ROOT CAUSE ISOLATION OF PROPAGATED OSCILLATIONS IN PROCESS PLANTS

XIAOYUN ZANG

A THESIS SUBMITTED TO
THE DEPARTMENT OF MECHANICAL ENGINEERING OF
THE FACULTY OF ENGINEERING OF
THE UNIVERSITY OF GLASGOW FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

Acknowledgements

First of all I would like to express my deep gratitude to my supervisor, Dr. John Howell, for his continuous help and supervisión.

I would like to give my thanks to Dr. Chunming Xia for his helpful recommendations; to Mrs. Nina F. Thornhill from UCL, for assisting with the application of the *N*-measure method and supply of data; to John Cox & Michael Paulonis from Eastman Chemical Company, Kingsport, TN, USA, for the supply of real plant data; and to Dr. Yihong Qiu, Dr. Chunmei Feng, Dr. Cristina D'Arrigo and Miss Bo Wang for their friendship.

Thanks also go to all the other people who have helped me during my stay at the University of Glasgow.

Special thanks go to my parents, my sister and my brother-in-law, for their love and encouragement.

I must thank my husband, Min Sheng for his love.

To my husband and my parents

Contents

List of Tables

List of Figures

Abbreviations

CLPA	Control Loop Performance Assessment
DFT	Discrete Fourier Transform
HOS	Higher Order Statistics
IAE	Integrate Absolute Error
IC	Independent Component
ICA	Independent Component Analysis
MVC	Minimum Variance Control
NGI	Non-Gaussian Index
NLI	Non-Linearity Index
PCA	Principle Component Analysis
PC	Principle Component
QPC	Quadratic Phase Coupling
SISO	Single Input Single Output

Abstract

Persistent whole-plant disturbances can have an especially large impact on product quality and running costs. There is thus a motivation for the automated detection of a plant-wide disturbance and for the isolation of its sources. Oscillations increase variability and can prevent a plant from operating close to optimal constraints. They can also camouflage other behaviour that may need attention such as upsets due to external disturbances. A large petrochemical plant may have a 1000 or more control loops and indicators, so a key requirement of an industrial control engineer is for an automated means to detect and isolate the root cause of these oscillations so that maintenance effort can be directed efficiently.

Although there has been considerable commercial and academic interest in methods for analysing the performance of control systems in these situations, they are usually designed to detect, and not to isolate or diagnose, faults or other root causes of poor performance. There are several possible causes for the presence of an oscillation in a control loop. Control valve non-linearity such as stiction, dead band, backlash or hysteresis, is one of the most possible causes. Another possible cause is bad tuning, which can destabilize the system. An external oscillatory disturbance is the third possible reason. Successful methods proposed in the literature have either detected the presence of sustained oscillations by analyzing the statistical properties of a signal, or diagnosed whether the detected oscillation has been caused by nonlinear problems in the loop. Only a few have tried to discriminate between valve problems and external disturbances, i.e. to isolate the root cause loop of whole-plant oscillations.

Previous methods for the isolation of the root cause of an oscillation include the D-factor and the Non-linearity index, both of which are based on an assessment of harmonic attenuation. Another method is Spectral ICA where the independent components align with harmonics also. Although reliant on the frequency response characteristics of harmonic propagation, all make assumptions about the attenuation of a plant, thus restricting their application. Major contributions of this PhD research include a detailed harmonic propagation analysis and the development of harmonic evaluation indices. The former provides an understanding needed in the application of any of the methods, whilst the latter leads to a new alternative method for the isolation of the source loop of plant-wide oscillations. Both loop-to-loop and loop-to-indicator propagation is modelled and the alternative method can detect and isolate oscillations in a multi-loop situation. Thus not only do they make use of controlled variables, but they also make use of indicator readings, set-points, and controller settings.

The propagation model that is proposed is represented by a log-ratio plot, which is shown to be 'bell' shaped in most industrial situations. Theoretical and practical issues are addressed to derive guidelines for determining the cut-off frequencies of the 'bell' from data sets requiring little knowledge of the plant schematic and controller settings. The alternative method for isolation is based on the bispectrum and makes explicit use of this model representation. A comparison is then made with other techniques. These techniques include nonlinear time series analysis tools like Correlation dimension and maximal Lyapunov Exponent and a new interpretation of the Spectral ICA method, which is proposed to accommodate our revised understanding of harmonic propagation. Both simulated and real plant data

are used to test the proposed approaches. Results demonstrate and compare their ability to detect and isolate the root cause of whole plant oscillations.

Being based on higher order statistics (HOS), the bispectrum also provides a means to detect nonlinearity when oscillatory measurement records exist in process systems. Its comparison with previous HOS based nonlinearity detection method is made and the bispectrum-based approach is preferred.

In the conclusions the various isolation methods are compared in several aspects and the bispectral method is found to be better than other isolating methods. Practical suggestions are also made to isolate propagated oscillations. The various methods, when applied together, will verify the diagnosis reached by the bispectral method. Process understanding, if applicable, also plays an important role when integrated with the data-driven analysis.

Chapter 1

Introduction

1.1 Background

It is important to detect and diagnose the causes of oscillations in process operation because a plant running close to a product quality limit is more profitable than a plant that has to back away because of variations in the product (Martin *et al.* 1991; Paulonis & Cox 2003; Qin 1998). They can also camouflage other behaviour that may need attention such as upsets due to external disturbances. A large petrochemical plant may have a 1000 or more control loops and indicators, so a key requirement of an industrial control engineer is for an automated means to detect and isolate the root cause of these oscillations so that maintenance effort can be directed efficiently. However it may not be easy to determine cause and effect particularly when physical influences propagate in the opposite direction to process flows, for instance due to recycle streams or when disturbances in the outflow stream of a tank cause deviations in the level of the tank.

Over the past decade there has been considerable commercial and academic interest in methods for analysing the performance of a controller. The development of minimum variance control (MVC) benchmark based, off-line, closed loop, performance assessment techniques (Harris & Seppala 2001) are now so well established that various vendors are offering commercial analysis products based on them. For example, Honeywell offer the *Loop Scout*™ software package, and Matrikon offer *ProcessDoctor*. These performance assessment techniques are

usually designed to detect, and not to isolate or diagnose, faults or other root causes of poor performance. As highlighted by Harris & Seppala (2001), a comprehensive approach for controller performance monitoring should also include the development of methods for diagnosing the underlying causes for changes in the performance of a control system. Another important limitation of these techniques is that the assessment assumes that the plant is in a steady state, in which the only variation in the controlled variable of a closed loop is as a result of stochastic noise. This condition is not the case when a plant is oscillating.

Controller performance monitoring has recently become popular in the continuous processing industries. An indication of this can be found in Computing and Control Engineering (2005) where a special section on control loop assessment and diagnosis has been published. The importance of root cause diagnosis is emphasized again.

Root causes of persistent oscillatory whole-plant disturbances include poorly tuned controllers, non-linearities such as saturation, dead band, or hysteresis in control valves or sensors (Shoukat Choudhury *et al.* 2004). The process itself might have nonlinearities that can cause limit cycle oscillations to build up within a control loop. Local instabilities can also arise as a result of control loop interactions when two controllers have a shared mass and/or energy store (e.g. pressure and level controllers may compete for control of the contents of a reactor). Structural disturbances might also be induced by coordinated transfers of mass and/or energy between different process units, especially when a recycle is present or by natural

resonances excited by noise, scheduled changes etc. Finally a disturbance in the loop might actually result from cyclic setpoint changes (Paulonis & Cox 2003).

If there are some nonlinearities in the control loop, the controller may not perform at the desired level. Nonlinearities degrade the performance of the controller in several ways. For example they may produce oscillations in process variables, shorten the life of the control valve, may upset process stability, and in most cases lead to inferior quality end-products thus causing larger rejection rates and reduced profitability. The nonlinearities may be present in the process itself or in the actuators or control valves. The thesis focuses on diagnosing oscillations induced by actuator nonlinearities that arise from control loops. Actuator or valve nonlinearities are typically due to faults such as stiction, backlash, saturation, deadzone, ruptured diaphragm, and/or corroded or eroded valve seats. These non-linearity induced oscillations have certain properties which change as the oscillations propagate around a plant. The thesis examines how these changes might be analysed to isolate the root cause.

1.2 Aims and Objectives

In particular this thesis describes control loop measurement based approaches for (1) the analysis of oscillation propagation based on an understanding of process dynamics and PI/PID controller settings, and (2) the detection and isolation of oscillations in a multi-loop situation. The aim is that these approaches should be as data-driven as possible, minimizing the need for detailed process model information, i.e. they should be based on available control loop measurements and controller settings. Measurements may include:

- Controlled variables
- Controller outputs
- Sensor readings from indicators
- Set-points

The objectives include:

- trend categorization: oscillatory or not, high or low frequency and so on;
- root cause isolation;
- initial diagnosis to determine, for instance, whether the oscillations are caused by a nonlinear valve problem or an aggressive controller, which are the most common causes for the severe deterioration of loop performance.

The approaches described are listed below.

(1) A loop-to-loop propagation model is proposed to describe how the harmonic content of an oscillation changes as it propagates through a plant. This model can be represented by a set of log-log plots, the pertinent features of which can be obtained from a knowledge of the controller type & settings, the frequency of the oscillations and control loop structures. A loop isolation procedure is then proposed based on an understanding of these models. A similar loop-to-indicator propagation model is developed that can help verify inferences made.

(2) Changes in harmonic content can be measured via the bispectral analysis.

(3) In addition nonlinear time series analysis tools such as Correlation dimension and maximal Lypunov Exponent are examined as alternative measures of the changes in the properties of an oscillation as it propagates.

(4) Knowledge of loop-to-loop propagation leads to a revision to the published Spectral independent component analysis (Spectral ICA) so that it is able to isolate a root cause in both low and high frequency cases.

(5) A bispectral approach can also give an indication of the scale of a non-linearity because its peak measures harmonic couplings.

1.3 Testing The Techniques That Meet These Aims And Objectives

Being data driven, the techniques that are proposed here were developed by analyzing both simulated and real data. Data was obtained from four sources.

(1) Crude and simple simulations were used to generate data to examine basic properties.

(2) Real data was obtained from an Eastman Chemical Plant. This contained low frequency oscillations, the cause of which had already been diagnosed by plant engineers.

(3) Real data was obtained from a SE Asian Refinery Plant, which contained high frequency oscillations. Unfortunately the source of these oscillations was not known.

(4) Two 4×4 multiloop simulations are used to validate the loop-to-loop propagation model, in particular the relationship between controller integral time and cut-off frequencies.

1.4 Outline Of The Thesis

Chapter 2 first briefly reviews some general methods for process monitoring and fault isolation, and then focuses on some specific methods that have had direct influences on the development of the approaches proposed here.

Whole plant oscillations are then examined in Chapter 3, and loop-to-loop & loop-to-indicator propagation models are developed.

The bispectral analysis method is introduced in Chapter 4, where its ability to extract the harmonic content of oscillating signals is examined. A bi-amplitude ratio index is proposed as a measure of the power ratio of the fundamental to the third harmonic and a procedure is then proposed to localize the source loop of propagated oscillations based on this measure.

Nonlinear time series tools can provide alternative measures that can be input into the same procedure. This is considered in Chapter 5, where a revision to the Spectral ICA method is also proposed. The performance of the various measures & methods are then compared on a simulated case and two sets of industrial data.

Bispectral analysis as a nonlinearity detection tool is then discussed in Chapter 6.

The various methods proposed in the thesis together with methods proposed by other authors are compared in the final chapter, Chapter7 and some future directions are advised.

1.5 Originality

Harmonic propagation modelling, the bi-amplitude ratio and its application, the application of nonlinear time series analysis tools on root cause isolation and the new interpretation of Spectral ICA are all original. The work has been presented at various symposia and conferences(Zang & Howell 2003; Zang & Howell 2004a; Zang & Howell 2004b), published in the International Journal of Adaptive Control and Signal Processing (Zang & Howell 2005a), and submitted to the Journal of Process Control (Zang & Howell 2005c) and Control Engineering Practice (Zang & Howell 2005b).

Chapter 2

Some Background To Propagated Oscillations In Process Plants

This chapter first briefly reviews some general methods for the detection and diagnosis of plant-wide oscillatory disturbances, and then focuses on some specific methods that have had direct influences on the development of the approaches proposed.

2.1 Overview

Severe deterioration of performance in process control loops is often indicated by oscillations. It is reported that about 30% of all control loops in a paper mill plant were oscillating because of valve problems (Hägglund, 1995). Ender (1993) reported similar results. The oscillations may cause increased energy consumption, waste of raw material and sometimes a less uniform product. Reducing or removing oscillations yields commercial benefits (Hägglund 1995; Martin *et al.* 1991; Shinskey 1990; Thornhill & Hägglund 1997; Thornhill *et al.* 2003b). Thus simple and efficient methods for oscillation detection and diagnosis is crucial (Xia 2003).

There are several possible causes for the presence of an oscillation in a control loop. Control valve non-linearity such as stiction, dead band, backlash or hysteresis, is

one of the most possible causes. The fundamental frequency of a limit cycle caused by a valve non-linearity can be estimated by employing a technique based on describing functions. A significant phase shift in the describing function usually implies a lower-frequency limit cycle, on the other hand, a small phase shift in the describing function usually implies a higher-frequency limit cycle, one whose frequency is closer to the loop ultimate frequency. Another possible cause is bad tuning, which can destabilize the system. An external oscillatory disturbance is the third possible reason. This situation often happens in a plant that contains loop interactions. The source oscillation will propagate to other interacting loops and result in secondary oscillations. Since interaction is inevitable in process plants, the isolation of the oscillation source becomes difficult.

A number of researchers and practitioners have indicated that more realistic estimates of the achievable performance of a plant can be obtained when one detects, diagnoses and removes the effect of oscillations (Harris & Seppala 2001; Horch A. 2000). Methods for detecting oscillations and valve stiction are described in Hägglund (1995), Bittanti *et al.*(1997), Horch & Isaksson (1998; 1999), Seborg *et al.*(1989), (Singhal & Salsbury 2005), Rossi & Scali (2005) and Tangirala, Sharh & Thornhill (2005). Methods for diagnosing stiction are suggested in Deibert (1994), Taha *et al.* (1996), Ogawa (1998), Horch & Isaksson (1998), Horch (1999) and Stenman *et al.* (2003). Less work has been carried out to isolate a loop on the basis of measurement time series records collected from controllers and sensors distributed throughout a plant. The literature is limited to Thornhill *et al.*(2001), Thornhill *et al.*(2002b), Dobson & Thornhill (2002), Thornhill *et al.*(2003b), Xia & Howell(2003a) and Xia & Howell (2005).

2.2 Oscillation Detection

Considerable research has focused on the automatic detection of oscillatory behaviour.

2.2.1 Zero-Crossings Analysis Based Approaches

One of the popular oscillation detection techniques in SISO systems is based on the analysis of zero-crossings of the loop operating data. Hägglund (1995) presented a real-time oscillation detection method that calculates the integrated absolute deviation (IAE) between successive zero crossings of the controller error signal. The idea is that when the controller error is oscillatory rather than random, such deviations are large and the interval between them is large, i.e., an oscillation signal has larger IAE values than a random one. The definition of IAE is given by

$$IAE_i = \int_{t_i}^{t_{i+1}} |Y(t)|\,dt$$

where $Y(t)$ is the controller error signal and t_i and t_{i+1} are times of successive zero crossings of $Y(t)$. These IAE deviations are compared to a threshold value which is based on the IAE value of a sinusoidal oscillation of amplitude a and on the highest frequency that might occur in the loop, i.e. the ultimate natural frequency ω_u. The IAE for such a sine wave ($a\sin(\omega_u t)$) is $\frac{2a}{\omega_u}$. When the ultimate frequency is not known, a good approximation for ω_u could be $\frac{2\pi}{T_i}$, where T_i is the controller integral constant. This is because, in a well-tuned controller, the ultimate frequency

at which a resonant oscillation occurs is similar to $\frac{2\pi}{T_i}$. Thus the condition for oscillation detection is:

$$IAE_i \geq \left[\frac{2a}{\omega_u} = \frac{aT_i}{\pi}\right] \tag{2.1}$$

In a real-time implementation, an oscillation is detected if the IAE deviation exceeds the threshold, with a set to one percent of the controller range over a supervision time of 50 times the presumed oscillation period.

Thornhill and Hägglund (1997) extended Hägglund's zero-crossings idea to an off-line analysis for oscillation detection. Forsman & Stratin (1999) improved the zero-crossings techniques to accommodate the detection of asymmetric and irregular oscillations.

Criticism 1

These methods are simple and easy to implement, however excessive noise will degrade their performance. Also the first two techniques were designed under the assumption of symmetric oscillations.

2.2.2 Auto-Correlation Analysis Based Approaches

Miao & Seborg (1999) have proposed a statistical-based approach to the detection of excessively oscillatory feedback control loops. The test is based on a decay ratio, which is obtained from the auto-correlation of the operating data, rather than from the oscillatory operating data itself. The underlying basis is that the auto-correlation

of an oscillatory signal is also oscillatory. Thus if the decay ratio is greater than a specified threshold, it is concluded that the signal is excessively oscillatory. Thornhill *et al.* (2003a) presented a method for the detection of oscillations in measurements from chemical processes including the case when two or more oscillations of different frequency are present simultaneously. They used a new method based on the regularity of the zero crossings of the filtered auto-covariance data to determine the presence of oscillations in selected frequency ranges. An automated filter selection algorithm for the selection of frequency range was proposed.

Criticism 2

This auto-correlation or auto-covariance based approach is an improvement on time domain zero-crossing methods in the sense that the calculation of auto-correlation essentially reduces noise effects. But the computational burden will be heavier, making it less suitable for online implementation. The application of zero-crossing analysis to auto-covariance data combines the advantages of both and is therefore superior to other off-line detection methods.

2.2.3 Spectral PCA Based Approaches

Thornhill *et al.* (2002a) have described the principle component analysis (PCA) of the power spectra of data from chemical processes. The novel feature of spectral PCA is that the rows of the data matrix, **X**, are the single-sided power spectra $P(f)$ of the signals over a range of frequencies up to the Nyquist frequency:

$$\mathbf{X}=\begin{pmatrix} P_1(f_1) & \cdots & P_1(f_N) \\ \cdots & \cdots & \cdots \\ P_m(f_1) & \cdots & P_m(f_N) \end{pmatrix}$$

N frequency channels ⟶

m process variables ↓

By performing PCA of this data, the power spectra in **X** can be decomposed into combinations of several dominant spectrum-like principle components (PCs):

$$\mathbf{X}=\begin{bmatrix} t_{1,1} \\ t_{2,1} \\ \vdots \\ t_{m,1} \end{bmatrix}\mathbf{v}_1{}' + \begin{bmatrix} t_{1,2} \\ t_{2,2} \\ \vdots \\ t_{m,2} \end{bmatrix}\mathbf{v}_2{}' + \cdots + \begin{bmatrix} t_{1,n} \\ t_{2,n} \\ \vdots \\ t_{m,n} \end{bmatrix}\mathbf{v}_n{}' + \mathbf{E} \qquad (2.2)$$

where $\mathbf{v}_1{}'$ to $\mathbf{v}_n{}'$ represent the dominant row-major spectrum-like PCs, **E** is an error matrix, which includes truncated principle components and $t_{i,j}$ (for i=1…m, j=1…n) are mixing factors. For example the i^{th} row power spectrum in **X** are approximately constructed by $\mathbf{v}_1{}'$ to $\mathbf{v}_n{}'$ with the relevant mixing factors ($t_{i,1}$ $t_{i,2}$,…,$t_{i,n}$).

A single dominant peak in a power spectrum, indicating an oscillation in the time domain, can be created by proper linear combinations of the PCs. The clustering of similar oscillatory trends is achieved by a scores plot, in which the point ($t_{i,1}$ $t_{i,2}$,…,$t_{i,n}$) maps to the i^{th} power spectrum. Similar spectra have similar *t*-coordinates. Therefore such groups form clusters. PCA of the auto-covariance functions of process variables gives similar or sometimes even superior performance compared to spectral PCA.

Criticism 3

Spectral PCA has superior performance to time domain PCA for the detection of clusters of data trends having similar features, even when time shifting is used to align the data trends. This is because spectra are invariant to the phase lags caused by time delays and process dynamics. Although the interpretation of what is detected, and the grouping of oscillation disturbances based on spectral PCs, is not straightforward because the same spectral peaks could be present in more than one PC and becomes difficult when the number of dominant PCs increases, it is still a good approach to the identification of the dominant spectral peaks. However it doesn't address the isolation of these oscillations, which is of greater interest to plant operating engineers.

2.3 Oscillation Diagnosis

The diagnosis of oscillations is complicated. First the source must be isolated to a control loop, or energy transfers or external etc., then a decision must be made as to whether it is caused by a poorly tuned controller or by a nonlinearity or by an external disturbance and so on.

2.3.1 On-line Real-time Diagnosis

Hägglund (1995) proposes a loop-by-loop test approach. He determines if the oscillations are being generated outside the control loop, or if they are generated inside the loop by disconnecting the feedback by switching the controller to manual mode. If the oscillation is still present, the disturbances must be generated outside the loop; otherwise, they were generated inside the loop. The inside problem could be either friction in the valve or a badly tuned controller. Whether friction is present or not can be determined by making small changes in the control signal and checking if the measurement signal follows.

Criticism 4

The on-line approach is applicable to control engineers for trouble-shooting. It is time consuming and may not be possible to find the source of the disturbances if the plant is tightly interacting and many loops are oscillating at the same time.

2.3.2 Measurement Record Based Approaches

Thornhill and Hägglund (1997) provide off-line diagnostic methods which use indicators from routine operating data: dynamic sp-pv maps for loops where the set point changes often, such as loops in a cascade mode; the CLPA index, regularity factor and oscillation-detection threshold. CLPA provides a grey indication because a low CLPA value does not always mean that no action is needed whilst sp-pv maps give a clear signature (Thornhill and Hägglund, 1997). The diagnosis procedure uses the indicators to guide a process control engineer towards suitable special off-line tests.

