), (f_2, f_2), (f_1, f_2)$, $(f_1, f_2 - f_1), (f_2 - f_1, f_1 + f_2)$ and $(f_2 - f_1, 2f_1)$ will be obtained due to nonlinear frequency couplings depending upon the contribution of nonlinear factor $n_l$.

***Mild Nonlinearity***

In this case, the amount of induced nonlinearity is small such that $n_l = 0.05$. The time trends, power spectra and squared bicoherence of $x$ and $y$ are shown in Figure 5.3. The visualization of time trends and power spectra in the left most and middle panels couldn't let us differentiate between the two signals. The three dimensional squared bicoherence plots in the right most panel clearly shows the nonlinearities present in signal $y$.

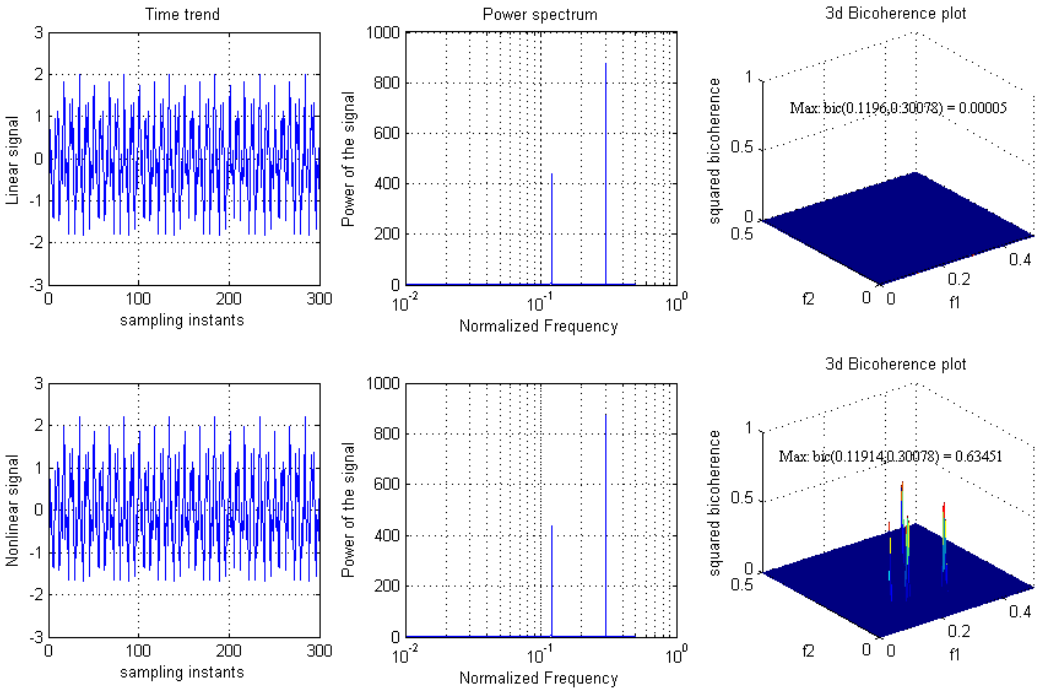



Figure 5.3: HOS analysis for the linear signal and nonlinear signal with mild nonlinearity

For the linear signal, the value obtained for NGI = 0 which means the signal is Gaussian and also linear. The last figure in first row of Figure 5.3 also shows the flat squared bicoherence which signifies the Gaussianity of the linear signal. On the contrary, the nonlinear signal is having NGI = 0.41 and NLI = 0.61 which clearly indicates that the signal is non-Gaussian and nonlinear. The squared bicoherence plot consists of four peaks at bifrequencies (0.12,0.12), (0.12,0.18), (0.30,0.30) and (0.12,0.30) corresponding to $(f_1, f_1), (f_1, f_2 - f_1), (f_2, f_2)$ and $(f_1, f_2)$. The other two peaks at $(f_2 - f_1, f_1 + f_2)$ and $(f_2 - f_1, 2f_1)$ are not visible due to their small magnitude for $n_l = 0.05$.