For loops in cascade mode where the set point moves, the presence of non-linear features can be revealed in an *sp-pv* map. Åström (1991) and McMillan (1995) have illustrated the typical patterns that appear on plots of signal versus stroke for non-linear characteristics such as valve dead band and stick slip. Figure 2-1 shows idealized shapes for dynamic *sp-pv* maps. If the *sp-pv* plot of a loop is similar to the idealized signature of a valve dead band, or a stick slip or an oversized valve, valve non-linearity is diagnosed; if it is similar to the idealized signature of a linear valve with phase lag, bad tuning is diagnosed.

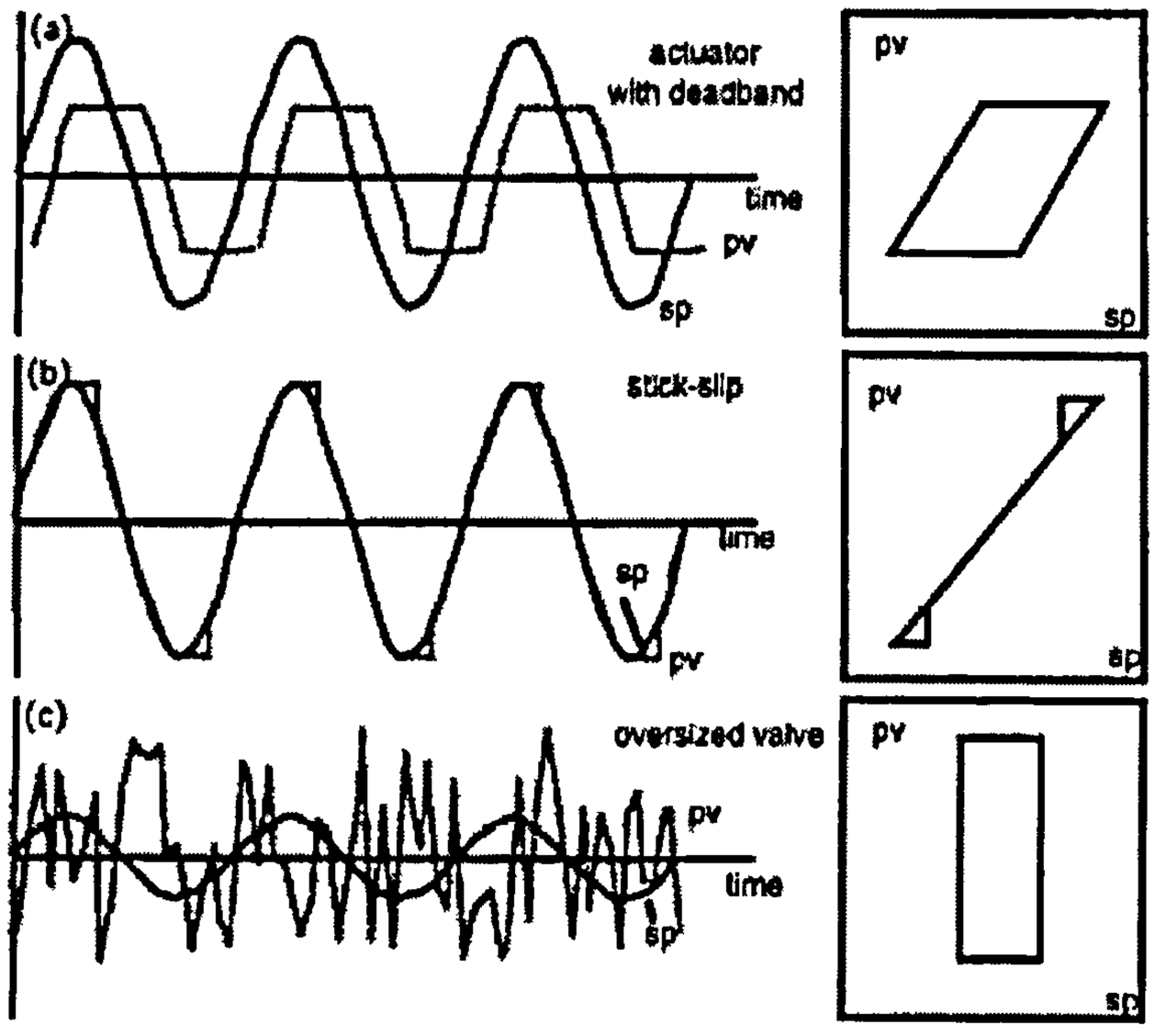


Figure 2-1: Idealized sp-pv maps for dynamic systems (Thornhill and Hägglund, 1997)

The basis of control-loop performance assessment (*CLPA*) described by Desborough & Harris (1992) is that the controller error should have no predictability over some given prediction horizon. The controller error sequence Y is decomposed as:

$$Y_t = \hat{y}_t + r_t \tag{2.3}$$

where $\hat{y}$ is the predictable component of the controller error and r the zero mean residuals. The aim of regulatory control is to remove any predictable components. That is, $\hat{y}$ should be small or zero. A common cause of a predictable component is a persistent oscillation. Desborough & Harris's performance index can be expressed as:

$$\eta = \frac{mse(\hat{y}_i)}{mse(Y_i)} = 1 - \frac{\sigma_r^2}{mse(Y_i)} \tag{2.4}$$

with $mse(\hat{y}_i)$ being the mean square value of the predictable component, $mse(Y_i)$ the mean square value of the controller error and σ_r^2 the variance of the (zero mean) residuals. When the control performance is good the controller error has little predictability and the index is 0 because $mse(Y_i) = \sigma_r^2$. The opposite is true for a poorly controlled loop in which the controller error is predictable.

Thornhill and Hägglund extended Hägglund's zero-crossings idea to an off-line analysis for oscillation detection by assessing the significance of a deviation relative to the r.m.s. value of the residuals (σ_r) that are provided by the CLPA algorithm rather than the absolute criterion (1% of the range). The off-line condition for a significant deviation relates the mean value of *IAE* over an interval ΔT between the zero crossings (t_{i+1}-t_i) to σ_r:

$$\frac{IAE_i}{\sigma_r \cdot \Delta T_i} \geq \xi \tag{2.5}$$

where ξ is the *oscillation detection threshold.* The normalized IAE deviations ($\frac{IAE_i}{\sigma_r \cdot \Delta T_i}$) are compared to a threshold value which is based on the normalized IAE value of a sinusoidal oscillation of amplitude a ($a\sin(\omega t)$) in the presence of noise having an r.m.s value of σ: $\frac{2a}{\pi\sigma}$.

The normalized IAE deviations can be plotted at the times t_{i+1} and the *regularity factor* is derived from the sequence of ratios between adjacent intervals Δk at which deviations cross the threshold. Thus:

$$R_i = \frac{\Delta k_{i+1}}{\Delta k_i} \tag{2.6}$$

and the regularity factor, q, which depends on the threshold, is:

$$q(\xi) = \frac{mean\ \ value\ \ of\ \ R}{standard\ \ deviation\ \ of\ \ R}. \tag{2.7}$$

Signals judged by eye to have regular oscillations have values of q above about 1.3.

The diagnosis method using interpretation of the *CLPA* index η, oscillation detection threshold ξ and regularity factor q is then: if $\xi > 2/\pi$ and q>1.3, non-linearity is diagnosed; if $\xi > 2/\pi$ and $q \leq 1.3$, an external disturbance is possible; if $\xi \leq 2/\pi$ and q>1.3 and $\eta > 0.15$, a tuning problem is diagnosed.

Criticism 5

The two off-line diagnosis methods are data-driven and need less loop information (e.g. the controller integral time) than online approaches. However, the dynamic *sp-pv* maps are only suitable for loops where the set point changes often, such as loops in cascade mode; the CLPA based method does not have a rigorous statistical interpretation and it is a grey signature because even when the controller error has little predictability, the r.m.s. value may still be too high where a hardware problem may have happened. The performance of the CLPA based method could also degrade under circumstances with excessive noise.

Thornhill and Hägglund (1997) also examined the features in the power spectrum of a controller error to produce a partial diagnosis: if the spectrum has harmonics a non-linearity might be present; if the spectrum has multiple peaks but no harmonics it is possible that the loop has been disturbed by neighbour loops; if there is only a single peak in the power spectrum, this is a signature of a tuning problem.

Criticism 6

The approach gives somewhat coarse results because many disturbed loops have obvious harmonic contents. The approach needs to be augmented by other diagnostic methods like *sp-pv* maps. Excessive noise may also disturb the appearance of harmonics in the power spectrum.

Horch (1999) presented a new method to identify stiction in control valves. The method is based on the cross-correlation between control input and process output. The motivation is as follows: the limit cycle caused by a sticking valve will result in a square wave profile of the controlled variable $y(t)$ and a triangular wave profile of the PI controller output $u(t)$ (see Figure 2-2) with a $\frac{\pi}{2}$ phase shift between them, provided that the process has no integral action.

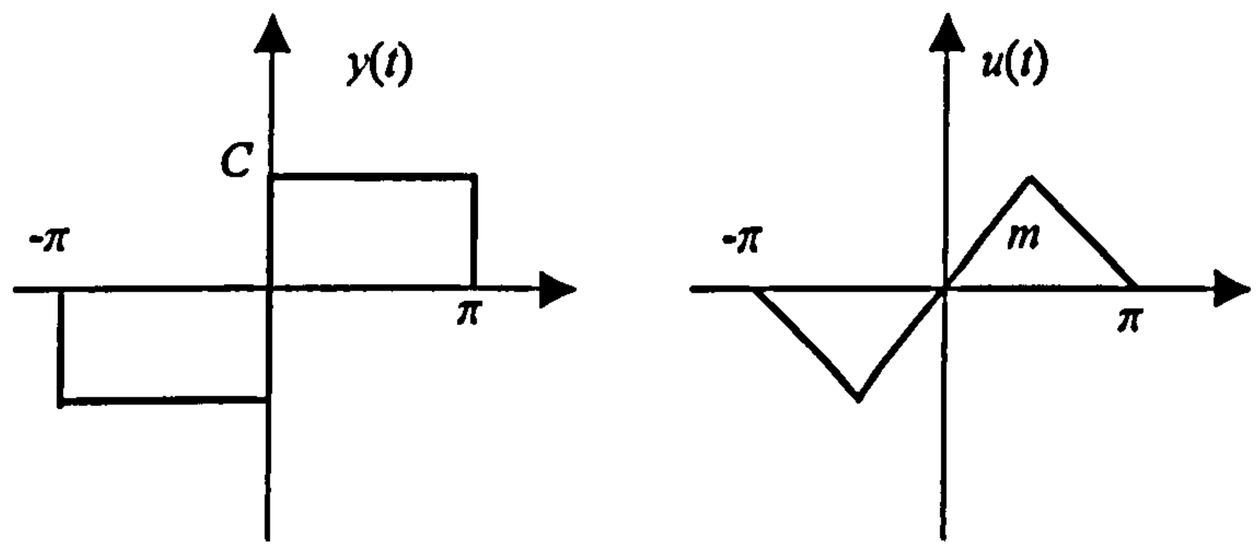


Figure 2-2: Ideal $y(t)$ and $u(t)$ signals in the presence of stiction

The detection condition is then:

(1) in the case of static friction (stiction) in a control valve, the $y(t)$ and $u(t)$ signals are shifted by approximately $\frac{\pi}{2}$ so that their cross-correlation is an odd function;

(2) oscillating external disturbances result in a phase shift of approximately π, so that the cross-correlation is an even function. This is because, in general, low frequency disturbances will be eliminated efficiently by the PI-controller since a controller with integral action has a high loop gain at low frequencies;

(3) an unstable loop which oscillates with constant amplitude (due to physical saturation) also results in an even cross-correlation function.

Both conditions (2) and (3) imply a high frequency oscillation such that the phase shift between $y(t)$ and $u(t)$, which is caused by the PI controller, is small and the only phase shift, i.e. π, is caused by the negative feedback sign.

A similar shape-based idea was used by Rengaswamy *et al.* (2001) to detect and diagnose different kinds of oscillations in control loops. The method is based on a semi-qualitative approach for the identification of different shapes of oscillation: square, triangular and non-triangular. The diagnosis of valve friction or hysteresis can then be achieved by examining the shape of the oscillation of the controlled variable and the controller output. For example: (1) a triangular controller output plus a non-triangular controlled variable corresponds to a situation of asymmetric friction; (2) a triangular controller output plus a square controlled variable corresponds to the case of hysteresis.

Criticism 7

Horch's sticking valve detection method is simple and straightforward, and process model information is not required. However it has some drawbacks: (1) the distortion of the ideal $y(t)$ and $u(t)$ signal shapes in the presence of stiction could degrade the detection performance; (2) a low frequency external oscillatory disturbance could still be dominant in the disturbed loop, even if it is largely attenuated, and such a low frequency oscillatory disturbance will result in an odd cross-correlation function, which will contradict the second detection condition; (3) the method is made under the assumption that the control loop oscillates in a way such that the controller output and process variable display symmetric triangular and square waves, which is not always the case.

Rengaswamy *et al.*'s method largely relies on the oscillation shape. A low degree of valve non-linearity may not cause the typical pattern, thus the diagnosis of the oscillation might not be effective in such cases. However, the oscillation detection algorithm has quite good performance, because it can recover shapes from data with up to 20% noise.

Shoukat Choudhury *et al.* (2004) developed two indices, the Non-Gaussian Index (*NGI*) and the Non-Linearity Index (*NLI*), based on Higher Order Statistics (HOS) theory to detect and quantify signal non-Gaussianity and non-linearity. The approach evaluates the *NGI* and *NLI* of routine operating data and can determine whether or not the measurement record is from a linear process driven by Gaussian

signal (*NGI*<0), or from a linear process driven by non-Gaussian signal (*NGI*>0 & *NLI*=0), or from a non-linear process (*NGI*>0 & *NLI*>0).

Both *NGI* and *NLI* are derived from the calculation of the squared bicoherence. Bicoherence is defined as:

$$bic^2(f_1,f_2)=\frac{|B(f_1,f_2)|^2}{E[|X(f_1)X(f_2)|^2]E[|X(f_1+f_2)|^2]} \tag{2.8}$$

where 'bic' is known as the bicoherence function, $X(f)$ is the discrete Fourier transform of time series $x(k)$, and $B(f_1, f_2)$ is the bispectrum in the bifrequency (f_1, f_2) defined as:

$$B(f_1,f_2)=E\left[X(f_1)X(f_2)X^*(f_1+f_2)\right] \tag{2.9}$$

where '*' denotes complex conjugate.

The test for Gaussianity is as follows:

- Null Hypothesis, H_0: The signal is Gaussian
- Alternate Hypothesis, H_1: The signal is not Gaussian.

Under the null hypothesis, the test for the average squared bicoherence can be based on the following equation:

$$P\left(2KL\overline{\hat{bic}^2}>c_\alpha^{\chi^2}\right)=\alpha \tag{2.10}$$

where, $c_\alpha^{\chi^2}$ is the critical value calculated from the central χ^2 distribution table for a significance level of α at 2L degrees of freedom since $\overline{\hat{bic}^2}=\frac{1}{L}\sum_{i=1}^{L}\hat{bic}_i^2$, L is the number of bifrequencies inside the principle domain of the bispectrum, and K is the number of segments used in data segmentation during bicoherence estimation. If the

number of bifrequencies in the principle domain is very large (more than 100) the normal approximation of the χ^2 distribution can be used:

$$P\left(\overline{\hat{bic}^2} > \frac{1}{4KL}\left[c_\alpha^z + \sqrt{4L-1}\right]^2\right) = \alpha \tag{2.11}$$

where c_α^z is the critical value of standard normal distribution at a significance level of α. NGI is therefore defined by

$$NGI = \overline{\hat{bic}^2} - \frac{1}{4KL}\left[c_\alpha^z + \sqrt{4L-1}\right]^2 \tag{2.12}$$

At a confidence level of α the following rule based decision can be obtained:

- if $NGI \leq 0$, the signal is GAUSSIAN
- if $NGI > 0$, the signal is NON-GAUSSIAN

It is shown in (Shoukat Choudhury *et al.* 2004) that if the signal is non-Gaussian and linear the magnitude of squared bicoherence should be a non-zero constant at all bifrequencies in the principle domain. At a 95% confidence level if the maximum squared bicoherence, $\hat{bic}^2_{max}$ is less than $\overline{\hat{bic}^2} + 2\sigma_{\hat{bic}^2}$, the magnitudes of squared bicoherence are assumed to be constant or the surface is flat. The *NLI* is defined as:

$$NLI = \left|\hat{bic}^2_{max} - \left(\overline{\hat{bic}^2} + 2\sigma_{\hat{bic}^2}\right)\right| \tag{2.13}$$

where, $\sigma_{\hat{bic}^2}$ is the standard deviation of the estimated squared bicoherence and $\overline{\hat{bic}^2}$ is the average of the estimated squared bicoherence. At a confidence level of 95% the following rule based decision can be obtained:

- if $NLI = 0$, the signal generating process is LINEAR
- if $NLI > 0$, the signal generating process is NONLINEAR

Criticism 8

The detected non-linearity cannot necessarily be attributed to valve problems because the nonlinear nature of the process itself may have caused a large *NLI*. In addition it can't discriminate between an internal non-linearity and an external disturbance due to non-linearity in another control loop, i.e. it cannot diagnose the root cause. The authors (Choudhury *et al.* 2004; Shoukat Choudhury *et al.* 2004) suggest using the specific patterns in the process output (*pv*) vs. the controller output (*op*) plot to help diagnose the causes of poor control loop performance. However the same authors argue in (Shoukat Choudhury *et al.* 2005) that the *pv-op* map captures not only the nonlinear valve characteristic but also the dynamics of the process and hence in some cases would fail to indicate a valve problem. They also suggest the mapping of valve position (*mv*) vs. *op*, but unfortunately this is often unavailable because the valve positions may not be recorded.

2.4 Oscillation Isolation

A common source of oscillations is a limit cycle caused by a control valve with a deadband or excessive static friction (Åström 1991; Shinskey 2000). A process variable oscillating for that reason can readily propagate the oscillation to other variables and disturb other control loops, hence causing a plant-wide disturbance. A focus upon non-linear root causes can thus be justified because valve friction causes the majority of cases, according to reported surveys (Ender 1993). The methods listed in this section are the few published studies that have been carried out to isolate a loop on the basis of measurement time series records collected from controllers and sensors distributed throughout a plant.

It is well known that non-linearity in a control loop can give rise to limit cycles (Thornhill and Hägglund, 1997). Limit cycles are periodic, and are usually non-sinusoidal. Harmonics having frequencies at integer multiples of a fundamental frequency appear in the power spectrum if the periodic component is non-sinusoidal. Valve non-linearity (stiction or hysteresis) induced oscillations in a control loop usually contain both even and odd harmonics. Such oscillations can be propagated to other loops due to high interactions in chemical process plant.

Xia & Howell (2003a) used spectral ICA to extract a significance index that links the sources to specific plant measurements in order to facilitate the isolation of the sources of the oscillations; Thornhill *et al.*(2003b) proposed an index for non-linearity that grows stronger closer to the source. These data-driven approaches are

listed here and their comparison with the root cause diagnosis approaches proposed in this thesis will be given in Chapter 7.

Thornhill *et al.* (2001) presented a distortion factor signature to aid the isolation of oscillations caused by limit cycles. The *distortion factor*, D, is defined as:

$$D \triangleq \sqrt{\frac{P_{tot} - P_{fund}}{P_{tot}}} \tag{2.14}$$

where P_{tot} is the total power in the fundamental plus harmonics and P_{fund} is the power in frequency channels k to $k+l$ occupied by the fundamental harmonic. The frequency range k to $k+l$ was determined by looking at the power spectra.

Root cause isolation based on distortion factors assumes that the harmonics of a limit cycle become smaller as the measurement point moves away from the root cause and that the time trends become more sinusoidal. The measurement having the highest distortion factor has more power in the harmonics and is thus a candidate for the root cause.

Criticism 9

The D factor approach is based on the same assumption as the non-linearity index method, i.e., on the low-pass filtering property of physical processes. As will be seen in Chapter 3, this does not hold in some cases. Furthermore, D cannot be determined in cases with no well-defined oscillation and no spectral peak.

Thornhill *et al.* (2003b) also proposed a time domain measurement based isolation approach, the Non-linearity index. To obtain the Non-linearity index of a set of data, the time series i.e. $x_1, x_2, \cdots x_N$ must first be projected into a multi-dimensional phase space:

$$\mathbf{X}_i = (x_i, x_{i+d}, x_{i+2d}, \cdots x_{i+(m-1)d}), \quad i = 1, 2, \cdots, N \tag{2.15}$$

where N is the data length, d is a time delay, m is known as the embedding dimension and $\mathbf{X}_i$ is called an embedding vector of m-dimension. For example with d=2 and m=3, $\mathbf{X}_1 = (x_1, x_3, x_5)$, $\mathbf{X}_2 = (x_2, x_4, x_6), \mathbf{X}_3 = (x_3, x_5, x_7)$ and so on. Thornhill *et al.* adopted a more engineering intuitive approach to the construction of the phase space. The data is pre-processed to have S samples per cycle, end-matched, mean centred and scaled to unit standard deviation. It is then embedded into a phase space where d=1 and m=S so that each embedding vector represents a single cycle of oscillatory data.

The non-linearity index, N, is then formed to test the null hypothesis that the data was output by a linear system driven by Gaussian white noise:

$$N = \frac{\bar{\Gamma}_{surr} - \Gamma_{test}}{3\sigma_{\Gamma_{surr}}} \tag{2.16}$$

where $\bar{\Gamma}_{surr}$ and $\sigma_{\Gamma_{surr}}$ represent the mean and standard deviation of a reference probability distribution and root mean square (r.m.s.) error Γ_{test} is derived by comparing the data with a smoothed version, obtained by applying a non-linear algorithm to it. The first S data points of this smoothed version are as measured (i.e. $\forall i : 1 \le i \le S, \hat{x}_i = x_i,$). Embedding vectors that are nearest neighbours to the first embedding vector, i.e. $(x_1, x_2, \cdots x_S)$, are now identified and the subsequent points

e.g x_{iS+1} (i.e. the i^{th} vector is a near neighbour to the first vector) are extracted. Estimate $\hat{x}_{S+1}$ is now the average of all these subsequent points. Estimation then moves on to the next point and so on until all points are estimated and Γ_{test} is derived. The nearest neighbour approach ensures that the average is only based on vectors that approximate to that vector that immediately precedes the data point to be estimated. Thornhill *et al.* have described this process as an *H*-step (*H* equals to *S*) ahead predictor, but strictly speaking such a predictor should only be based on embedding vectors that have happened in the past, whereas off-line estimation can be based on vectors from both the past and the future. The reference probability distribution is parameterised by simulation: the r.m.s. errors of n surrogate data sets are analysed to obtain the reference mean, $\overline{\Gamma}_{surr}$, and the standard deviation, $\sigma_{\Gamma_{surr}}$. These *n* surrogate data sets are derived from the pre-processed data. The pre-processed data is first Fourier transformed, then *n* surrogate data sets are generated by inverse Fourier transforming the same power spectrum, but with different random phases.

The non-linearity test determines whether a time series could plausibly be the output of a linear system driven by Gaussian white noise, or whether its properties can only be explained as the output of non-linearity. Thornhill *et al.* relate non-linearity of a time series to the harmonic content of this data. If the plant acts as a mechanical low-pass filter, the process will filter out harmonics and process variable will become more sinusoidal further away from the source, therefore the non-linearity is strongest nearest to the source and the process variable with the largest *N* is likely to be located closest to the non-linear root cause.

Criticism 10

The non-linearity test approach has been successfully applied to some industrial cases, however, the assumption that the plant acts as a mechanical low-pass filter does not always hold. As will be discussed in Chapter 3, harmonics may be amplified through oscillation propagation routes and some loops further away from the source may contain more harmonics. Moreover, excessive non-Gaussian noise will degrade the performance because the surrogate data pertains to Gaussian random signal. Additionally sub-sampling the original data may lose useful information, as compression could have an effect on non-linearity (Thornhill *et al.* 2004).

Xia & Howell (2003a) applied spectral ICA to extract dominant independent components from spectral data. In the spectral ICA model, the rows of the data matrix, **M**, are single-sided power spectra $P(f)$ of the observations over a range of frequencies up to the Nyquist frequency (one-half of the sampling frequency):

$$\mathbf{M}=\begin{pmatrix} P_1(f_1) & \cdots & P_1(f_N) \\ \cdots & \cdots & \cdots \\ P_m(f_1) & \cdots & P_m(f_N) \end{pmatrix}$$

N frequency channels $\longrightarrow$

m process variables $\downarrow$

M is then decomposed into n dominant independent components:

$$\mathbf{M}=\mathbf{A}_{m\times n}\begin{bmatrix}\mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_n'\end{bmatrix}=[\mathbf{a}_1,\mathbf{a}_2\cdots\mathbf{a}_n]\begin{bmatrix}\mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_n'\end{bmatrix}=\begin{bmatrix}a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1}\end{bmatrix}\mathbf{y}_1'+\begin{bmatrix}a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2}\end{bmatrix}\mathbf{y}_2'+\cdots+\begin{bmatrix}a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n}\end{bmatrix}\mathbf{y}_n' \quad (2.17)$$

where $\mathbf{y}_j'$ is the j^{th} independent component (IC), **A** is an $m \times n$ real mixing matrix and $m \geq n$. An independent component is most dominant when its spectrum has a narrow-band peak. These components are related to the spectral data through significance indices (*SI*) $a_{i,j}$: a significance index of unity suggests the source of the j^{th} IC oscillation is most probably closest to the i^{th} process variable; a significance index of zero implies no correlation.

Criticism 11

Spectral ICA is very helpful in the identification of the root cause of low-frequency oscillations because in these cases the dominant IC is likely to be the fundamental harmonic and the spectra collected from the problem loop should display this fundamental most strongly, whereas other loops will also display harmonics. However, there are cases where the harmonics in loops, which do not contain the source, are attenuated, so that the significance index of the problem loop associated with the fundamental IC is unlikely to be the largest. Further explanation of such cases will be given in Chapter 5 where spectral ICA will be modified in such cases for root cause isolation.

2.5 Conclusions

Previous methods for the isolation of the root cause of an oscillation that are based on harmonic attenuation include the D- factor and Non-linearity index (Thornhill *et al.* 2001; Thornhill *et al.* 2003b)). In Spectral ICA (Xia & Howell 2005) the independent components align with harmonics also. Although reliant on the frequency response characteristics of harmonic propagation, all make assumptions about the attenuation of a plant, thus restricting their application. Researchers at Eastman have started to look at harmonic content for hardware problem diagnosis, for example Eastman's large-scale controller performance assessment system diagnoses process nonlinearities by finding harmonic peaks in the spectrum (Paulonis & Cox 2003), but there has not yet been a more thorough evaluation. Major contributions of this PhD research are its detailed harmonic propagation analysis and the development of harmonic evaluation indices which are helpful in the isolation of the source loop of plant-wide oscillations.

Chapter 3

Plant-Wide Oscillations

This chapter presents an overview of whole-plant oscillations, which is divided into the following sub-sections:

- their properties;
- other types of disturbances;
- their propagation.

Harmonic propagation is emphasized, because this is the main focus of this thesis.

3.1 Properties of Oscillations

Plant-wide oscillations might contain a single peak in their power spectrum or a few peaks at their fundamental and harmonic frequency channels. Moreover, the fundamental frequency might be either fixed or variant.

3.1.1 Bad-Tuning and Non-Linearity Induced Oscillations

Although there are other possible reasons that may cause sinusoidal or non-sinusoidal oscillations, the most common experience is that bad tuning may cause sinusoidal oscillations, and that valve non-linearity such as stiction, hysteresis may cause oscillations with harmonics.

3.1.1.1 Bad Tuning Induced Oscillations

Sinusoidal oscillations are common in over-tuned loops which behave in a resonant manner. A single peak in the power spectrum is therefore taken to be a signature of a tuning problem (Thornhill & Hägglund 1997). Figure 3-1 gives a unity feedback control system without setpoint change. When the system is noise free or disturbance f is a stationary random noise, a sustained oscillation happens when the controller gain is increased such that the open loop gain is unity when the open loop phase is -180°. The Nyquist plot (Figure 3-2) shows what happens in the frequency response of the system. The oscillation frequency is called the ultimate frequency (ω_u). A badly tuned controller is particularly visible in nonlinear plants where a change in operating point might result in a loop gain which is too-high (Hägglund 1995). However, controllers in process-control plants are often tuned conservatively, and bad tuning is not the most likely cause of oscillations.

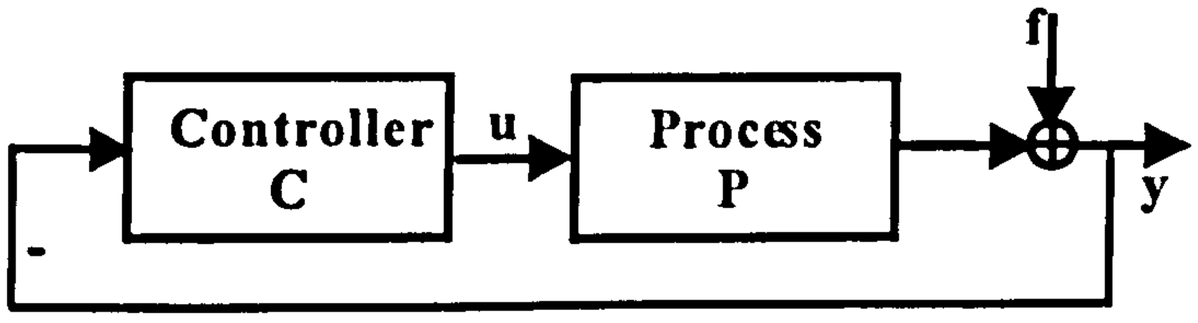


Figure 3-1: A unit feedback control loop

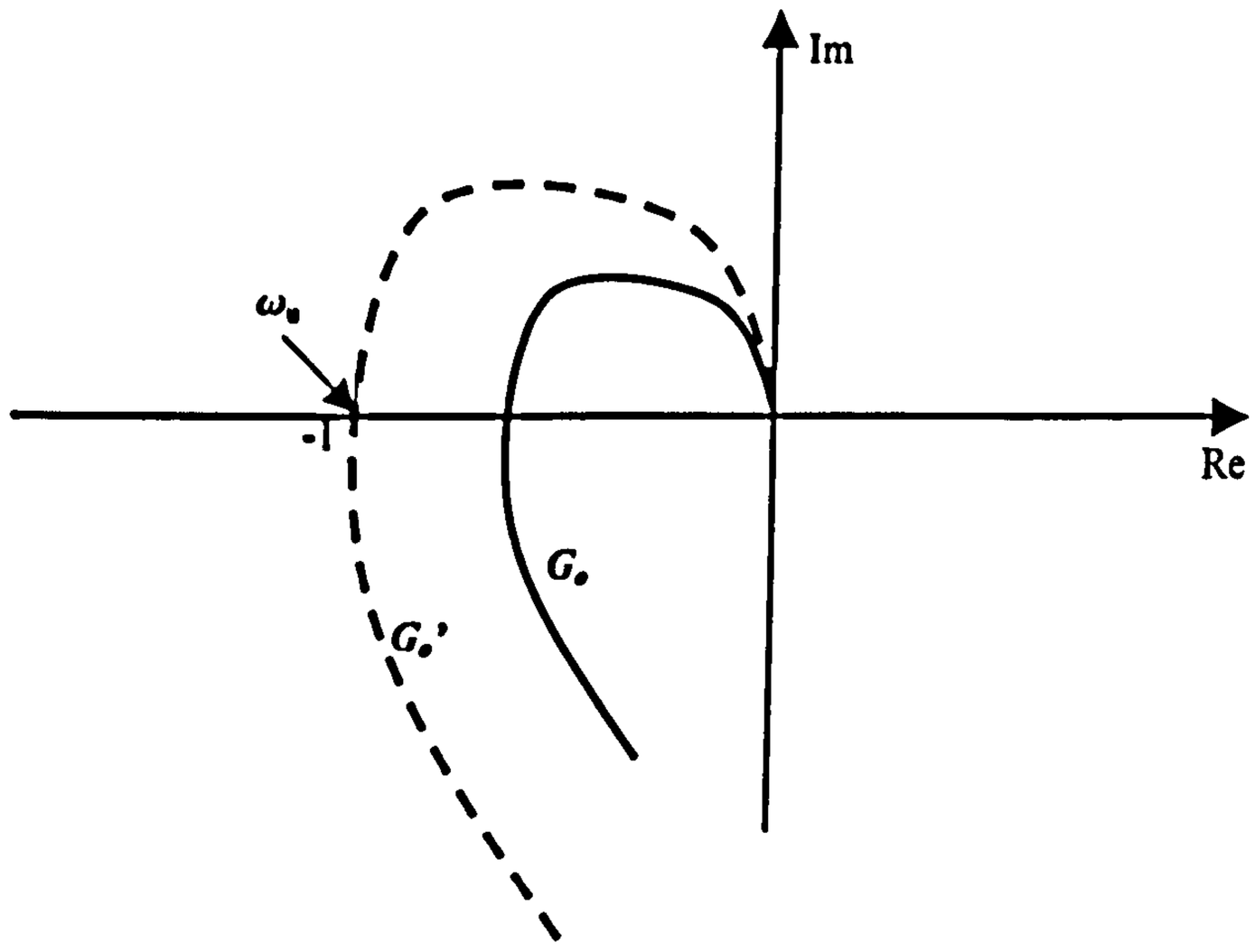


Figure 3-2: Nyquist plots of open loops: open-loop transfer function G_o=CP; G_o' is a badly-tuned open loop obtained by increasing the gain

3.1.1.2 Valve Non-Linearity Induced Oscillations

Among many types of non-linearities in control valves, stiction and hysteresis/backlash are the most common reasons for oscillations in control loops (Hagglund 2002; Rengaswamy *et al.* 2001).

Valve Stiction Model

Many studies (Choudhury *et al.* 2004; Horch & Isaksson 1998; Olsson *et al.* 1998; Olsson & Astrom 2001; Taha *et al.* 1996; Wallen 1997) have been carried out to define and detect static friction or stiction. The commonly used valve friction model is the so-called physical model (Choudhury *et al.* 2004). For a pneumatic sliding stem valve, the force balance equation based on Newton's second law can be written as:

$$M\frac{d^2x}{dt^2} = \sum Forces = F_a + F_r + F_f + F_p + F_t \tag{3.1}$$

where M is the mass of the moving parts, x is the relative stem position, $F_a = Au$ is the force applied by the pneumatic actuator where A is the area of the diaphragm and u is the actuator air pressure or the valve input signal, $F_r = -kx$ is the spring force where k is the spring constant, $F_p = -\alpha\Delta P$ is the force due to the force imbalance across the value where α is the plug area imbalance and ΔP is the fluid static pressure difference across the valve, F_t is the extra force required to force the valve to be into the seat and F_f is the friction force. F_t and F_p have a negligible contribution to the model and hence are ignored. The friction model includes static and moving friction:

$$F_f = \begin{cases} -F_c\,\mathrm{sgn}(v) - vF_v & if \quad v \neq 0 \\ -(F_a + F_r) & if \quad v = 0 \quad and \quad |F_a + F_r \leq F_s| \\ -F_s\,\mathrm{sgn}(F_a + F_r) & if \quad v = 0 \quad and \quad |F_a + F_r > F_s| \end{cases} \tag{3.2}$$

where F_c is Coulomb friction, F_v is viscous friction and F_s is the maximum static friction.

The disadvantage of the physical model of a control valve is that several parameters must be known. Manufacturer's values are listed in (Choudhury *et al.* 2004). Choudhury *et al.* (2004) developed an empirical model for valve stiction that produces input-output (I/O) behaviour similar to that of more complicated physics-based models. The I/O characteristics of a sticking control valve are presented in Figure 3-3. It consists of four components: deadband, stickband, slip jump and the moving phase. When the valve comes to a rest or changes the direction at point A, the valve sticks. After the controller output overcomes the deadband (AB) and the

stickband (BC) of the valve, the valve jumps to a new position (point D) and continues to move. The deadband and stickband represent the behaviour of the valve when it is not moving, though the input to the valve keeps changing. Slip jump represents the abrupt release of potential energy stored in the actuator chambers due to high static friction; this is converted into kinetic energy and the valve starts to move. Once the valve slips, it continues to move until it stops, then eventually sticks again.

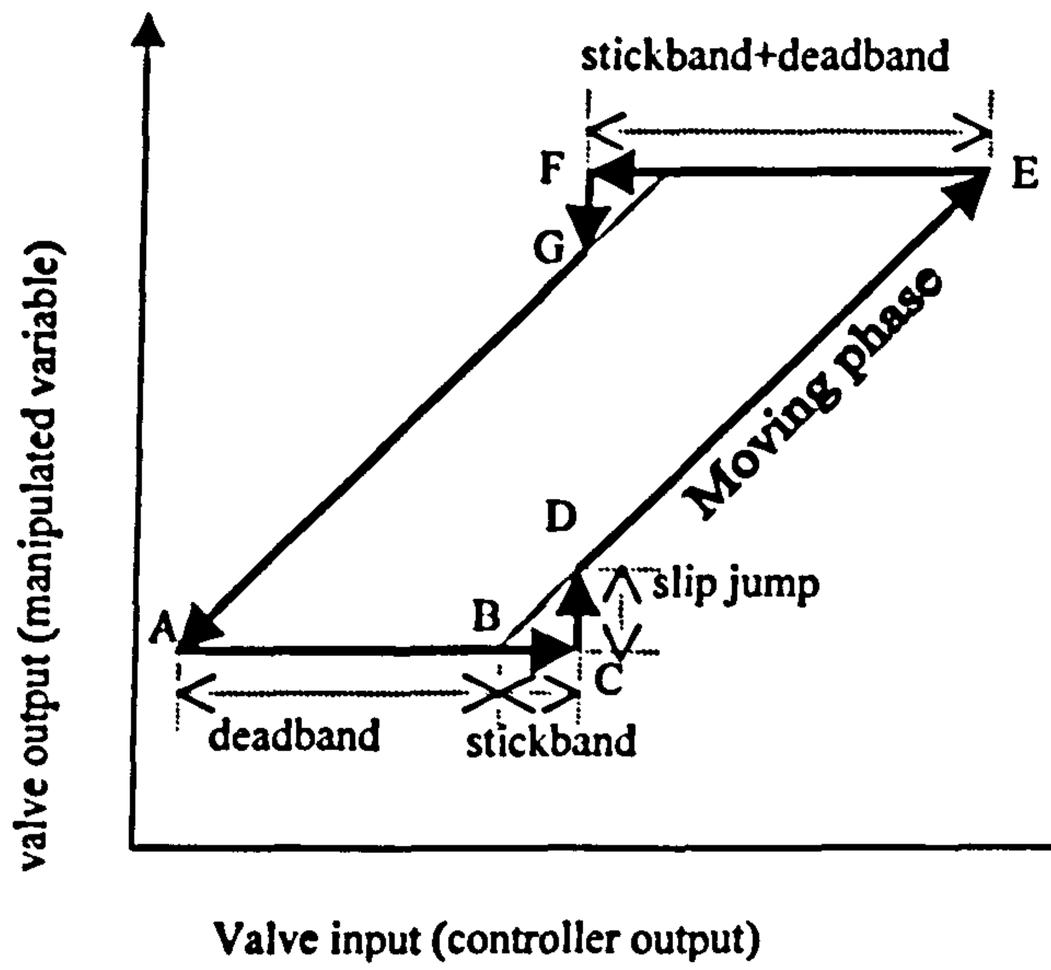


Figure 3-3: Typical input-output behaviour of a sticky valve

Based on the observed valve stiction behaviour, a data-driven model has been established by Choudhury *et al.* (2004) which produces the same behaviour as the physical model. The model consists of two parameters –namely the size of the deadband plus stickband *s* and slip jump *j*. If the direction of the slope of the input signal changes or remains zero for two consecutive instants, the valve is assumed to be stuck and does not move. When the cumulative change of the input signal is

more than the amount of the deadband plus stickband (s), the valve slips and starts moving. The output is calculated using the equation:

$$output = input - sign(slope) * (s - j)/2 \tag{3.3}$$

and depends on the type of stiction present in the valve:

- deadband: j=0, represents the pure deadband case without any slip jump;
- stiction (undershoot): when $j<s$, the valve output can never reach the valve input;
- stiction (no offset): when $j=s$, the algorithm models pure stick-slip behavior;
- stiction (overshoot): when $j>s$, the valve output overshoots the valve input due to excessive stiction.

Of interest here is the sinusoidal response of this data-driven valve stiction model, and in particular in its harmonic content. Its Fourier series expansion is performed in Appendix A: only odd harmonics are present in the valve output signal for a pure sinusoidal signal input. This is consistent with the result of Fourier analysis of a valve with deadband published by Wilton (2000), who assumed j=0. Note that Choudhury *et al.*(2004) analysed only the fundamental Fourier component.

Valve Backlash Model

Backlash is present in every mechanical system where the driving element is not directly connected with the driven element (Horch A. 2000). Figure 3-4 shows the time domain behaviour for a backlash element. When the input changes direction, an initial change in input has no affect on the output. The amount of side-to-side play is referred to as the deadband (d). The deadband is centered about the output. There are three modes:

– disengaged, where the input does not drive the output and the output remains constant;

– engaged in a positive direction, where the input is increasing (has a positive slope) and the output is equal to the input minus half the deadband width;

– engaged in a negative direction, where the input is decreasing (has a negative slope) and the output is equal to the input plus half the deadband width.

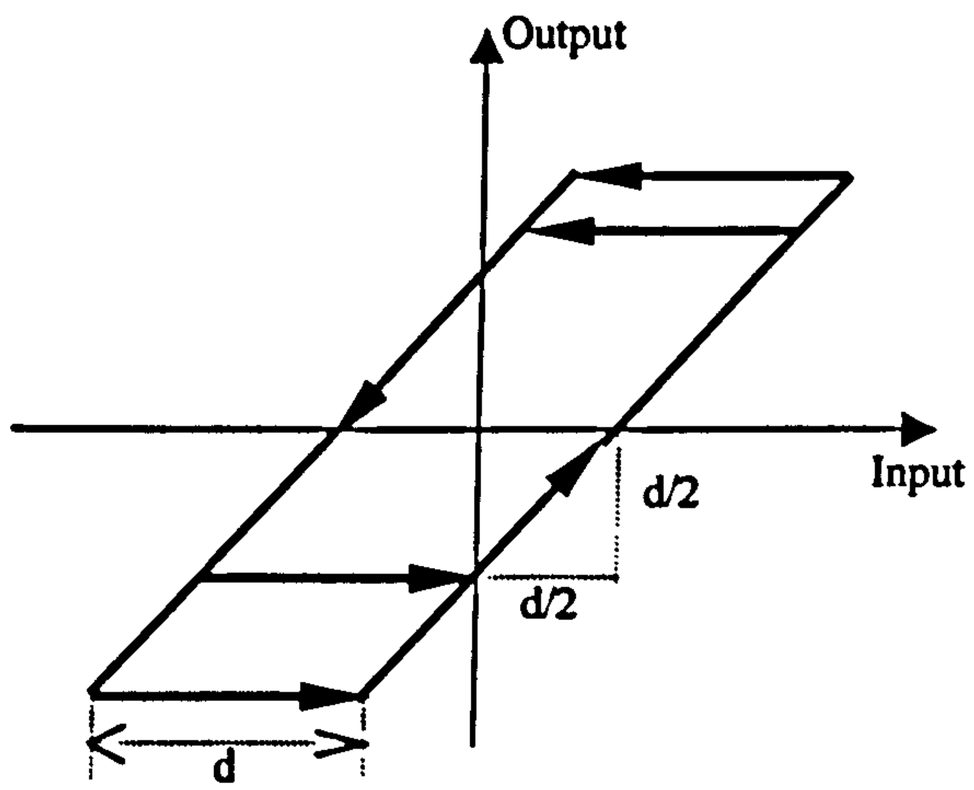


Figure 3-4: Time-domain characteristic of a backlash element

The Fourier series expansion of the output of the backlash model, when forced by a pure sinusoidal input, should be the same as a deadband model (i.e. the stiction model with j=0) because the sinusoidal input has fixed amplitude and phase. Again no even harmonics are seen in the output and only odd harmonics are present.