***Strong Nonlinearity***

The amount of nonlinearity induced in this case is 25 % i.e. $n_l = 0.25$. The time trends, power spectra and squared bicoherence of the linear signal and the nonlinear signal with strong nonlinearity are shown in Figure 5.4. As seen in the earlier case, the time trends of both signals couldn't let us differentiate between the two signals. The presence of three more small peaks in the power spectrum of nonlinear signal at frequencies 0.18, 0.24 and 0.42 indicates the strong presence of nonlinearity. The HOS analysis for nonlinear signal with strong nonlinearity possess NGI = 0.74 and NLI = 0.93 which clearly indicate the increase in nonlinearity.

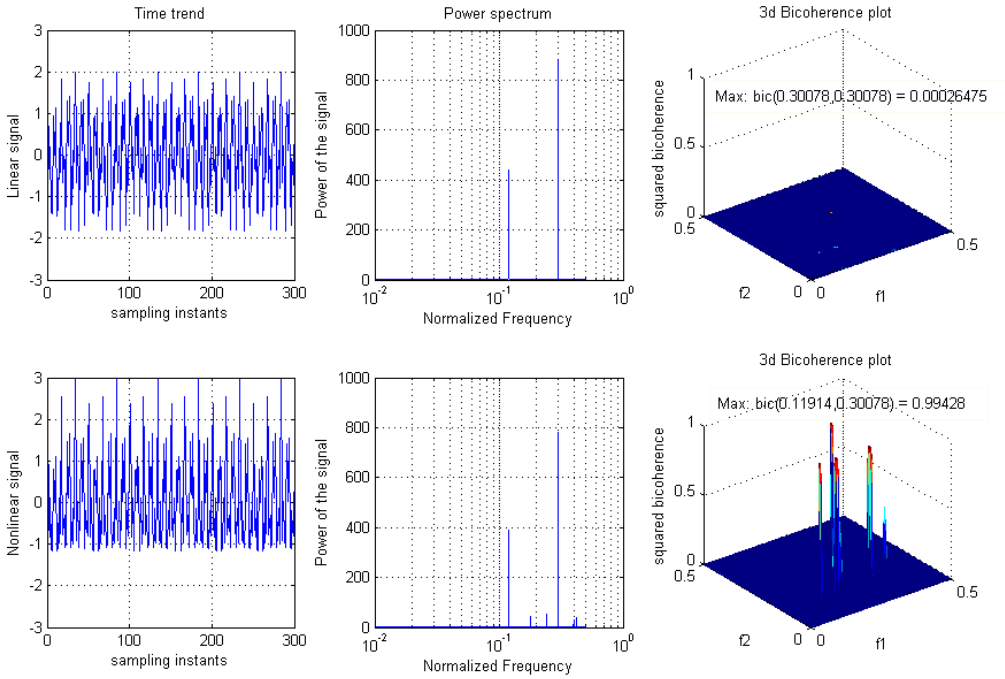


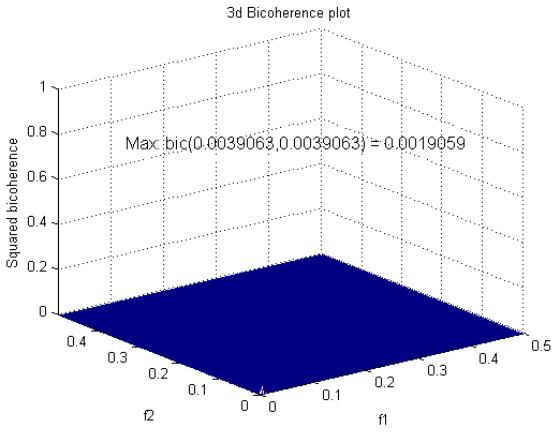

Figure 5.4: HOS analysis for the linear signal and nonlinear signal with strong nonlinearity

The squared bicoherence plot for the nonlinear signal with strong nonlinearity clearly illustrates the presence of all six peaks at (0.12,0.12), (0.12,0.18), (0.30,0.30), (0.12,0.30), (0.18,0.24) and (0.42,0.18) corresponding to bifrequencies $(f_1, f_1)$, $(f_1, f_2 - f_1), (f_2, f_2), (f_1, f_2), (f_2 - f_1, 2f_1)$ and $(f_1 + f_2, f_2 - f_1)$. In essence, it is clearly understood that the squared bicoherence could clearly indicate the interactions among different frequencies that leads to nonlinearity in a signal.

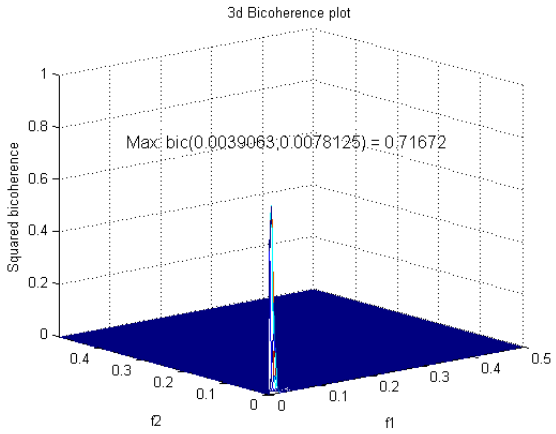
### 5.5.2 Industrial Case Study 1 – *Entech Challenge Problem*

The case study described in section 3.3.1 consists of a paper and pulp process wherein the hardwood and softwood are mixed to get the desired composition. The schematic of the process is shown in Figure 3.5. The time trends and power spectra of 12 tags have been illustrated in Figure 3.6. It was found that the three frequencies at 0.02 $min^{-1}$, 0.002 $min^{-1}$ and 0.004 $min^{-1}$ lead to plant-wide oscillations. The proposed GA based optimization method described in chapter 4 found the loops 10, 1 and 8 as root causes behind plant-wide oscillations associated with the frequencies 0.002 $min^{-1}$, 0.02 $min^{-1}$ and 0.004 $min^{-1}$ respectively. Now, we will diagnose these loops to find the reason behind their oscillations using higher order statistics based squared bicoherence.

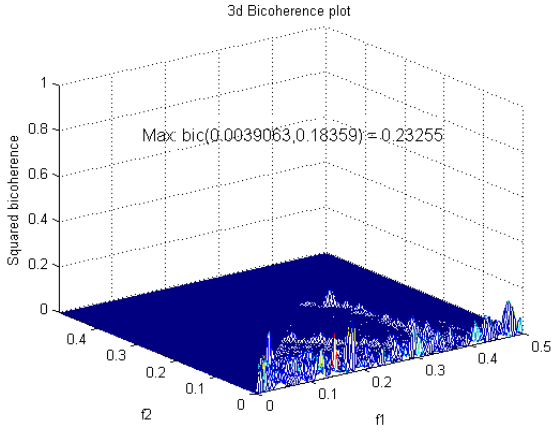
Figure 5.5 shows the squared bicoherence for the loops 1, 8 and 10 in order. The flow controller of loop 1, affected by the frequency 0.02 $min^{-1}$ has flat squared bicoherence as can be seen in Figure 5.5 (a).