3.1.1.3 The Shape of Oscillations

Several papers have been published on detecting and measuring valve stiction or hysteresis by inspecting the shapes of the control error and the controller output signals during sustained oscillations (Gerry & Ruel 2001; Rengaswamy *et al.* 2001;

Rossi & Scali 2005; Ruel 2000; Ruel 2002; Singhal & Salsbury 2005). They have suggested that the controller output should be a saw-tooth or triangular wave for a sticking or hysteresis valve and a sinusoid for an aggressive controller. Symmetric triangular waveforms contain only odd harmonics, however, in practice even harmonics are also observed. A possible reason for this is that triangular waves that are observed in controller outputs are asymmetric and these contain both even and odd harmonics. Figure 3-5 shows examples of an asymmetric waveform observed as a result of valve stiction and a sinusoidal waveform observed as a result of aggressive control (Singhal and Salsbury 2005). The latter was analysed to distinguish between the shapes of the signals caused by an aggressive controller and those by a sticking valve using the ratio of the areas before and after the peak of the control error signal. Appendix B gives the Fourier analysis of an asymmetric triangular waveform.

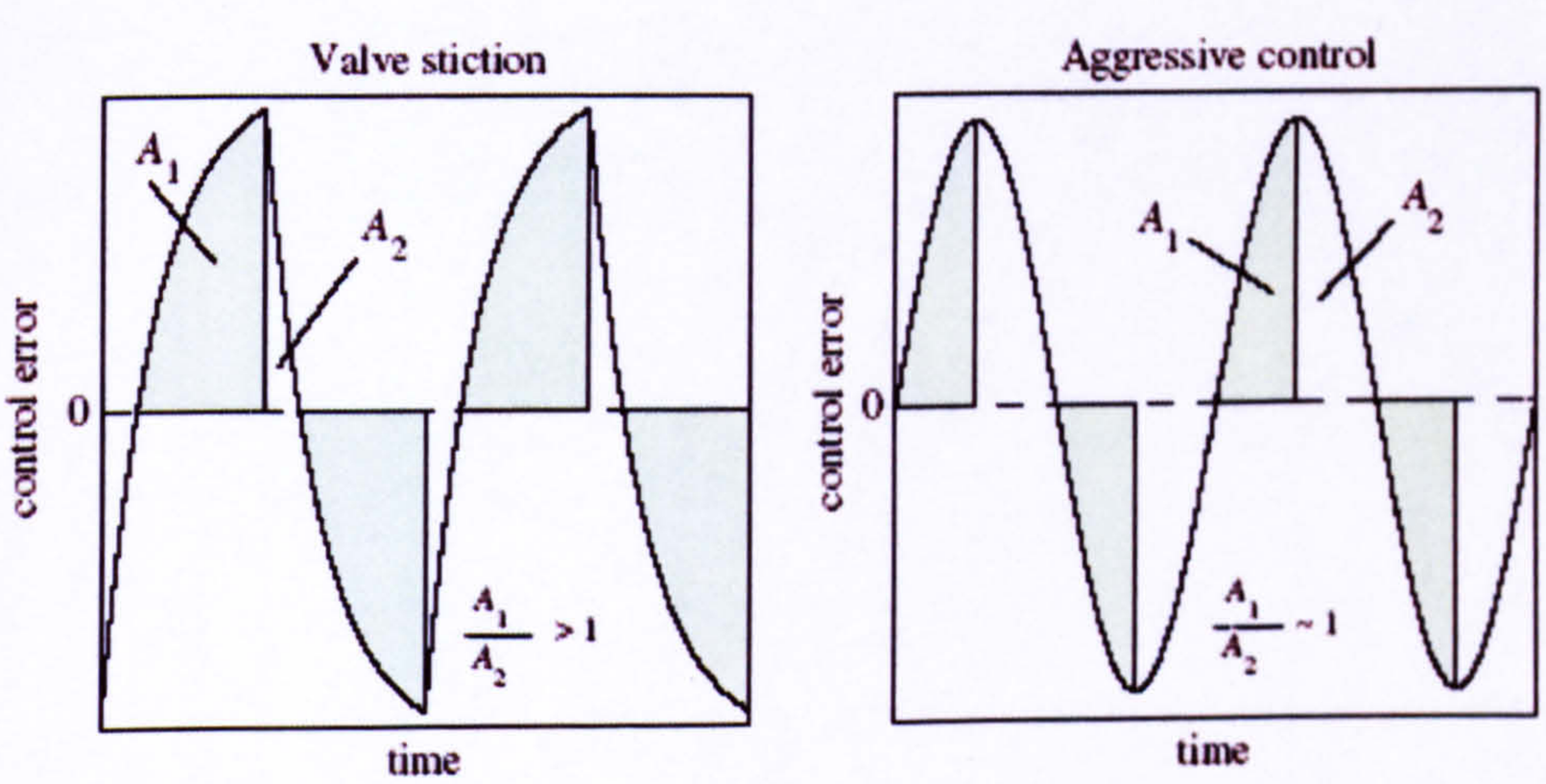


Figure 3-5: Control error signal shapes for valve stiction and aggressive control

Possible reasons for the shift in oscillation phases so that asymmetric waveforms are observed in practice are:

- process dead time;
- valve dead time;
- poor valve position controller;
- random load disturbance;
- valve sticking stops between points D and E in Figure 3-3 while travelling in the same direction due to low or zero velocity (EnTech 1998; Kano *et al.* 2004).

Note that the dead time in a control loop is inevitable because the valve response always has dead time (EnTech 1998)which affects the phases of the harmonics in the signal. When the phase-shifted signal goes through the non-linear valve, the time when the valve sticks and slips will shift accordingly so that the resulting valve output *m(t)* will shift its phase and will become asymmetric. Singhal and Salsbury (2005) derived the analytical expressions of control error signal shapes from first and second-order plus time-delay plants with valve stiction, which also showed that when the effective time-delay is nonzero, the shape is asymmetric.

To demonstrate these observations a simulation was performed with a process transfer function of $\frac{2}{(3s+1)(10s+1)}e^{-3s}$, a PI controller of $0.5\left(1+\frac{1}{12.3s}\right)$ and with coulomb friction . Figure 3-6 shows the time records and power spectra of the process variable (PV) and controller output (OP). The process dead time has shifted the phase of the controller output wave and both even and odd harmonics are seen in the PV and OP measurements.

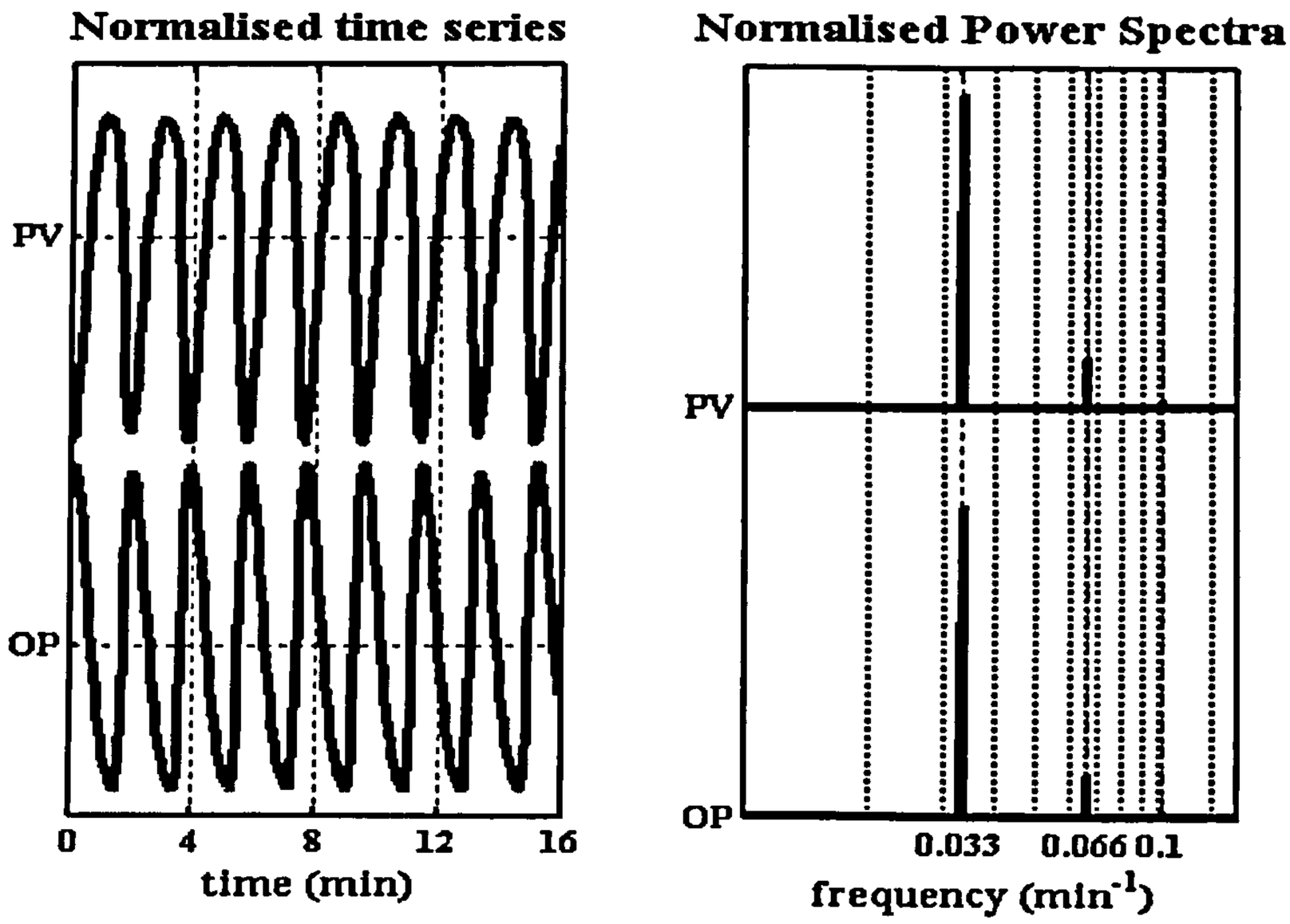


Figure 3-6: The shapes and power spectra of OP and PV

3.1.2 Fixed and Variable Frequency

In theory, a valve nonlinearity induced oscillation will contain a single fundamental frequency plus its harmonics. However in practice process variation or controller retuning will cause the fundamental frequency and its harmonics to alter, resulting in a variable-frequency oscillation. Both situations are analysed here by referring to Nyquist plots of the open loop transfer function and of a valve nonlinearity describing function.

3.1.2.1 Fixed Frequency

The literature often models the nonlinear elements as a describing function to give insight into the formation of the stable limit cycle (Figure 3-7). The condition for an oscillation to arise in a negative feedback loop when the loop gain is –1 is

$$G_o(i\omega) = -\frac{1}{N(X_m)} \tag{3.4}$$

where $G_o(i\omega)$ is the open loop frequency response which includes the controlled system and the controller, and $N(X_m)$ is the describing function, which is defined as

$$N = \frac{Y_f}{X} \tag{3.5}$$

where $X = X_m \sin\omega t$ is the input to the nonlinear element and Y_f is the magnitude of the fundamental component of the output from the non-linearity. The frequency (point A) at which the loci of $G_o(i\omega)$ and $-1/N(X_m)$ cross (Figure 3-7) should then be the fundamental frequency of the oscillation that arises.

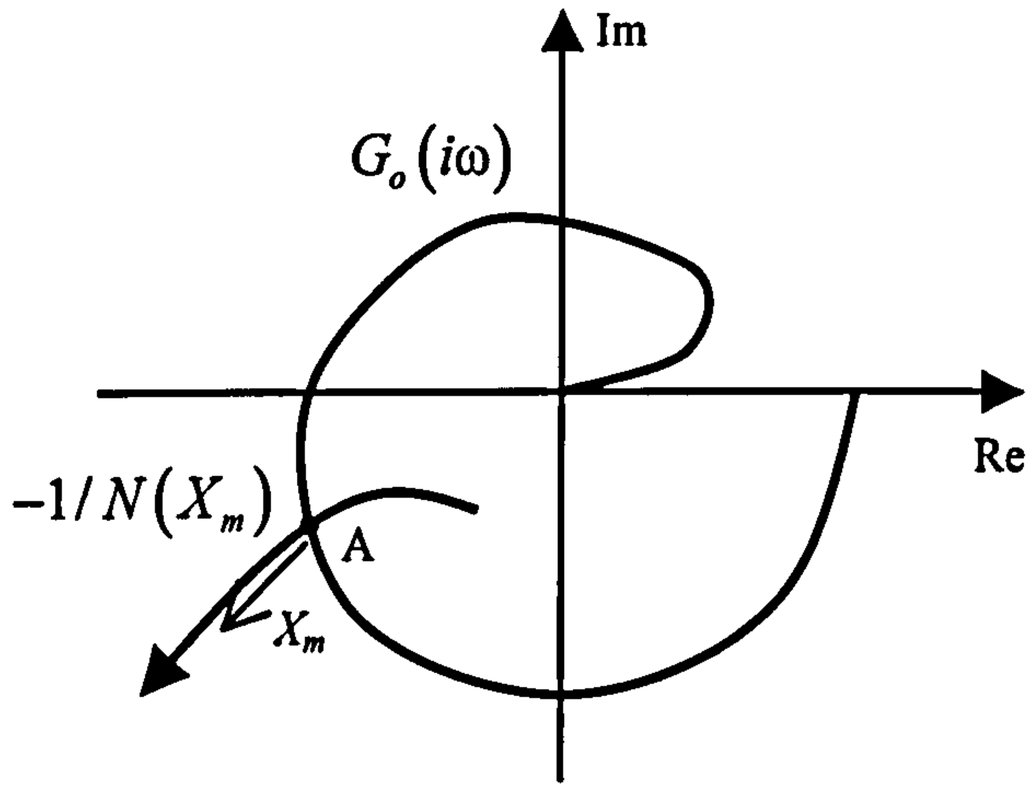


Figure 3-7: Existence of limit cycles

3.1.2.2 Variable Frequency

It can be seen from Section 3.2.2.1 that the fundamental frequency of the limit cycle depends on the process dynamics, the controller and valve characteristics. Either process dynamics or controller change will change the point of intersection of the

open loop transfer function G_o with the describing function, and hence change the fundamental frequency of the oscillation.

For example Figure 3-8 shows the effect of changing the open loop Nyquist plot from G_o to G_o' for instance by de-tuning the controller gain. The oscillation point changes from A to A' correspondingly, and the frequency will change from ω_1 to ω_2. Similarly the fundamental frequency may vary when process variations change the Nyquist plot of G_o.

A key feature of a variable fundamental frequency is in the broadening of the harmonic peaks in their power spectrum. An example of this can be seen in Figure 3-9, which was generated by analysing $\sin(0.01t\beta(t))$, where $\beta(t)=1+0.1\sin(0.1t)$.

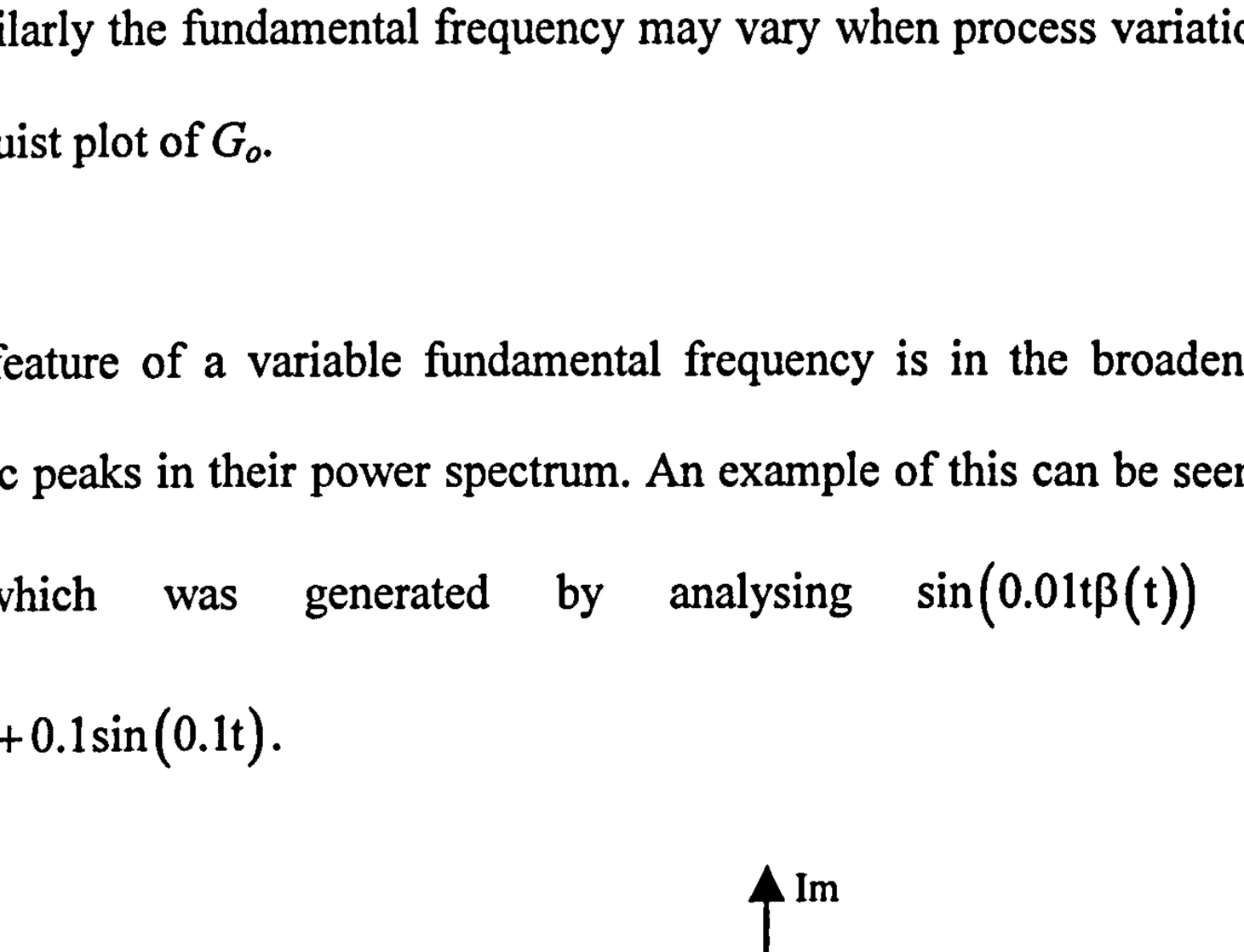


Figure 3-8: Nyquist plots of open loops: open-loop transfer function $G_o=CP$; G_o' is a de-tuned open loop; $N(X_m)$ is a describing function of the non-linear component

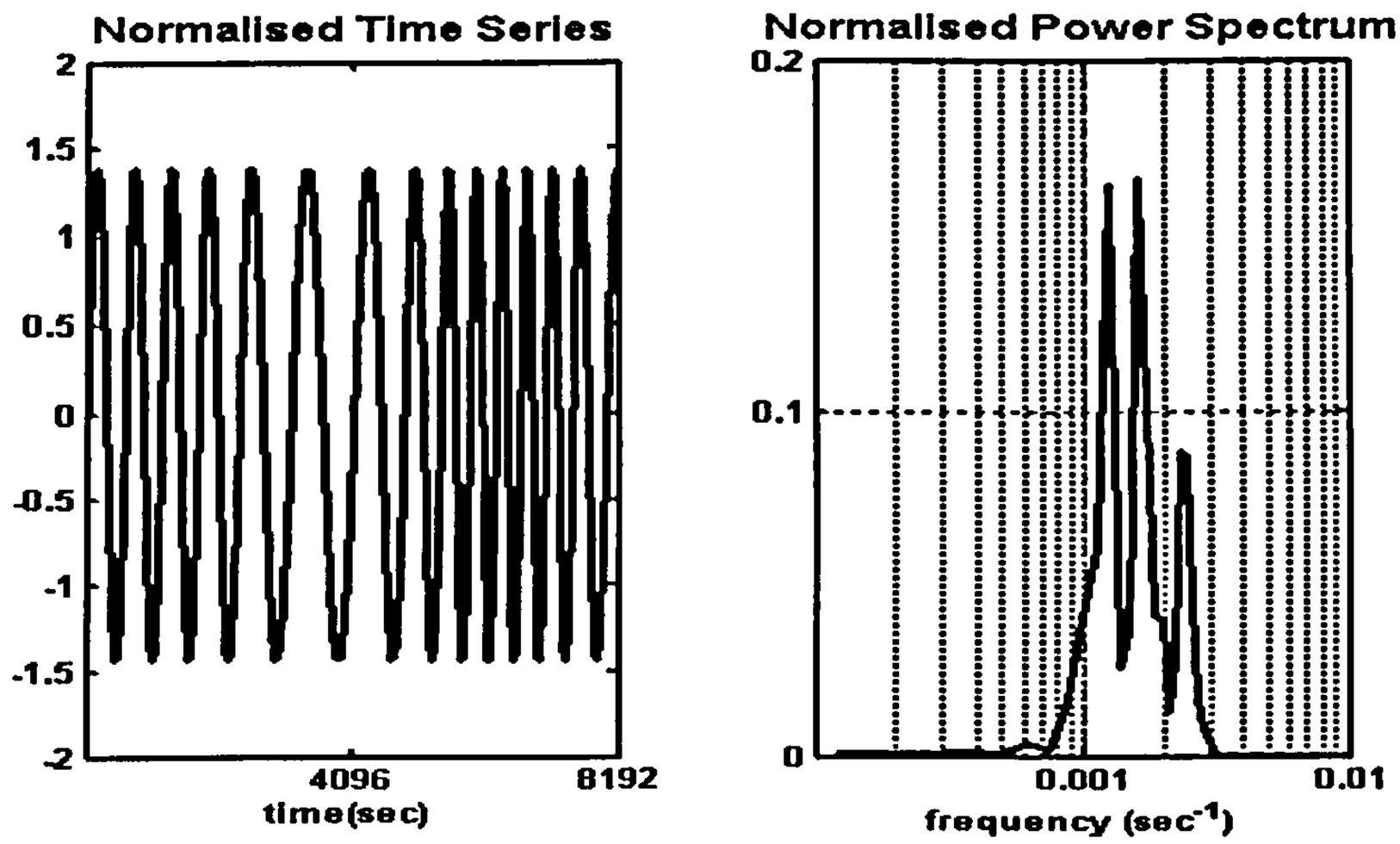


Figure 3-9: Time series and power spectrum of a variable frequency signal

3.2 Other Types of Disturbances

Besides these oscillatory disturbances, regular or non-periodic disturbances can also be observed in process plants. These can result from disturbances at plant boundaries; sensor drift; drift in system kinetics; controller induced step changes; non-stationary disturbances due to tight coupling in control loops; etc.