(a) Squared bicoherence for loop 1




(b) Squared bicoherence for loop 8




(c) Squared bicoherence for loop 10

Figure 5.5: Higher order statistical analysis for Entech data set

The HOS analysis for loop 1 resulted in NGI = 0, which means the loop is linear and according to the flow chart proposed by Choudhury *et al.* (2006) [7], in Figure 5.2 the possible cause is improper controller tuning as has also been confirmed in [59]. Figure 5.5 (b) shows the squared bicoherence for loop 8 affected by the frequency 0.004 $min^{-1}$. The careful examination of power spectra of loop 8 in figure 3.6 shows that it is affected by both frequencies 0.002 $min^{-1}$ and 0.004 $min^{-1}$. The HOS analysis results shows that NGI = 0.4252 and NLI = 0.6974 which is also clearly seen as a sharp peak in Figure 5.5 (b). This shows that loop 8 fails to regulate the flow through it due to the presence of nonlinearity as has also been confirmed in [59] and [52]. The squared bicoherence for level controller identified as loop 10, affected by the frequency 0.002 $min^{-1}$ is as shown in Figure 5.5 (c). The values for NGI and NLI obtained using HOS analysis of loop 10 are 0.17 and 0.22 respectively. This shows that the loop is suffering from nonlinearity in the valve and probable cause could be stiction, backlash, etc. Thus, the HOS analysis for Entech data set correctly identifies the persisting problems in the loops identified as root cause.

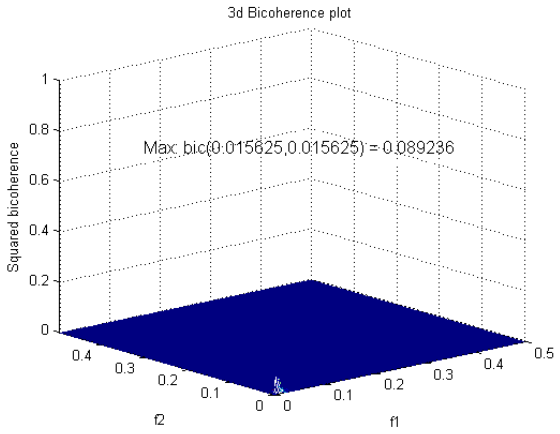
### 5.5.3 Industrial Case Study 2 – *Paprican, Canada*

The process described in section 3.3.2 is of a boiler process provided by *Paprican, Canada*. The process consists of 11 controllers associated with each control loop and schematic of the process is shown in Figure 3.9. The time trends and power spectra for each of the control loops have been illustrated in Figure 3.10. The analysis of this data set using proposed GA based plant-wide oscillation detection resulted in three frequencies.

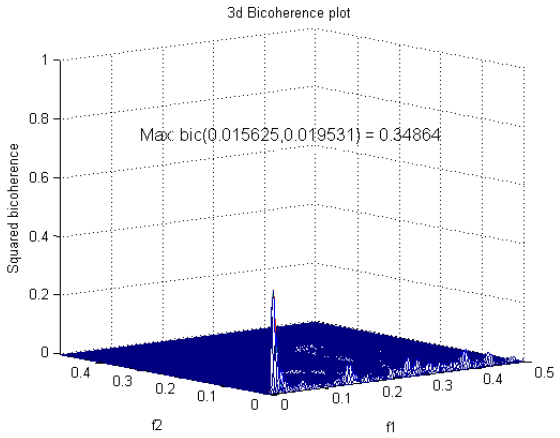
The loops identified as root causes of plant-wide oscillations are 2, 5 and 8 each associated with the frequency 0.1 $min^{-1}$, 0.014 $min^{-1}$ and 0.12 $min^{-1}$ respectively. The squared bicoherence for the three loops are shown in Figure 5.6. Loop 2 associated with the frequency 0.1 $min^{-1}$ is analyzed using HOS and it is found that NGI = 0. The flatness in the squared bicoherence plot of Figure 5.6 (a) also concludes the same. Thus, there is no nonlinearity associated with loop 2 and it is Gaussian and linear. The probable cause of oscillations associated with loop 2 with reference to Figure 5.2 could be either improper controller tuning or external disturbances.

The middle panel of Figure 5.6 shows the squared bicoherence for loop 5. The frequency associated with the oscillations caused by this loop was found to be $0.014\ min^{-1}$. A peak in the squared bicoherence plot signifies the presence of nonlinearity in the loop. The implementation of HOS analysis for loop 5 resulted in NGI = 0.2433 and NLI = 0.3378. This clearly shows the presence of nonlinearity in the valve. In Figure 5.6 (c), the squared bicoherence for loop 8 is shown. The plot shows the corruption of data with relatively high amount of noise. The frequency associated with loop 8 was found to be 0.12 $min^{-1}$. The HOS analysis resulted in the values of NGI and NLI as 0.2725 and 0.4257 respectively. The presence of nonlinearity even in this loop is confirmed by the analysis made using HOS.

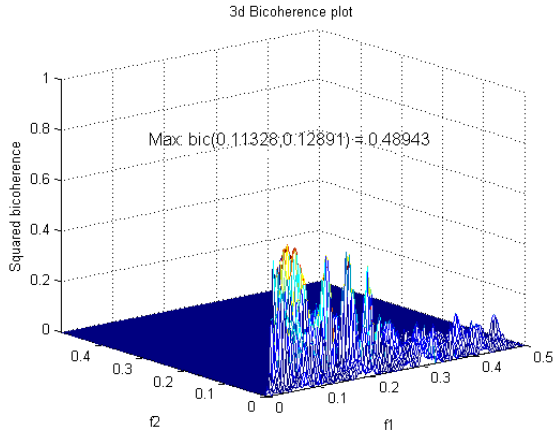
The diagnosis of the industrial boiler set using HOS found the nonlinearities in loops 5 and 8. Also, the diagnosis showed that loop 2 is linear and probable cause could be either improper controller tuning or external disturbances. The linearity of loop 2 and nonlinearity in loop 8 is also confirmed by the plant personnel.




(a) Squared bicoherence for loop 2




(b) Squared bicoherence for loop 5




(c) Squared bicoherence for loop 8

Figure 5.6: Higher order statistical analysis for Industrial Boiler data set

## 5.6 Conclusion

To complete the methodology of detection and diagnosis of plant-wide oscillations, the subject of diagnosis for the loops identified as root causes has been dealt within this chapter on the basis of Higher Order Statistical Analysis (HOSA). This chapter starts with introduction to HOS. This is followed by in detail explanation of two most important statistical quantities known as Bispectrum and Bicoherence which could help in the diagnosis of root cause. The drawback of bispectrum is found to be unboundedness of its magnitude. This problem is dealt in [6] by developing a measure known as bicoherence whose magnitude is bounded in the range of 0 to 1. Later, the procedure for estimation of squared bicoherence is explained with emphasis on the removal of spurious peaks. Two important measures known as non-Gaussianity index (NGI) and nonlinearity index (NLI) based on squared bicoherence are presented which were originally developed by Choudhury *et al.* (2008) [5] to measure the Gaussianity and linearity of loops identified as root cause. The chapter ends with the implementation of HOSA to an example and two industrial case studies presented earlier. In conclusion, HOSA was able to differentiate the presence of oscillations in loops due to nonlinearity in valve or improper controller tuning and external oscillatory disturbances.

# CHAPTER 6

# CONCLUSIONS AND FUTURE RESEARCH

## 6.1 Conclusion

The work presented in this thesis focuses upon performance monitoring of controllers in a plant. Precisely, the aim is to develop a non-invasive methodology which could detect and diagnose the occurrence of oscillations in the whole plant. The proposed methodology is completely data based and could work based on routine operating data. Existing methods in the literature to deal with the problem of plant-wide oscillations has been studied in detail via extensive simulations on examples and industrial case studies in order to understand their shortcomings. Of these methods, three are data based (SPCA, ICA and NMF) and one is based upon the process understanding (Control loop digraphs). The data based methods presented utilizes the concept of spectral decomposition into basis shapes and their respective basis weights.