Xia & Howell (2003b) have classified loop status as either (1) well-behaved and in steady state; (2) well-behaved but with controller compensation; (3) undergoing a short-term transient; (4) undergoing a trend that is disturbed in some non-stationary manner; (5) cycling at a relatively low frequency; (6) cycling at a fundamental frequency similar to the natural frequency of the loop; (7) out of control (critical). Non-periodic disturbances could be fitted into any of the classifications (2), (3), (4) and (7). The detection of these disturbances can be achieved by analysing qualitative and quantitative *loop status statistics* (Xia & Howell 2003b).

3.3 Harmonic Propagation

The previous chapter reviewed time-domain based and frequency-domain based methods for isolating the root cause of non-linearity induced oscillations. All of these are based on the premise that a non-linear hardware fault in a control loop will introduce non-sinusoidal oscillations in the controller outputs and process variables, and that these oscillations contain power spectra at integer multiples of a fundamental frequency, i.e., harmonics. They are based on the heuristic that physical systems act as low-pass filters and hence a recorded measurement further away from the non-linear fault will contain fewer harmonics than a measurement near the root cause. However, the propagation of the oscillation from one control loop to another control loop cannot be simply deemed as propagation through a physical system that is inherently lowpass, for instance because the commonly used PI/PID controller will introduce harmonic amplification and thus the propagation route from one loop process variable to another loop process variable can not necessarily be modelled as a low-pass filter.

Plant measurement sensors are not only installed as components in control loops but also to provide additional indicators of plant operation. The propagation of an oscillation to other parts of the plant can therefore be observed in the various control loops or in data collected by these additional indicators. The root cause might derive from a non-linear valve in a SISO loop or a cascade control system, which could be part of the regulatory layer in a multivariable control scheme providing that the set-point changes do not reflect oscillatory behaviour. The power spectra of its controller output and valve opening measurement records would then

contain a fundamental plus harmonic content. Power spectra of measurements recorded at other locations around the loop would possess different harmonic content, because of plant filtering. In this section the harmonic propagation throughout the plant is analysed. Loop to loop propagation, loop to indicator propagation and cascade loop propagation are all considered.

3.3.1 The Loop to Loop Propagation Shapes

The propagation of the oscillation to other parts of a plant and, in particular, to other control loops clearly depends on the paths of interaction that exist in the plant. There is a spectrum of possible scenarios: at one extreme there are tightly coupled plants where all loops affect all loops, and at the other extreme there are plants with serial, daisy-chained arrangements where one loop disturbs another, which disturbs another and so on. An example of the first extreme might be a reactor with level, pressure and temperature control loops, which respond in tandem. If the source of the oscillation was to be in one of the loops, the tight coupling could result in all loops behaving almost identically, making it impossible to isolate the cause. The analysis here assumes that any coupling is relatively weak and is performed by viewing the source oscillations as external disturbances to other loops. To do this, a plant is hypothesised (Figure 3-10) that contains a number of interacting control loops where Loop *s* contains the source of the oscillation. This oscillation will then propagate to other loops. Propagation from loop *s* to loop *k* is considered here: C_s and C_k represent the controllers in the two loops, P_{ss}, P_{ks}, P_{sk} and P_{kk} process transfer functions, U_s and U_k controller outputs and Y_s and Y_k controlled process variables.

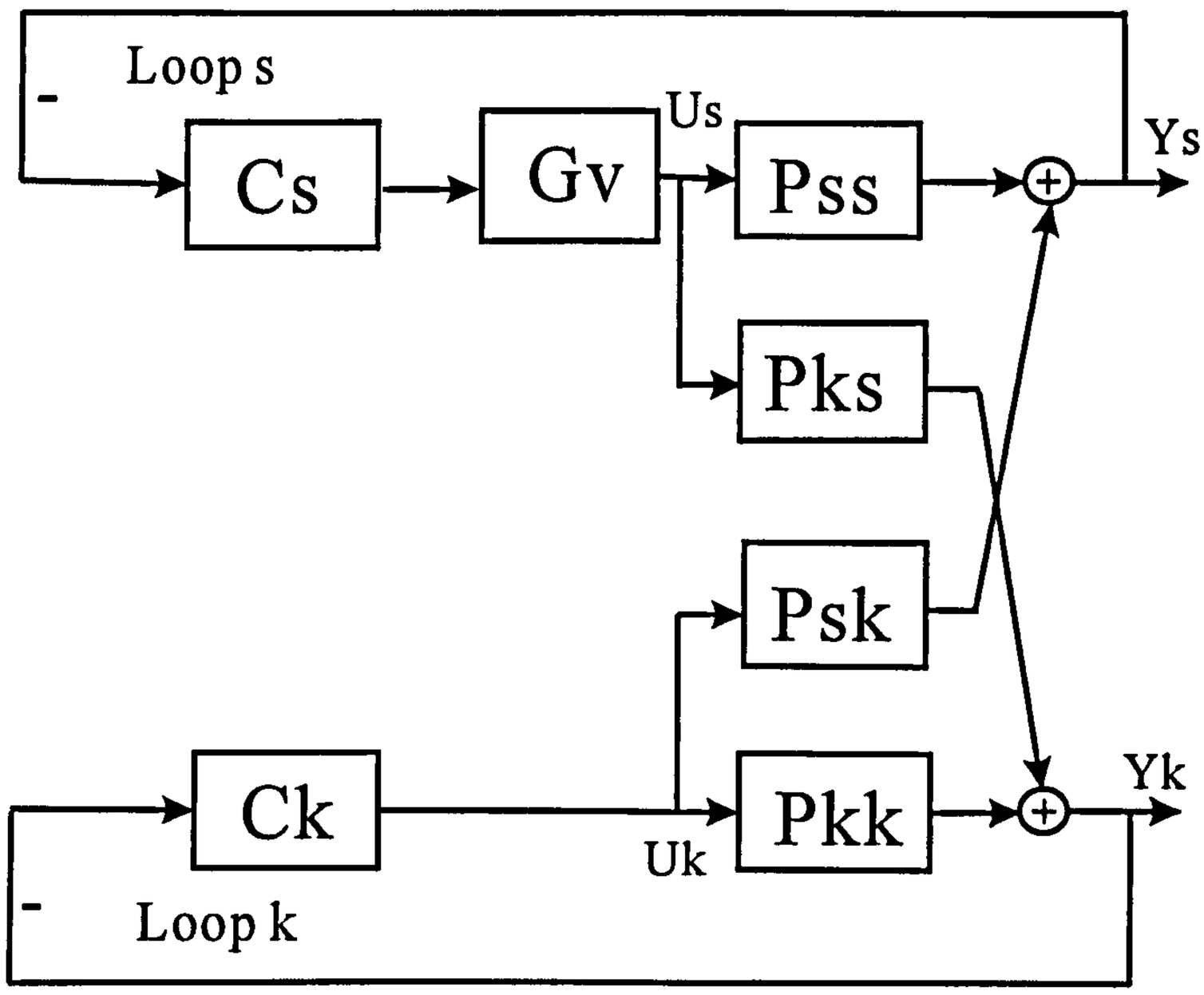


Figure 3-10: Block diagram for two interacting control loops

Harmonic analysis assumes that the plant is operating in steady state. Thus the analysis here is restricted to the case where there are no set-point changes or disturbances acting on these loops, so Y_s and Y_k should remain relatively constant unless one of the loops limit cycles. Suppose that Loop s limit cycles at a fundamental frequency ω_l as a result of a valve non-linearity, then the equivalent process faced by U_s (i.e. $G_s : Y_s = G_s U_s$) can be derived (Zhang *et al.* 2002) as

$$G_s = P_{ss} - \frac{P_{sk} P_{ks} C_k}{1 + C_k P_{kk}}$$

i.e.
$$Y_s = \left[P_{ss} - \frac{P_{ks} P_{sk} C_k}{1 + C_k P_{kk}} \right] U_s \qquad (3.6)$$

The dynamics of the open-loop transmission from U_s to Y_s, $G_s(s)$, is known as the effective open-loop process (EOP) (Huang *et al.* 2003). Similarly the propagation of the oscillation from U_s to Y_k can be described by:

$$Y_k = \frac{P_{ks}}{1+C_k P_{kk}} U_s. \tag{3.7}$$

The ratio, $r(\omega)$, of the magnitudes of the sinusoidal components observed at Y_s relative to that observed at Y_k at any frequency ω is then given by:

$$r(\omega) \equiv \frac{|Y_k|}{|Y_s|}\Big|_\omega = \left| \frac{P_{ks}(j\omega)}{P_{ss}(j\omega)+C_k(j\omega)\left[P_{ss}(j\omega)P_{kk}(j\omega)-P_{sk}(j\omega)P_{ks}(j\omega)\right]} \right| \tag{3.8}$$

Equation (3.8) can also be written as

$$r(\omega) = \left| \frac{P_{ks}(j\omega)}{G_s(j\omega)\left[1+C_k(j\omega)P_{kk}(j\omega)\right]} \right| \tag{3.9}$$

The plot of $20\log_{10} r(\omega)$ versus $\log_{10}\omega$ is informative and in recognition of this it is given a name, the *log-ratio* plot of loop s to loop k. Its shape is dependant on the sub-plots of $20\log_{10}\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ *vs.* $\log_{10}\omega$ and $20\log_{10}\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$ *vs.* $\log_{10}\omega$.

Analysis of the combination of these sub-plots suggests that the shape of the log-ratio plot can take one of four possibilities:

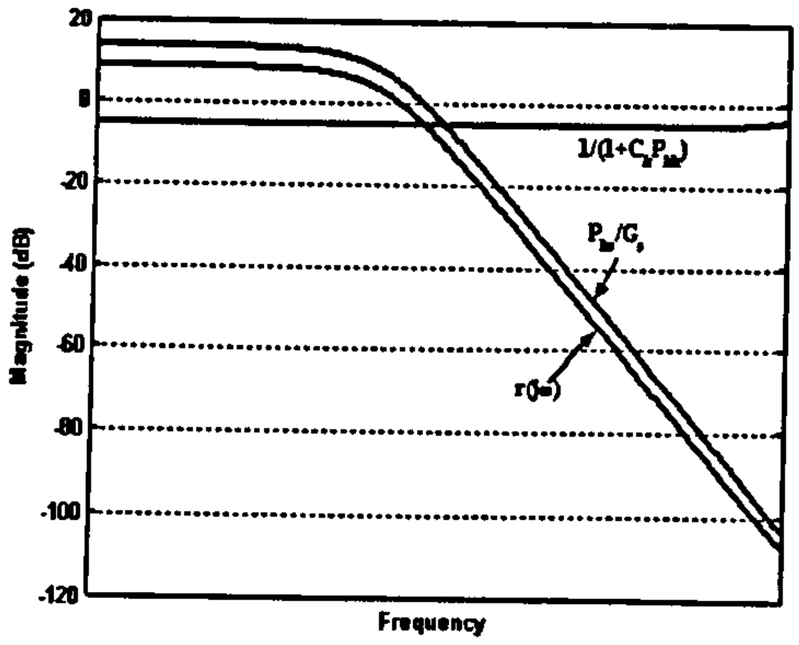
(a) a plateau at low frequencies and a negative slope at high frequencies

(b) a positive slope at low frequencies and a plateau at high frequencies

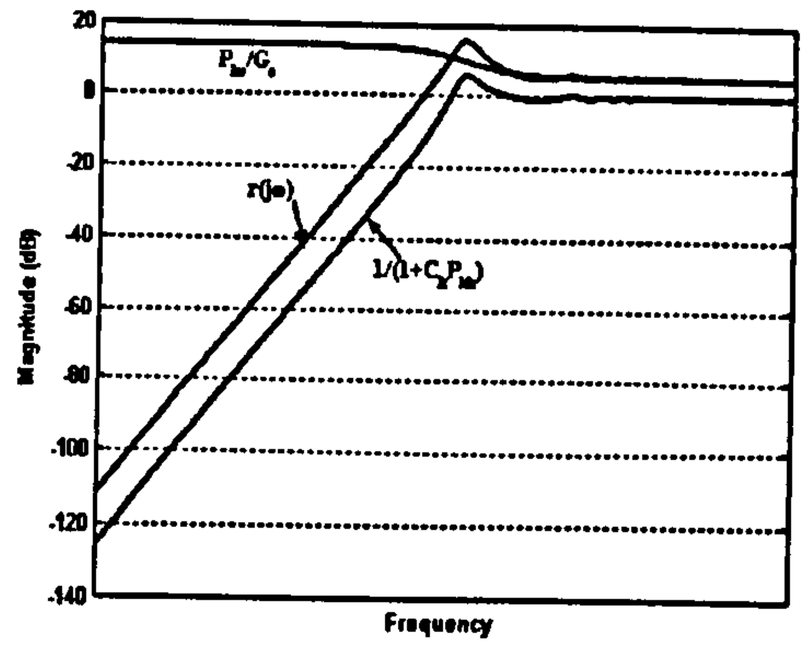
(c) a plateau at both low and high frequencies

(d) a positive slope at low frequencies and a negative slope at high frequencies

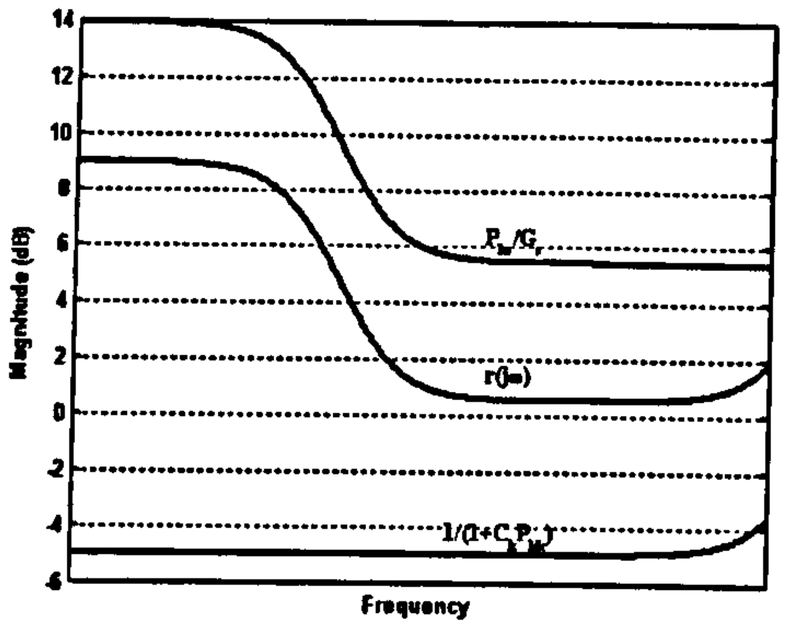
Figure 3-11 shows examples of these four possibilities, including the sub-plots from which the plots were constructed.



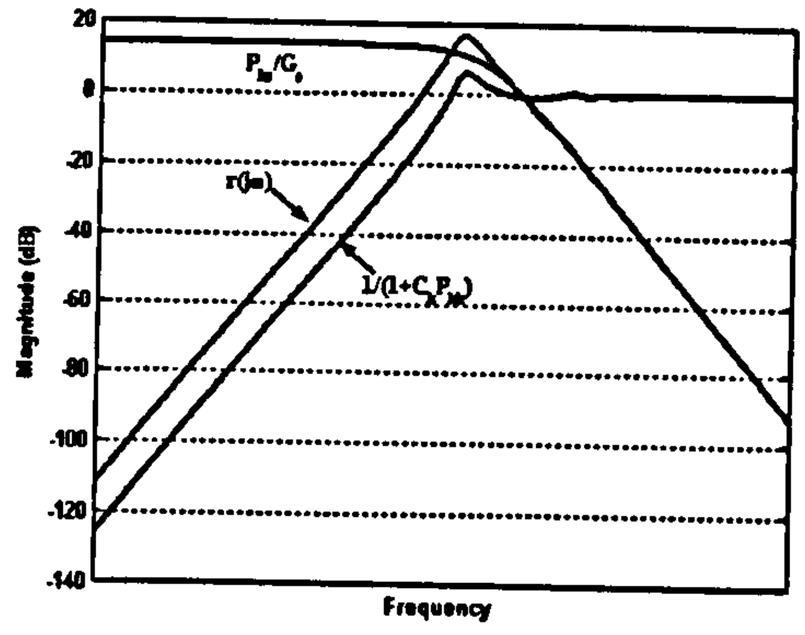
(a)



(b)



(c)



(d)

Figure 3-11: Typical log-ratio plots of $\left|\frac{P_{ks}(j\omega)}{G_s(j\omega)}\right|$, $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$ and r(ω)

A low frequency asymptote with a positive slope will be observed ((b) and (d)) if C_kP_{kk} contains one or more poles at the origin (e.g. as with a PI/PID controller or P controller with an inherent integral mode in the process P_{kk}), whereas the low frequency asymptote is likely to be horizontal ((a) and (c)) if C_kP_{kk} contains no poles at the origin (e.g. as with a P controller with no integral mode in the process). A negative slope at high frequencies ((a) and (d)) is to be expected, because it is likely that the order of P_{ks} > order of G_s: G_s is influenced, predominantly, by $P_{ss}(s)$ and transfer function $P_{ks}(s)$ is likely to contain components that are similar to those contained in $P_{ss}(s)$, because elements of both represent common physical processes. The orders of $P_{ks}(s)$ and $G_s(s)$ might be the same if the two loops are tightly

coupled, in which case the high frequency asymptote will be horizontal. Thus the most common shape is likely to be (d), i.e. a *bell*, because usually process plant interactions are low pass and controllers have integral action. Shape (a) might be relevant for plants with P controllers whereas shapes (b) and (c) might be relevant in tightly coupled plants. Shape (d) is the most common and hence the focus of the next section.

3.3.2 Analysis of the Bell Shape

Although the frequency at which the apex of the 'bell' occurs must be influenced by the cut-off frequencies of the two terms $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ and $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$, it has been found by looking at examples that this frequency is predominantly affected by the cut-off frequency of the target loop disturbance transfer function $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$. It is found in practice that the apex of the bell is predominantly determined by the resonant frequency of $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$ and is not shifted by the dynamics of $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ greatly. An insight into this can be gained by looking at the models used to generate Figure 3-11(d). Sub-plot $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ will only affect the peak if its cut-off frequency is relatively low and its negative slope is steeper than the positive slope of $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$. Suppose that the slope

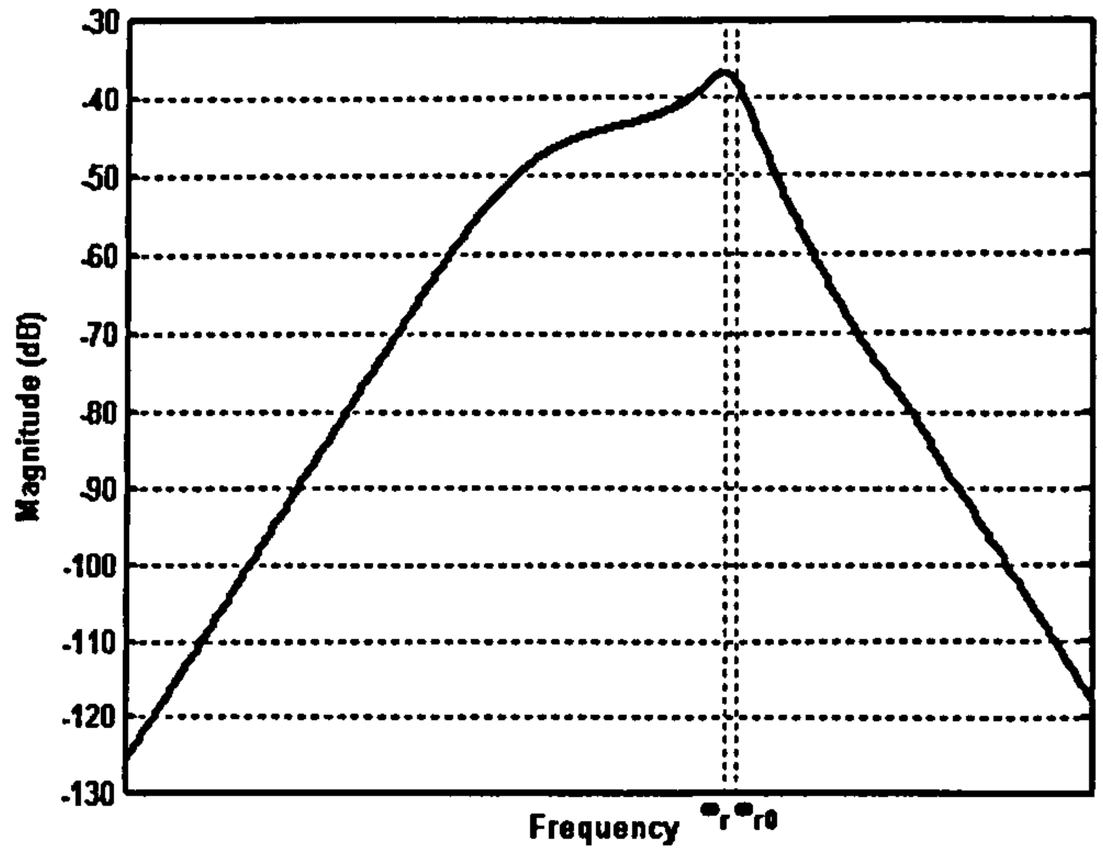
before the apex frequency (ω_{r0}) of $\left|\frac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$ is 20n dB/dec, then the rank of $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ must be at least (n+1) to shift the apex frequency of the bell (ω_r). In addition (n+1) of the cut-off frequencies of $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ must be less than ω_{r0} i.e. at least (n+1) time constants must be larger than $2\pi/\omega_{r0}$. A process with such large time constants is likely to have filtered out the oscillations observed so that the loop is unlikely to be on the list of candidate source loops, and hence would not be subject to the procedure proposed here.

To demonstrate this, a simple example is constructed assuming that the open loop transmission G_s can be approximated by either a FOPDT or SOPDT model (Huang *et al.* 2003):

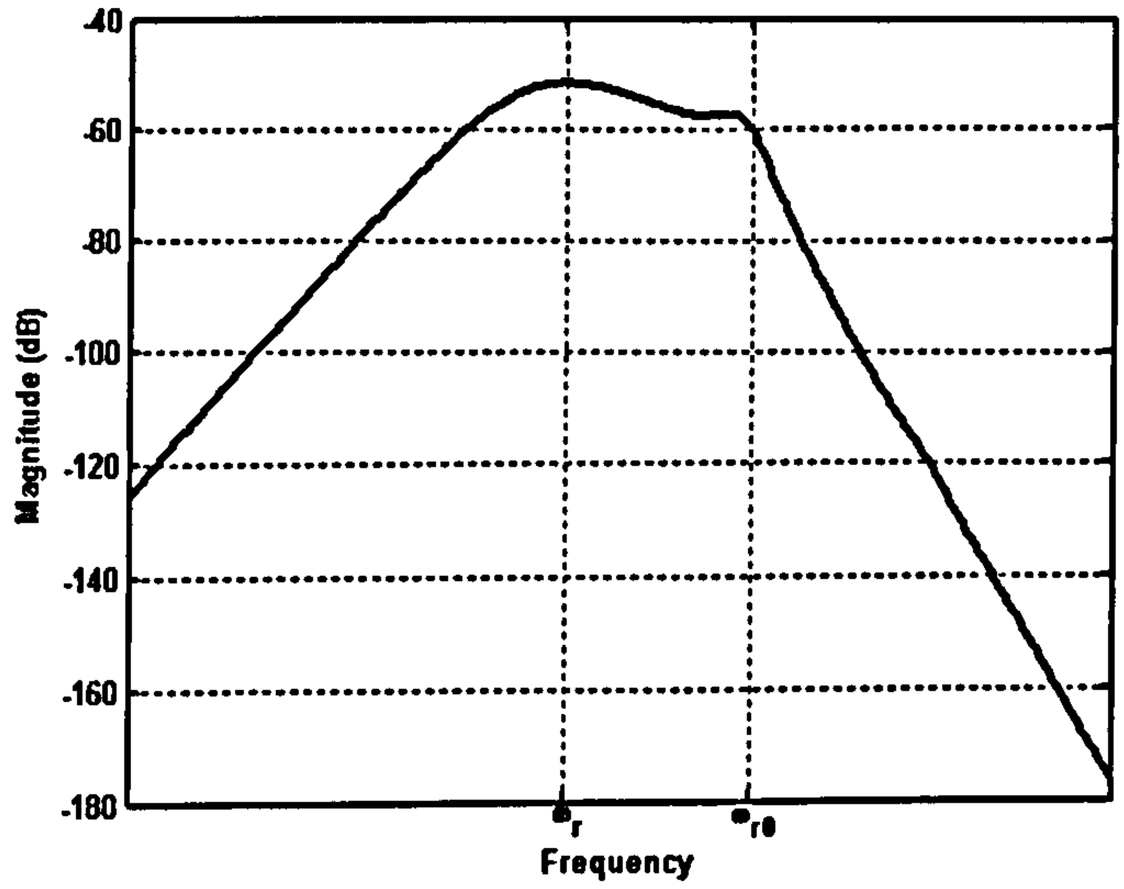
$$G_s^* = \frac{k_p e^{-\theta s}}{\tau s+1}\ ,\ G_s^* = \frac{k_p(\tau_c s+1)e^{-\theta s}}{(\tau_a s+1)(\tau_b s+1)} \text{ or } G_s^* = \frac{k_p(\tau_1 s+1)e^{-\theta s}}{\tau^2 s^2+2\tau\zeta s+1}.$$

Term $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ can now be approximated by $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|} \approx \left|\frac{k}{(\tau_1 j\omega+1)(\tau_2 j\omega+1)\ldots(\tau_n j\omega+1)}\right|$ and an extreme case can be examined by approximating this term by a multiple pole with a cut-off frequency of half the target loop resonance frequency. Figure 3-12 shows the resulting shift in the apex frequency when $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|} = \frac{1}{(\tau s+1)^2}$ and $\frac{|P_{ks}(j\omega)|}{|G_s(j\omega)|} = \frac{1}{(\tau s+1)^3}$, where $\tau=4\pi/\omega_{r0}$. A

significant shift in peak frequency is only obtained with a third order pole and in both cases significant attenuation is also obtained.



(a) $\dfrac{|P_{ks}(j\omega)|}{|G_s(j\omega)|} = \dfrac{1}{(\tau s+1)^2}$



(b) $\dfrac{|P_{ks}(j\omega)|}{|G_s(j\omega)|} = \dfrac{1}{(\tau s+1)^3}$

Figure 3-12: The effect of $\dfrac{|P_{ks}(j\omega)|}{|G_s(j\omega)|}$ on the apex frequency of $\left|\dfrac{1}{1+C_k(j\omega)P_{kk}(j\omega)}\right|$

3.3.3 Estimating the Target Loop Disturbance Transfer Function Resonant Frequency

In general the target loop disturbance transfer function resonant frequency, ω_{rok}, is not normally known, although the closed loop resonant frequency, ω_{rk}, and damping factor, ξ_k, might have been estimated previously, for instance during commissioning. If this is the case, then these parameters can be used to estimate ω_{rok}. To do this $\frac{1}{1+C_k(s)P_{kk}(s)}$ and $\frac{C_k(s)P_{kk}(s)}{1+C_k(s)P_{kk}(s)}$ can be approximated by second order models, which have identical natural frequencies, ω_{nk}, but different resonant peak frequencies, ω_{rok} and ω_{rk} respectively: $\omega_{rok} = \frac{\omega_{nk}}{\sqrt{2}}\sqrt{1+\sqrt{1+8\xi_k^2}}$ and $\omega_{rk} = \omega_{nk}\sqrt{\left(1-2\xi_k^2\right)}$, so that these two equations can be combined to estimate ω_{rok}. If this information is not available then it should be possible to estimate ω_{rok} from knowledge of the ultimate frequency, ω_{uk}, since it is likely (but not guaranteed) that $\omega_{rk} < \omega_{rok} < \omega_{uk}$. Hägglund (1995) has pointed out that if a PI/PID controller is properly tuned, then usually the integral time constant of the k^{th} loop, T_{Ik}, approximates to the ultimate oscillation period $\frac{2\pi}{\omega_{uk}}$ of the closed loop system. Thus $\frac{2\pi}{T_{Ik}}$ should be a reasonable estimate of ω_{rok}. Either way the estimation of ω_{rok} must be viewed with caution, and any isolation procedure that is proposed must accommodate this uncertainty.

Finally it is worth pointing out that a ripple is sometimes observed in log-ratio plots. This arises because of the effect of the time delay. Figure 3-13 shows two examples of the same process but with different time delays and hence controller settings, obtained on the basis of Ziegler-Nichols criteria. The ripple is clearly visible in both plots.

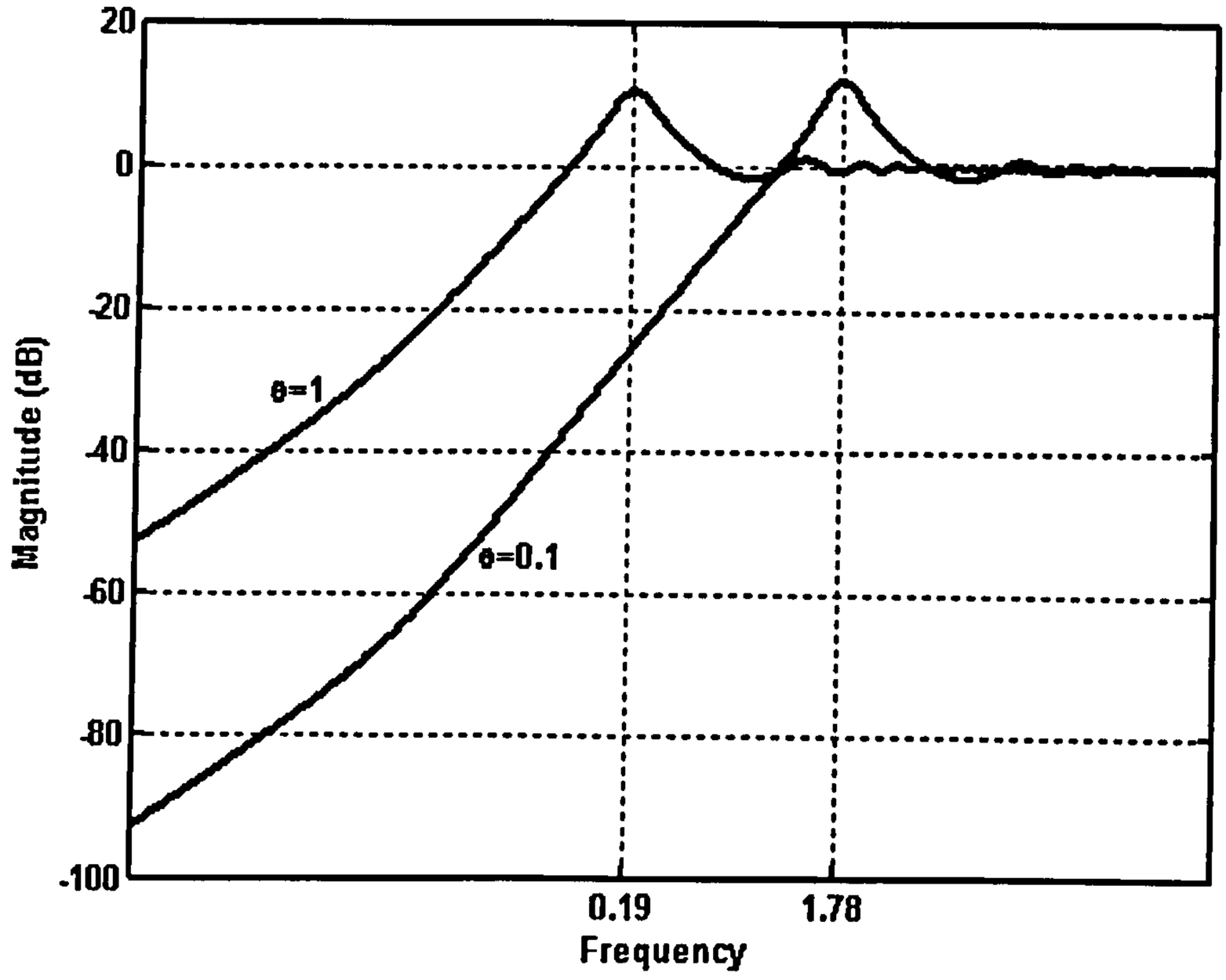


Figure 3-13: Example of 1/(1+CP) with two different time delays

3.3.4 Loop to Indicator Propagation

Suppose that the oscillation in Loop 1 also propagates to an indicator, installed elsewhere on the plant, which outputs signal Y_3. The propagation of the oscillation from U_1 to Y_3 can be described by $Y_3 = P_{31}U_1$ if the propagation path can be represented by P_{31}. The ratio, $r(\omega)$, of the magnitudes of the sinusoidal components of Y_1 to Y_3 at any frequency ω is then given by:

$$r(\omega) \equiv \frac{|Y_3|}{|Y_1|}\Big|_{\omega} = \left|\frac{P_{31}(j\omega)U_1(j\omega)}{P_{11}(j\omega)U_1(j\omega)}\right| = \left|\frac{P_{31}(j\omega)}{P_{11}(j\omega)}\right| \tag{3.10}$$

Ratio $r(\omega)$ is likely to be on a positive slope of ω.vs.$r(\omega)$ if P_{11} and P_{31} are both low pass: $\omega_{11}<\omega<\omega_{31}$ where ω_{11} and ω_{31} are their respective bandwidths; ratio $r(\omega)$ is likely to be on a horizontal slope at low frequencies and on a negative slope otherwise.

3.3.5 Cascade Loop Propagation

If the source is located within a cascade control system, the outer-loop plant dynamics determines how the harmonics propagate from the inner to the outer loop. Referring to Figure 3-14, the ratio $r(\omega)$ of the magnitudes of the sinusoidal components of Y_1 to Y_2 at any frequency ω is simply: $r(\omega) \equiv \frac{|Y_2|}{|Y_1|}\Big|_{\omega} = \left|P_{22}(j\omega)\right|$.

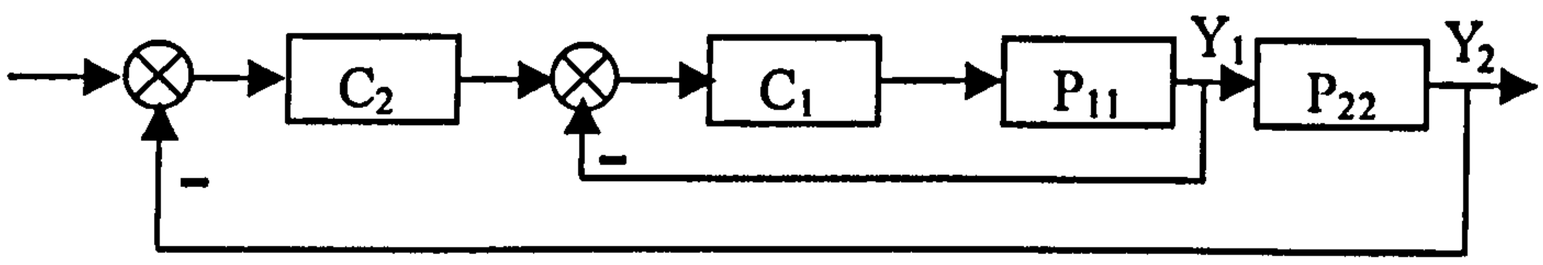


Figure 3-14: Cascade control loop schematic

3.3.6 Conclusions

The change in harmonic content of oscillations as they propagate through a plant is modeled by examining ratios of observed signals. Heuristics are then proposed to enable these models to be characterised without recourse to detailed plant identification. This will result in harmonic-content based procedures in the next two chapters to isolate the source of an oscillation.

Chapter 4

The Bi-spectrum

This chapter starts by introducing the *Bi-spectrum*, its:

- definition and estimation
- application to multiple oscillations
- application to variable-frequency oscillations

Bi-spectral analysis is then proposed as an aid to the isolation of whole-plant oscillations. The aid is centred on a new biamplitude ratio index.

4.1 What It Is

The power spectrum provides information on the second-order properties (i.e. variance, energy) of a signal, whereas the bispectrum provides information on the signal's third-order properties (i.e. skewness). Bispectral analysis is a subset of Higher Order Statistical (HOS) analysis, a rapidly expanding area of signal processing which shows great potential for a wide variety of practical applications (Fackrell *et al.* 1995). There are three main reasons for using HOS analysis: to extract information due to deviations from Gaussianity, to recover the true phase character of a signal, and to detect and quantify nonlinearities in time series (Nikias & Petropulu 1993). Applications of bispectral analysis have been found in fields as diverse as condition monitoring (Collis *et al.* 1998; Zhu *et al.* 2005), system identification (Tugnait & Zhou 2000; Wang *et al.* 2004), and fault diagnosis of rotating machinery (Eugene Parker *et al.* 2000; Li *et al.* 2005).

As reviewed in Chapter 2, Shoukat Choudhury *et al.* (2004) proposed a method to detect system non-linearity based on the *bicoherence*, which is a normalized version of the *bispectrum*. This section gives a detailed description of the properties of the bispectrum, and then analyses the reasons why the bispectrum instead of bicoherence, is more appropriate for oscillation detection and diagnosis.

4.1.1 Definition

Power spectral analysis treats each frequency component independently of all others to provide a measure of the distribution of signal energy across frequencies. This independence is the linear assumption upon which all power spectral methods rely. In practice it is known that many systems are not adequately represented by such linear models, and this prompts the investigation of higher order properties of the signal. The phenomenon of harmonic distortion is an example of a process in which frequency components couple together (Fackrell *et al.* 1995). The bispectrum is a third-order measure which can be interpreted as measuring the amount of coupling at frequencies f_1, f_2 and f_1+f_2. This coupling phenomenon has been observed in real signals, and 'quadratic phase coupling' (QPC) is the term used to describe it.

The bispectrum can be formed by extending the definitions of conventional second-order power spectral measures. The power spectrum of a discrete time series $x(n)$ is described in terms of the signal's discrete Fourier transform (DFT):

$$P(f) = E\left[X(f)X^*(f)\right] \tag{4.1}$$

where $E[]$ denotes the expectation operator, '*' denotes complex conjugation and f is the discrete frequency variable. The bispectrum $B(f_1, f_2)$ at the bifrequency (f_1, f_2) is also defined in terms of the signal DFT:

$$B(f_1, f_2) = E\left[X(f_1)X(f_2)X^*(f_1 + f_2)\right] \tag{4.2}$$

Note that the bispectrum is complex, and has two frequency indices f_1 and f_2. The magnitude of the bispectrum, $|B(f_1, f_2)|$, can be plotted in three dimensions with two orthogonal axes defining the f_1, f_2 plane, and the bispectral content rising from that plane. It is well understood that the DFT, and hence the power spectrum contains redundant information above the Nyquist frequency $f_s/2$, where f_s is the sampling frequency. In a similar way several symmetries exist in the (f_1, f_2) plane, and it is not necessary to compute $B(f_1, f_2)$ for all (f_1, f_2) (Hinich & Wolinsky 1988). The non-redundant region is called the principal domain, is shown in Figure 4-1 and is defined by $\{f_1; f_2 : 0 \le f_1 \le f_s/2, f_1 \le f_2, 2f_1 + f_2 = f_s\}$ (Hinich and Wolinsky 1988). For the analysis of stochastic signals it is useful to divide the principle domain into two triangles called the inner and outer triangles. Statistical tests for stationarity, non-Gaussianity and the detection of aliasing have been devised (Hinich and Wolinsky 1988) based on summing the bispectral content in one or other of these triangles. However, the tests are not useful in the analysis of periodic signals because these signals cannot satisfy the constraints required by the tests.

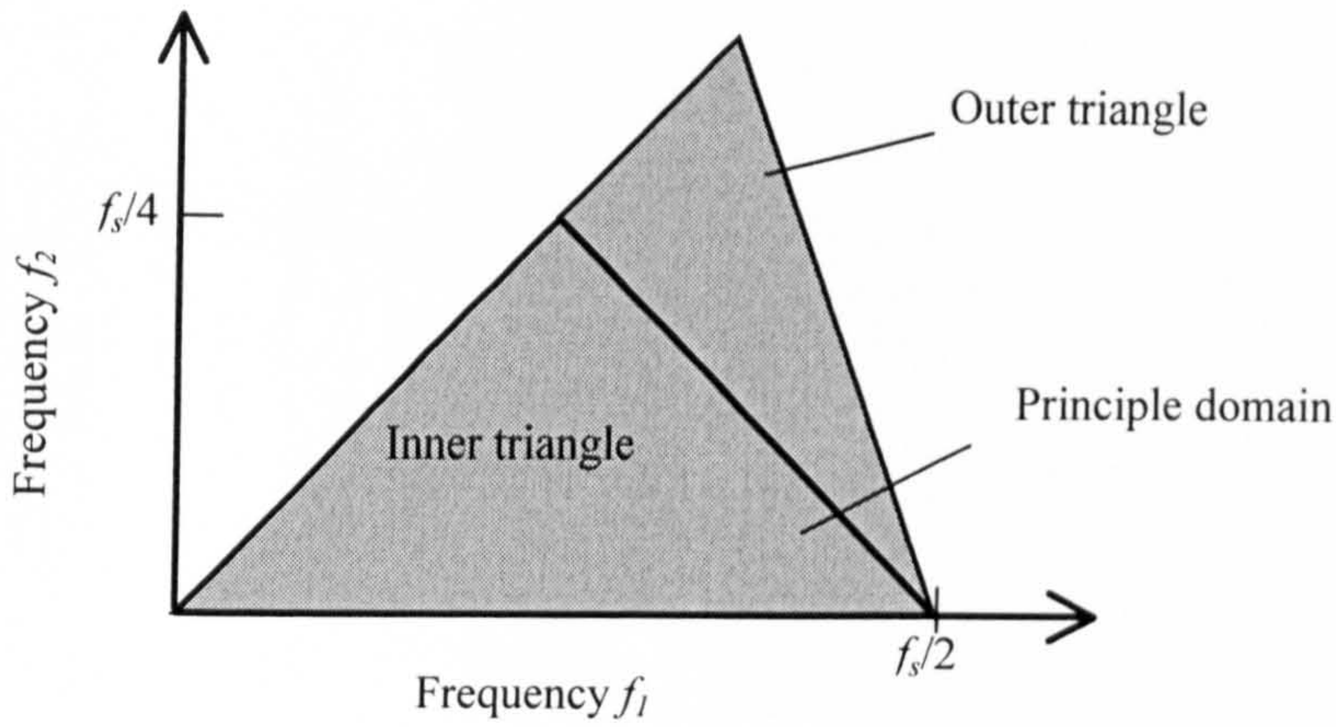


Figure 4-1: The non-redundant part of the bispectral plane, showing the principle domain, composed of the inner triangle and the outer triangle.

The bicoherence is a normalised bispectrum which measures QPC on an absolute scale between 0 and 1. It is defined by

$$bic^2(f_1, f_2) = \frac{|B(f_1, f_2)|^2}{E[|X(f_1)X(f_2)|^2]E[|X(f_1 + f_2)|^2]} \tag{4.3}$$

The bicoherence at any frequency pair f_1, f_2 can be interpreted as the fraction of power at frequency f_1+f_2 which is phase coupled to the component at f_1+f_2.