Although Spectral Principal Component Analysis is regarded as the most common method to detect plant-wide oscillations, it suffers from a drawback of negativity in basis shapes. To overcome the negativity issue in basis shapes, ICA and NMF methods were proposed but also have drawbacks associated with them. The ICA method imposes the assumption of statistical independence of sources which is often very difficult to verify. NMF method doesn't impose any such assumption but has a problem of getting stucked

in the local minimal solution. In this thesis, a GA based optimization method is proposed to deal with the problem of local minimal solution by searching the solution space globally. The proposed GA based method can detect plant-wide oscillations efficiently and its performance is found to be significantly improved in terms of spectral reconstruction of the original matrix.

In order to complete the methodology, the loops identified as root causes of plant-wide oscillations are diagnosed further utilizing the concept of Higher Order Statistics developed earlier in the literature [6]. This can identify the persisting problem in the loops identified as root causes of plant-wide oscillations.

## 6.2 Recommendations for Future Research

Plant-wide oscillation is one of the important aspects of the Plant Performance Monitoring and Assessment. This is an emerging field in the current scenario which has been developed to a large extent but still some of the areas are to be explored. Following are some of the recommendations for future research:

(a) Alternative optimization techniques such as Simulated Annealing and Particle Swarm Optimization can also be explored to deal with the problem of plant-wide oscillation detection.

(b) The present work considers the dynamics of the plant to be linear and presence of nonlinearity in valve. Two more cases could be studied wherein the first case considers the nonlinear dynamics of a plant and nonlinearity in valves and second case considers linear dynamics of plant and linearity in valves.

(c) There is always an assumption of external disturbance to be linear, but there is a possibility of nonlinear disturbance entering from either upstream or downstream of the process. This work can be extended to study the effects of nonlinear external disturbances.

# REFERENCES


[1] Bialkowski, W. L. (1992). Dreams vs reality: A view from both sides of the gap. In: *Control Systems*. Whistler, BC, Canada, pp. 283 – 294.

[2] Chatfield, C., & Collins, A. J. (1980). *Introduction to multivariate analysis*. London, UK: Chapman and Hall.

[3] Choudhury, M. A. A. S. (2004). Detection and Diagnosis of Control Loop Nonlinearities, Valve Stiction and Data Compression. Ph.D. thesis. University of Alberta, Canada.

[4] Choudhury, M. A. A. S., Kariwala, V., Thornhill, N. F., Doukec, H., Shah, S. L., Takadac, H., and Forbes, J. F. (2007). Detection and Diagnosis of plant-wide oscillations. *Canadian Journal of Chemical Engineering* 85, 208 – 219.

[5] Choudhury, M. A. A. S., Shah, S. L., and Thornhill, N. F. Diagnosis of Process Nonlinearities and Valve Stiction – Data Driven Approaches. *Springer* (2008).

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

[7] Choudhury, M. A. A. S., Shah, S. L., and Thornhill, N. F., and Shook, D. (2006c). Automatic detection and quantification of stiction in control valves. *Control Engineering Practice* 14, 1395 – 1412.

[8] Collis, W. B., White, P. R., and Hammond, J. K. (1998). Higher-order spectra: The bispectrum and trispectrum. *Mechanical Systems and Signal Processing* 12, 375 – 394.

[9] Daniel D. Lee and H. Sebastian Seung (1999). "Learning the parts of objects by non-negative matrix factorization". *Nature* 401 (6755): 788–791.

[10] Daniel D. Lee and H. Sebastian Seung (2001). "Algorithms for Non-negative Matrix Factorization". *Advances in Neural Information Processing Systems* 13. pp. 556–562.

[11] Desborough, L., and Miller, R. (2002). Increasing customer value of industrial control performance monitoring – Honeywell's experience. In: *AIChE Symposium Series*. Number 326. 153 – 186.

[12] Ender, D. (1993) Process control performance: Not as good as you think. *Control Engineering*, 40, 180 – 190.

[13] Ettaleb, L. (1999). Control Loop Performance Assessment and Oscillation Detection. Ph.D. thesis. The university of British Columbia.