4.1.2 Estimation

Bispectrum applications have been limited to date because of the difficulty in interpreting bispectral features: the bispectrum is sensitive to many of the computational factors which affect the power spectrum, but in more complicated ways. The bispectrum can be estimated in a similar way to the Welch-periodogram method for power spectrum estimation, but the lengths of data required to obtain

consistent estimates are longer than those required for power spectrum estimation. The variance of the bispectral estimates has the property:

$$\sigma^2(f_1, f_2) = \frac{N_3^{\ 2}}{MK} P(f_1)P(f_2)P(f_1 + f_2) \tag{4.4}$$

where P is the power spectrum and $N_3=M/(2J+1)$ is the number of smoothing windows in each record. It can be seen that the variance of the estimate depends on the second order spectral properties. The bicoherence was proposed as an alternative, that does not have this property. Alternatively the variance of the estimate can be reduced by an averaging procedure and frequency domain smoothing (Nikias and Petropulu 1993). The estimation procedure involves the application of the Fast Fourier Transform, followed by segment averaging:

$$\hat{B}_{segment\ \ averaged}(f_1, f_2) = \frac{1}{K}\sum_{i=0}^{K-1} X_i(f_1)X_i(f_2)X_i^*(f_1 + f_2) \tag{4.5}$$

and frequency-domain smoothing:

$$\hat{B}(f_1, f_2) = \frac{1}{(2J+1)^2}\sum_{k_1=-J}^{J}\sum_{k_2=-J}^{J} \hat{B}_{segment\ \ averaged}(f_1 + k_1, f_2 + k_2) \tag{4.6}$$

where the signal $x(n)$, $0 \le n \le N-1$ is segmented into K records with M samples each, i.e., $N = K \times M$, f_1 and f_2 represent the indices of frequency, i is the index of the segment, X^* is the complex conjugate of X, ^ denotes the estimator and $2J+1$ is the length of the smoothing window that is applied to each record.

In practice it is found that the estimates for periodic signals have lower variance than those for stochastic signals (Fackrell *et al.* 1995), and therefore the choices of the parameters (K, J, etc.) for bispectrum estimation for oscillation isolation (this chapter) and for non-linearity detection (Chapter 6) are different. In this chapter the

magnitude bispectrum is used to isolate the root cause loop of plant-wide oscillations. All candidate measurements are well-defined deterministic signals and stochastic noise is not dominant, so only a few segments and smoothing windows are needed. Unless otherwise stated, a segment length of 1024, a 50% overlap, window length 1 (J=1, i.e. no frequency domain smoothing), and a DFT length of 1024 are used throughout Chapter 4. In Chapter 6 the bispectrum is proposed to discriminate between bad tuning induced oscillations and non-linearity induced oscillations. Stochastic noise is more dominant especially with the example given, so it is necessary to have enough segment averaging and frequency domain smoothing to get good estimates of the stochastic part of the signals. A segment length of 256, a 50% overlap, a Rao-Gabr window of length 21 (J=10), and a DFT length of 256 are used throughout Chapter 6.

4.1.3 Why Bispectrum?

The bicoherence has been applied widely because the estimate of its variance is independent of the signal energy (Fackrell 1996). However, the bispectrum is preferred here because of the considerations listed below.

- The bicoherence is sensitive to division by a small number, as this can artificially increase calculated values and hence be misleading. In practice the peak of the bicoherence is found not to be at the quadratic phased couple bifrequencies. Magnitude bispectra, however, can show explicit peaks at frequencies where phase coupling happens.

- The oscillations of interest are well-defined deterministic signals, and as has been mentioned already, the estimates of deterministic signals have lower variances than those for stochastic signals. Therefore there is less necessity to normalize the bispectrum (i.e. bicoherence) to reduce the estimate variance.
- The calculation of the bispectrum is numerically more efficient than that of the bicoherence. This is very important because deterministic signals don't need segmentation (Nikias and Petropulu 1993) and the estimate of the bicoherence for the entire length of data with few segments is more time and memory consuming due to more matrix calculations. The bispectrum estimator, however, can estimate the bispectrum at desired frequency channels easily with few segments.

4.2 Magnitude Bispectrum of Oscillations

By definition (Equation 4.2), if the components present at frequencies f_1, f_2, (f_1+f_2), and (f_1-f_2) are spontaneously excited modes, each will be characterized by statistically independent random phases. Thus through statistical averaging in Equation 4.2, the bispectrum will take a zero value. On the other hand, if the sum and/or difference component is generated through some nonlinear interaction, then phase coherency exists and the statistical averaging will not lead to a zero bispectrum value. Valve non-linearity induced oscillations contain both even and odd harmonics, that is, quadratic coherency exists in components of the oscillations at frequencies f_0 (fundamental frequency), f_1 (second harmonic, $f_1=2f_0$), f_2 (third harmonic, $f_2=3f_0$), etc. Non-zero peaks are expected to be seen in these frequency channels in the magnitude bispectrum plots.

4.2.1 Oscillations With Single Fundamentals

Non-linear source induced oscillations have a fundamental frequency, which is at the crossing point of the non-linearity describing function and the open loop transfer function of the plant (Chapter 3). The measurement records contain a fundamental with its harmonics. A simple example here is used to demonstrate how the magnitude bispectrum of an oscillating signal shows peaks at the quadratic coupled bifrequencies (fundamentals and harmonics). Consider a signal with a fundamental together with its second and third harmonics:

$$s(i) = \sin(2\pi f_0 i) + \alpha_1 \sin(4\pi f_0 i + \pi/5) + \alpha_2 \sin(6\pi f_0 i + \pi/7), i = 1, \cdots, 4096 \quad (4.7)$$

where f_0 is the fundamental frequency: f_0=0.05, and α_1, α_2 are the amplitudes of harmonics: $\alpha_1 = 0.5, \alpha_2 = 0.1$. The magnitude bispectrum of this signal is given in Figure 4-2, where the contour plot of the whole plane (including the redundant area) is shown on the left panel and the surface plot of only the non-redundant area is shown on the right panel.

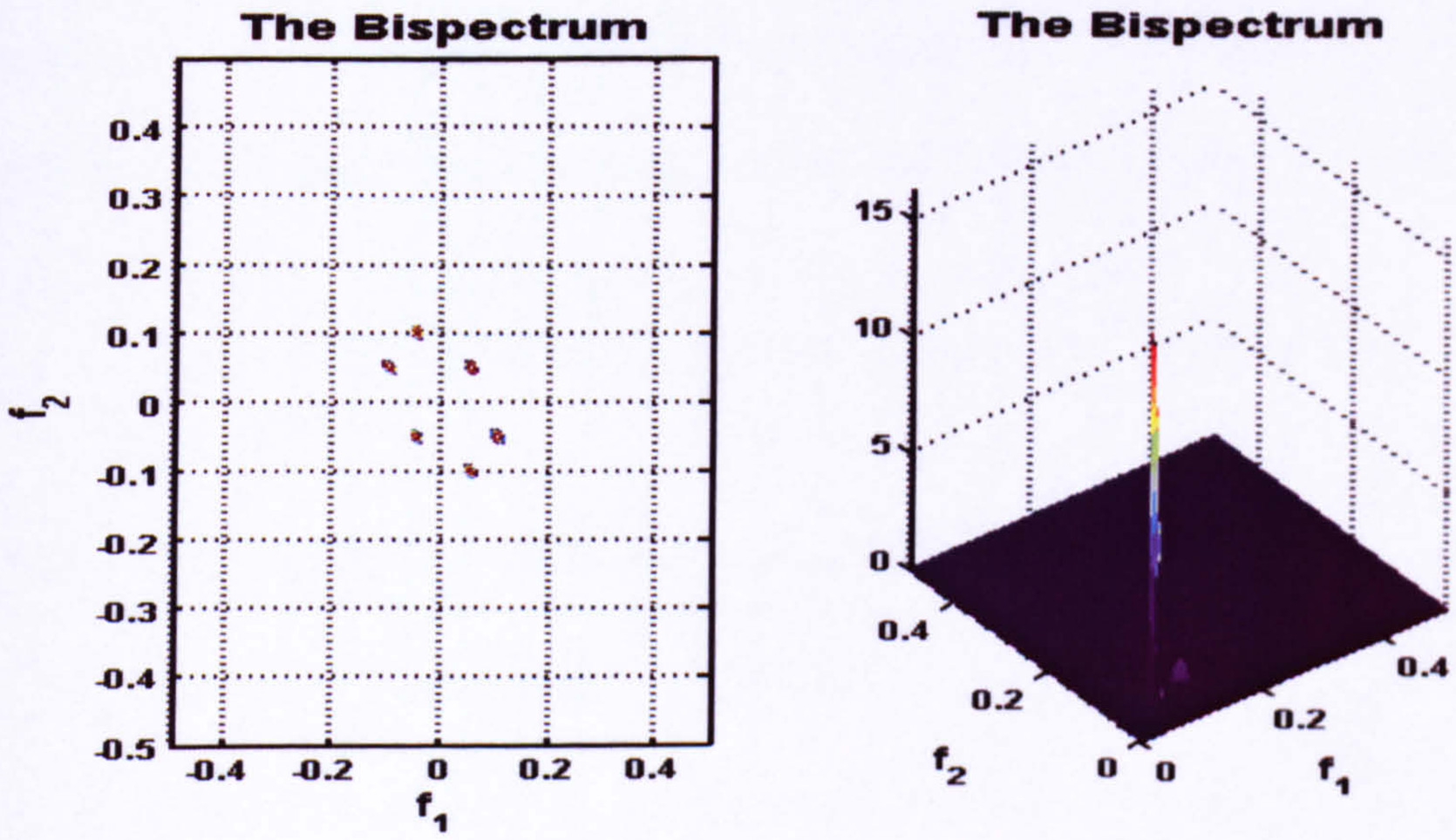


Figure 4-2: Magnitude bispectrum of a signal with single fundamental frequency: left panel—contour plot, right panel—surface plot quadrant.

It can be seen that there is a strong peak at the quadratic coupled bifrequency (f_0, f_0), i.e. (0.05,0.05) in the non-redundant area. With a weak peak at quadratic coupled bifrequency ($2f_0, f_0$), i.e. (0.1, 0.05) it cannot be visible in the contour plot but can be seen in the surface plot. The magnitude bispectra at other frequency channels, however, turn out to be zero, which validates the prediction that no quadratic coupling happens at those bifrequencies.

4.2.2 Multiple Oscillations

A number of non-linear oscillatory sources might be present in a plant, giving rise to multiple oscillations. These oscillations are likely to have different fundamental frequencies so that measurement records might contain two or more fundamentals together with their respective harmonics. The bispectrum has the inherent capability of detecting and diagnosing such multiple oscillations. For instance consider the bispectrum of a signal with two fundamentals and their second and third harmonics:

$$\begin{aligned} s(i) = & \sin(2\pi f_0 i) + \alpha_1 \sin(4\pi f_0 i + \pi/5) + \alpha_2 \sin(6\pi f_0 i + \pi/7) \\ & + \beta_0 \sin(2\pi f_1 i) + \beta_1 \sin(4\pi f_1 i + \pi/4) + \beta_2 \sin(6\pi f_1 i + \pi/8), \\ & i = 1, \cdots, 4096 \end{aligned} \tag{4.8}$$

where f_0 and f_1 are the two separate fundamentals: f_0=0.07, f_1=0.12, $\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$ are the amplitudes of harmonics: $\alpha_1 = 0.5, \alpha_2 = 0.1, \beta_0 = 0.8, \beta_1 = 0.3, \beta_2 = 0.08$. The magnitude bispectrum of the signal is given in Figure 4-3. It can be seen that both oscillations have peaks at bi-frequencies (f_0, f_0), ($2f_0, f_0$) and (f_1, f_1), ($2f_1, f_1$), which makes the separation of these oscillations straightforward. In Section 4.3.2 automatic steps to identify these fundamental frequencies will be given by locating their corresponding bifrequencies. Alternatively Thornhill *et al* (2003a) has proposed an automated frequency domain filtering method to categorize multiple oscillations into multiple single-source oscillations. Their frequency domain filter sets the power in unwanted frequency channels to zero, destroying the power spectrum of the stochastic part of the signal. Hence it is not recommended when dealing with real data corrupted by stochastic noise.

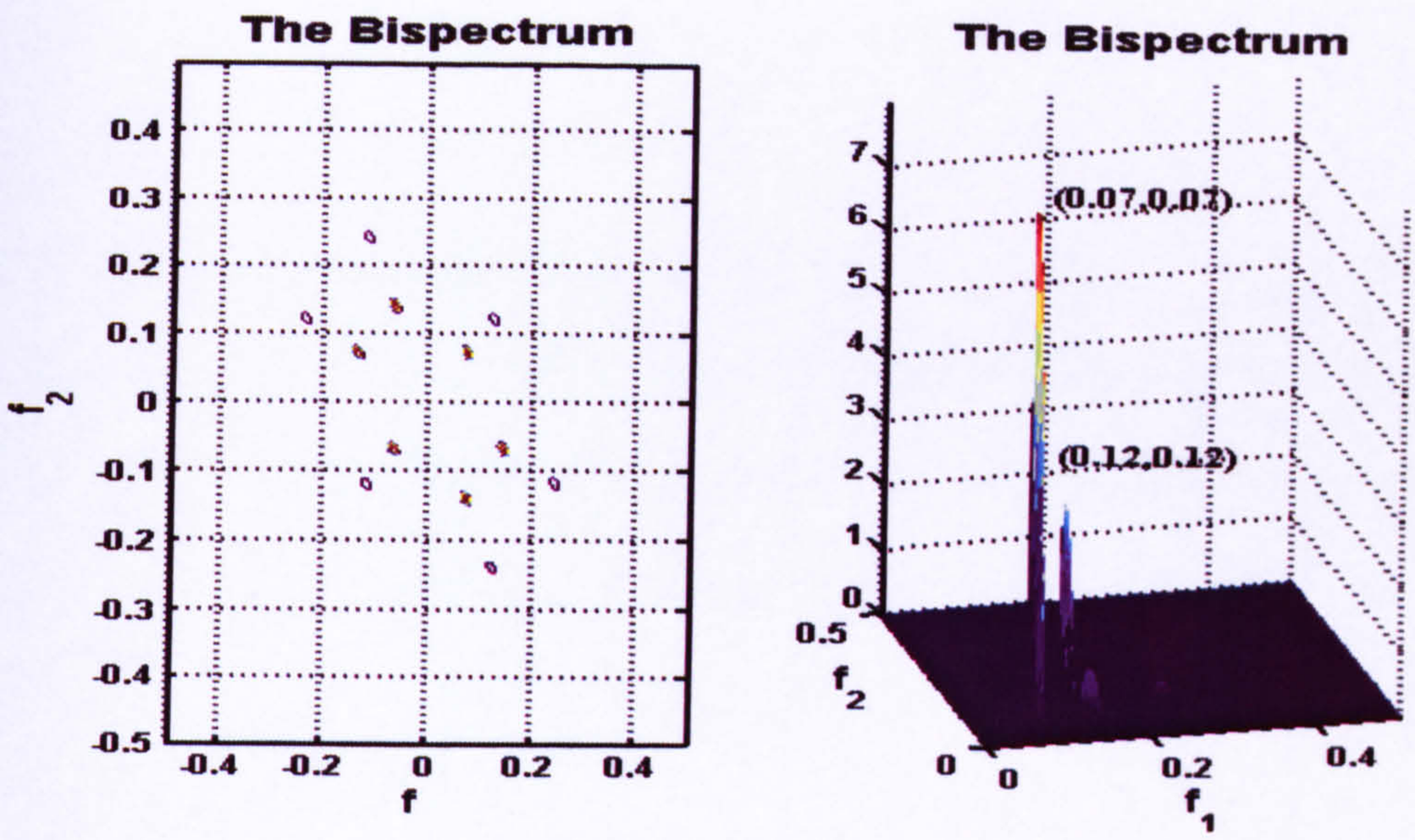


Figure 4-3: Magnitude bispectrum of a signal with two fundamental frequencies: left panel—contour plot, right panel—surface plot.

4.2.3 Variable-Frequency Oscillations

Multiple oscillations have multiple fundamentals whereas variable-frequency oscillations have one fundamental at one time and a different fundamental at another. Since the frequency varies with time, a simple method to deal with such kinds of data is to split the data into stages in which a fixed fundamental frequency can be assumed in each. This cannot be used in the unlikely event that the frequency varies constantly or the variation is so large that its peak encompasses the positions located by both the fundamental and the second harmonic.

An example of this broadening is shown in Figure 4-4 (a) which pertains to a normalized time series and a power spectrum of simulated oscillatory data with frequency variations ($\sin\left(0.01t\beta(t)\right)$, where $\beta(t)=1+0.1\sin\left(0.1t\right)$). It can be seen from the power spectrum plot that the range of the fundamental frequency widens

when the three dominant peaks exist closely. The frequency pertaining to the third peak happens to be twice the frequency pertaining to the first one, and a nonzero bispectrum is expected, which is explained by the magnitude bispectrum plot in Figure 4-4 (b). A similar nonzero bispectrum could happen when a few frequencies are coupled such as $f_1+f_2=f_3$. Special care should be taken when detecting a non-linearity using bispectrum. One way of eliminating the possibility of variable frequency would be to examine the power of the first few harmonic frequencies, as opposed to just the second.

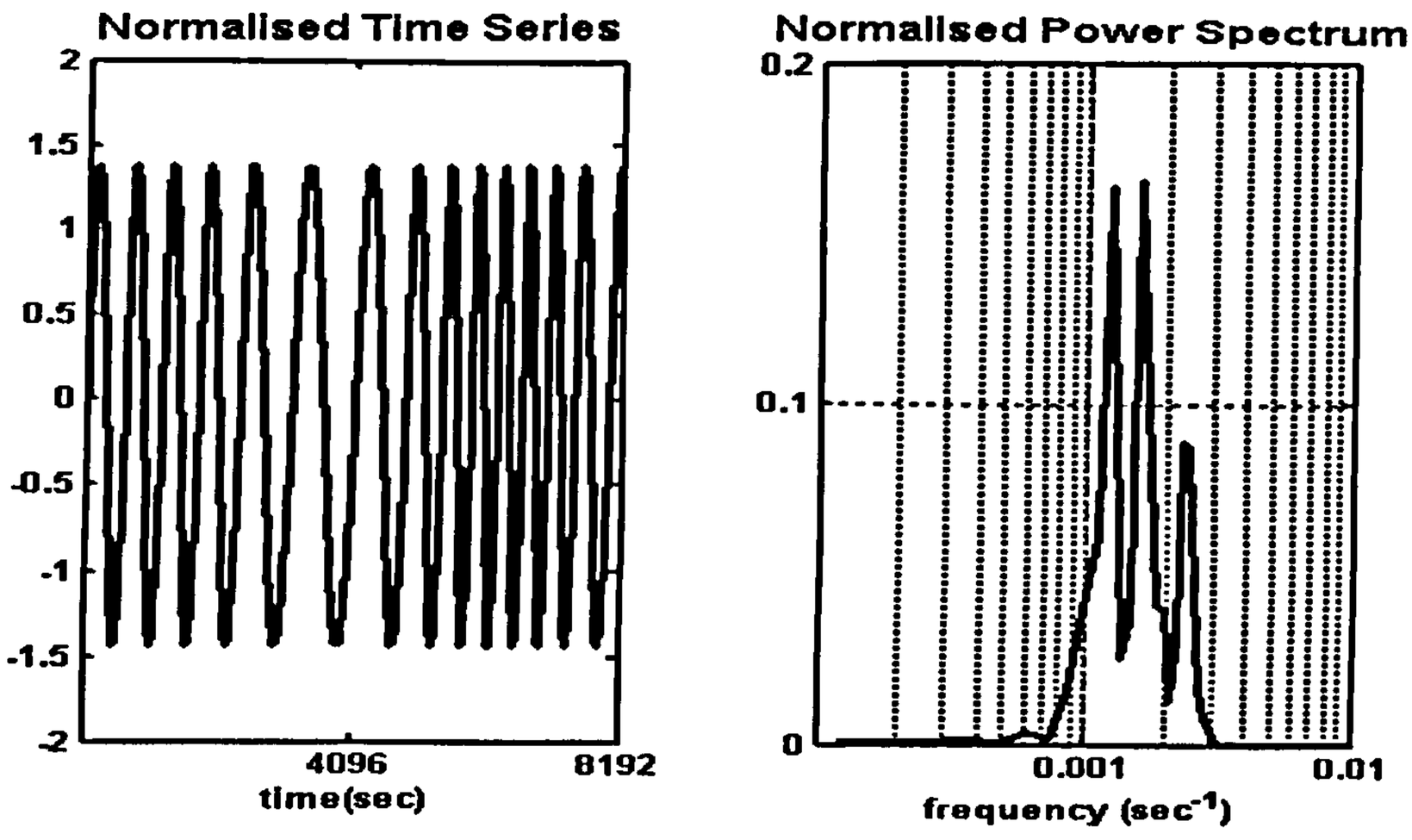


Figure 4-4 (a): Time series and power spectrum of variable frequency, simulated data