[14] Ettaleb, L., Davies, M. S., Dumont G. A. and Kwok, E., Monitoring oscillations in a Multiloop System. In: *Proceedings of the IEEE International Conference on Control Applications*. Dearborn, MI. September 15 – 18, 1996, 859 – 863.

[15] Fackrell, J.W.A. (1996). Bispectral Analysis of Speech Signals. Ph.D. thesis. The University of Edinburgh. UK.

[16] Forsman, K., and Stattin, A. (1999) A new criterion for detecting oscillations in control loops, *European Control Conference*, Karlsruhe, Germany.

[17] Goldberg, D. E., *Genetic Algorithms in search, optimization, and machine learning*, Addison-Wesley, 1989.

[18] Hagglund, T. (1995). A control loop performance monitor. *Control Engineering Practice* 3(11), 1543 – 1551.

[19] Harris, T. (1989) Assessment of closed loop performance. *Canadian Journal of Chemical Engineering*, 67, 856 – 861.

[20] Harris, T., Boudreau, F., & MacGregor, J. F. (1996). Performance assessment of multivariate feedback controllers. *Automatica*, 32, 1505 – 1518.

[21] Hinich, M. J. (1982). Testing for gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis* 3, 169 – 176.

[22] Holland, J. H., *Adaptation in natural and artificial systems*, Addison-Wesley, 1975.

[23] Horch, A. (1999). A simple method for detection of stiction in control valves. *Control Engineering Practice* **7**, 1221 – 1231.

[24] Hoyer, P. O., (2004) Non-negative matrix factorization with sparseness constraints, Journal of Machine Learning Research 5, 1457 – 1469.

[25] Huang, B., and Shah, S. L., (1999). *Performance Assessment of Control Loops: Theory and Applications*. Springer-Verlag. Germany.

[26] Huang, B. & Shah, S.L. (1998). Practical issues in multivariable feedback control performance assessment. *Journal of Process Control*, 8, 421 – 430.

[27] Hurri, J., Gävert, H., Särelä, J., Hyvärinen, A. (1998). *The FastICA package for MATLAB*. Retrieved from http://www.cis.hut.fi/projects/ica/fastica/.

[28] Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent component analysis*. New York: Wiley.

[29] Jarvis, R. M., and Goodacre, R., (2005) Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics* 21(7), 860 – 868.

[30] Jelali, M. (2006). An overview of control performance assessment technology and industrial applications. *Control Engineering Practice* 14, 441 – 466.

[31] Jiang, H. (2008). Detection and diagnosis of poor control performance. Ph.D. thesis. University of Alberta, Canada.

[32] Jiang, H., Choudhury, M. A. A. S., and 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.

[33] Kim, Y. C., and Powers, E. J. (1979). Digital bispectral analysis and its applications to nonlinear wave interactions. *IEEE Transactions on Plasma Science* PS-7, 120 – 131.

[34] Lee, J., M., Yoo, C., & Lee, I., B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control*, *14*(5), 467 – 485.

[35] Lynch, C., & Dumont, G. A. (1996). Control loop performance monitoring. *IEEE Transactions on Control Systems Technology*, 18, 151 – 192.

[36] Mah, R. (1989). *Chemical Process Structures and Information Flows*. Butterworth Publishes. 41, 42, 43.

[37] Man, K. F., Tang, K. S., and Swong, S. (1999) *Genetic Algorithms*, Springer.

[38] Matsuo, T., Sasaoka H., and Yamashita, Y. (2003). Detection and diagnosis of oscillations in process plants. *Lecture Notes in Computer Science* 2773, 1258 – 1264.

[39] Matsuo, T., Tadakuma, I., and Thornhill, N. F. (2004). Diagnosis of a unit-wide disturbance caused by saturation in a manipulated variable. In: *Proceedings of IEEE*

*Advanced Process Control Applications for Industry Workshop 2004*. Vancouver, Canada.

[40] Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2003a). A system framework for the development and analysis of signed digraphs for chemical processes. : Algorithms and analysis. *Ind. Eng. Chem. Res.* 42, 4789 – 4810. 40

[41] Miao, T., & Seborg, D. E. (1999). Automatic detection of the excessively oscillatory feedback control loops. In: *Proceedings of IEEE conference on control applications*, Kohala Coast-Island, USA.

[42] Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1992.

[43] Mitchell, M., *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, 1998.

[44] Nikias, C. L., and Petropulu, A. P. (1993). *Higher-Order Spectra: A Nonlinear Signal Processing Framework*. Prentice Hall. New Jersey.

[45] Paulonis, M. A., and Cox, J. W. (2003). A practical approach for large scale controller performance assessment, diagnosis, and improvement. *Journal of Process Control*, 13(2), 155 – 168.

[46] Qin, S. J. (1998) Control performance monitoring – A review and assessment. *Computers & Chemical Engineering*, 23, 173 – 186.

[47] Rao, T. S., and Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Vol. 24. Springer-Verlag. New York. Lecture Notes in Statistics.

[48] Rosenblatt, M., and Van Ness, J. W. (1965). Estimation of the bispectrum. *Annals of Mathematical Statistics*, 36, 420 – 436.

[49] Snášel, V., Krömer, P., Platoš, J., and Hŭsek, D., (2008), On the Implementation of Boolean Matrix Factorization, In *19th International Conference on Database and Expert Systems Application*, DEXA, 554 – 558.

[50] Snášel, V., Krömer, P., and Platoš, J.,(2008) Investigating Boolean Matrix Factorization, In *Data Mining using Matrices and Tensors (DMMT)*, Las Vegas, USA.

[51] Taha, O., Dumont, G. A., and Davies, M. S. (1996). Detection and diagnosis of oscillations in control loops. In *Proceedings of the 35th Conference on Decision and Control*. Kobe, Japan.

[52] Tangirala, A. K., Kanodia, J., & Shah, S. L. (2007). Non-negative matrix factorization for detection of plant-wide oscillations. *Ind. Eng. Chem. Res.*, 46 (3), 801 – 817.

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

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

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

[56] Thornhill, N. F., & Hagglund, T. (1997). Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5, 1343 – 1354.

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

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

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

[60] Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37 – 52.

[61] Xia, C., & Howell, J. (2003). Loop status monitoring and fault localization. *Journal of Process Control*, 13, 679 – 691.

[62] Xia, C., & Howell, J. (2005). Isolating multiple sources of plant-wide oscillations via independent component analysis. *Journal of Process Control*, 15, 1027 – 1035.

[63] Xia, C., Howell, J., and Thornhill, N. F., (2005) Detecting and isolating multiple plant-wide oscillations via spectral independent component analysis. *Automatica* 41, 2067 – 2075.

[64] Xia, C., Zheng, J., and Howell, J. (2007). Isolation of whole plant oscillations via non-negative spectral decomposition. *Chinese Journal of Chemical Engineering*, 15, 353 – 360.

[65] Yuan, Z., (2009) Advances in Independent Component Analysis and Non-negative Matrix Factorization, Ph.D. thesis. Helsinki University of Technology, Finland.

[66] Zang, X., and Howell, J. (2007). Isolating the source of whole-plant oscillations through bi-amplitude ratio analysis. *Control Engineering Practice*, 15, 69 – 76.

# VITAE

- *Name:* **Mohammed Nadeemullah Shareef**
- *Nationality:* Indian
- *Date of birth:* 3rd July
- Received Bachelor of Engineering in Instrumentation Engineering from M.J.C.E.T, Osmania University, Hyderabad, India in 2006.
- Received Master of Science (MS) in Systems Engineering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia in 2010.
- *Email:* nadeem.shareef@gmail.com
- *Permanent Address:* H.No.: 16 - 3 - 919, "SAHARA", Chanchalguda, Hyderabad – 500024, India.