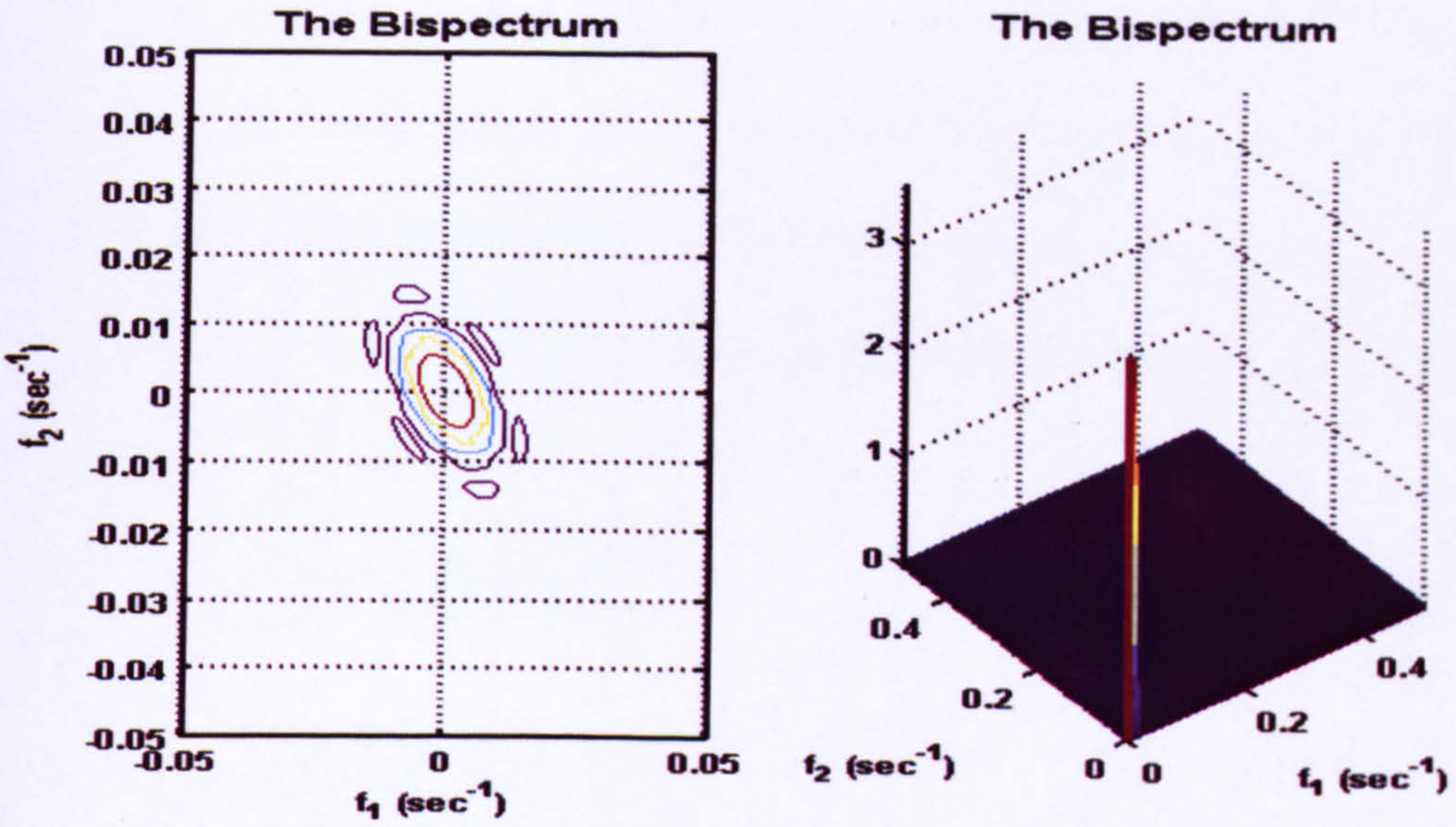


Figure 4-4 (b): The bispectrum of variable frequency, simulated data

Another example pertains to frequency variations observed in real plant data that result in a widening of the fundamental peak in the power spectrum. Figure 4-5 (a) pertains to real plant oscillatory data which was caused by a sticking valve. The power spectrum showed that the frequency varied from 0.0028 to 0.0031 as a result of the oscillation being faster in the second twenty-four hours than in the first twenty-four hours. Possible explanations are that some controller change or operating point change happened in the middle. The bispectrum (Figure 4-5 (b)) shows peaks which are due to quadratic coupling caused by the valve nonlinearity. When the bispectrum is used to isolate the root cause loop of such oscillations, it is recommended that the first half or the last half of the data should be used in order to avoid cross-coupling between these two fundamentals. Figure 4-6 gives the time series, power spectrum and bispectrum of the first half of the data. It can be seen from the power spectrum plot that the fundamental frequency is now limited to only one peak. Although the difference between the bispectra in Figure 4-5 (b) and in Figure 4-6 (b) is not visibly clear because they both have a distinct peak at

bifrequency (f_0, f_0) and much smaller peaks at other bifrequencies, the scale in the later figure is larger, which gives indication of stronger coupling between the fundamental and second harmonics in the first half of the signal. The automatic interpretation of its bispectrum will become simpler and clearer.

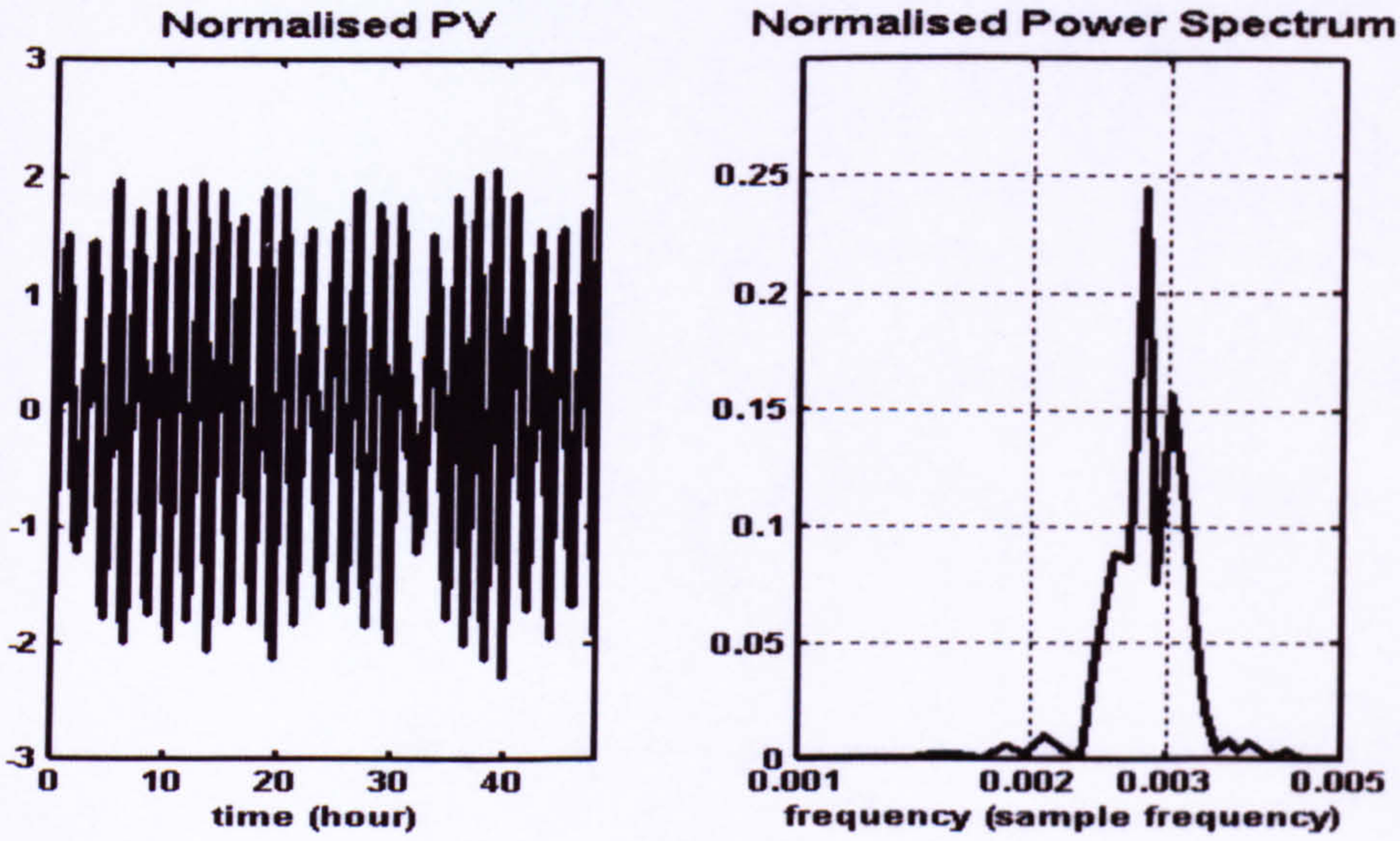


Figure 4-5 (a): Time series and power spectrum of the variable frequency, real data

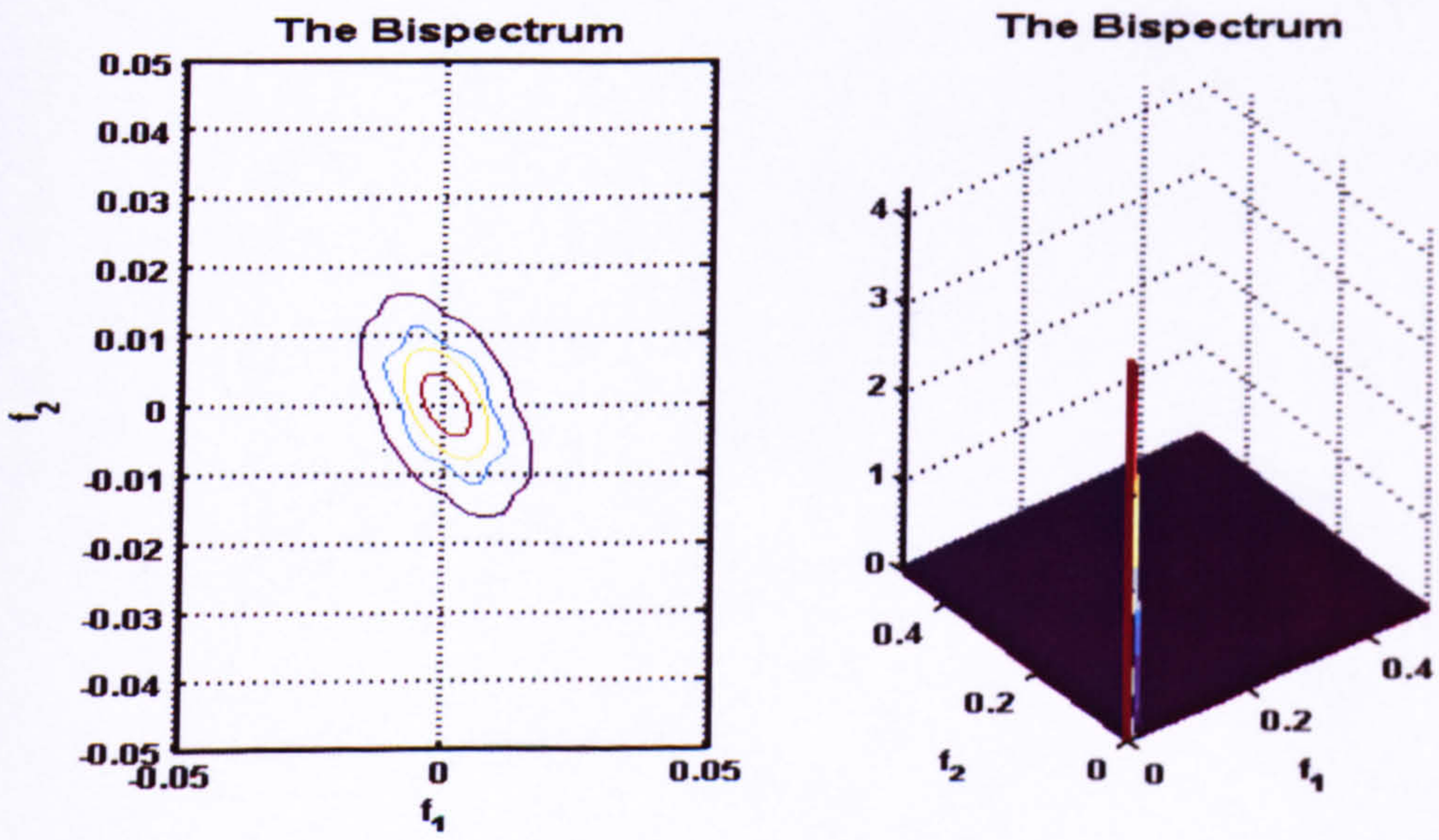


Figure 4-5 (b): The bispectrum of the variable frequency, real data

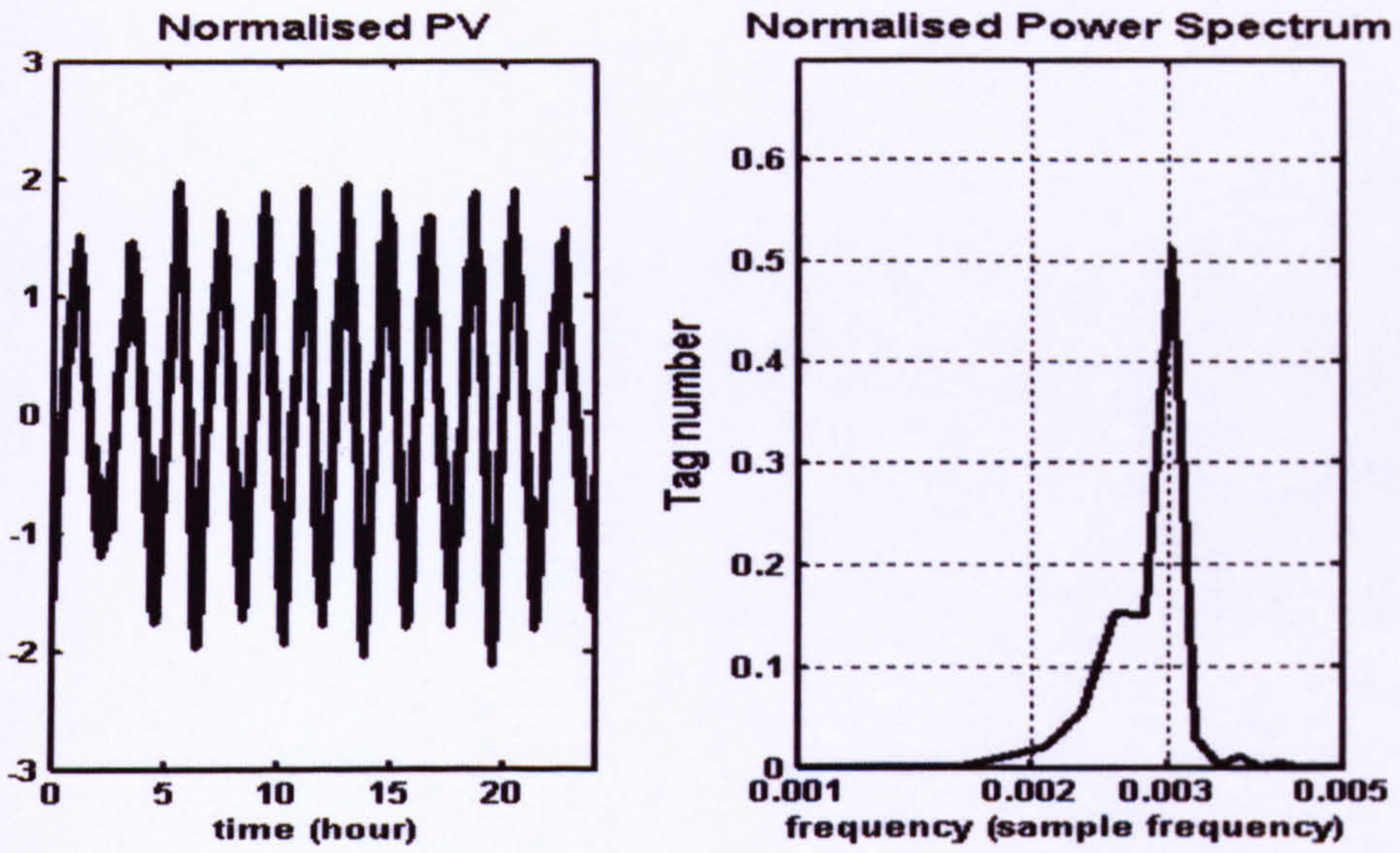


Figure 4-6 (a): Time series and power spectrum of the first half of the variable frequency, real data

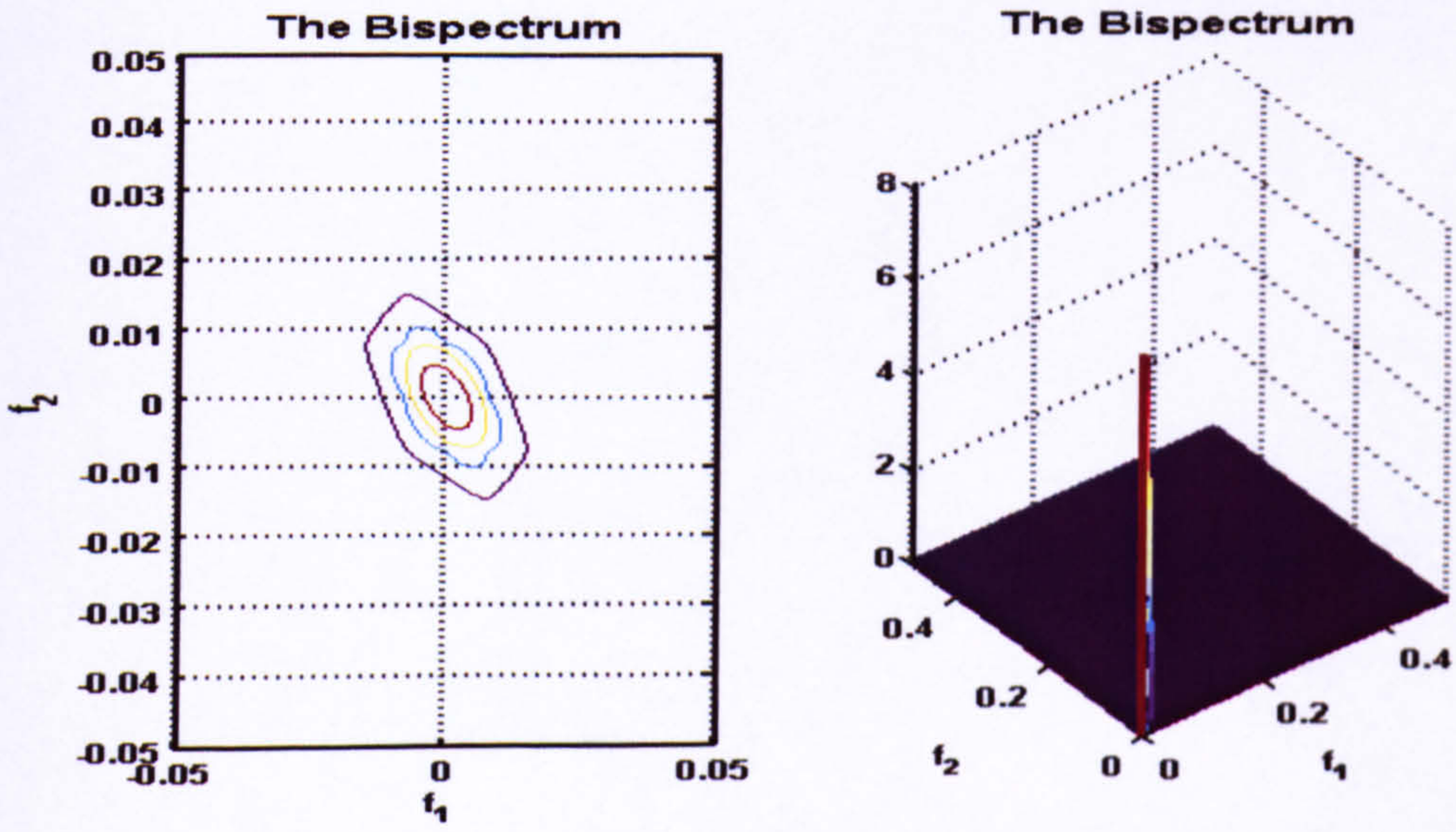


Figure 4-6 (b): The bispectrum of the first half of the variable frequency, real data

4.2.4 Noise Effects

Although the bispectrum of Gaussian noise is theoretically zero, the presence of non-Gaussian noise in the signals can affect the bi-amplitude ratio in at least three ways. Any bispectrum based approach must accommodate these eventualities. These three possible situations are listed below.

Situation 1: The harmonics are hidden by this noise to the extent that they will be indistinguishable in the bispectrum. The net effect might then be that the bi-amplitude at the fundamental bi-frequency is now not the dominant peak in the magnitude bispectrum plot.

Situation 2: Low-frequency noise, e.g. caused by drifting, might also result in the presence of a dominant peak at the corresponding low bi-frequency in the bi-amplitude plot i.e. at a location other than at the fundamental bi-frequency.

Situation 3: Noise might affect the power of the harmonics disproportionately. Generally the Fourier series of an oscillating signal will have a larger fundamental, than second harmonic, than third harmonic and so on. Hence the bi-amplitudes should obey the relationship: $B(f_0, f_0) > B(2f_0, f_0) > B(3f_0, f_0)$ unless corrupted by noise.

4.3 Root Cause Isolation Based On Bi-spectral Analysis

The '*bell*' shaped harmonic propagation described in Chapter 3 implies that the harmonics can either be amplified or attenuated through disturbed loops depending on the frequency of the oscillation relative to the apex of the bell. Based on this knowledge, the source of the oscillations can be found by comparing the harmonic content of detected oscillating loops. In this section a new index, the bi-amplitude ratio index, is presented, which measures the relative harmonic contents among oscillating signals in plant-wide oscillations.

4.3.1 The Bi-Amplitude Ratio

It is important to note that the bi-amplitude ratio is only intended as an index for oscillatory signals. These signals will have normalised measurement records (usually process variables), composed, primarily, of the sum of a fundamental, its harmonics and random noise:

$$x(n) = \sum_{i=0}^{M-1} \alpha_i \cos(2\pi f_i n + \phi_i) + v(n), \quad 1 \le n \le N \tag{4.9}$$

where f_i is the i^{th} normalized frequency (frequency/sampling frequency), α_i is the amplitude of the i^{th} harmonic, ϕ_i its phase, M is the number of significant harmonics, $v(n)$ is Gaussian noise with zero mean and variance σ^2 and is uncorrelated with the signal. Note that all f_i's are distinct and that index i=0 corresponds to the fundamental harmonic, i=1 the second harmonic and so on.

Consider the Fourier transform of a single sinusoidal component, then

$$X(f_i) = F(\alpha_i \cos(2\pi f_i n + \phi_i)) = \frac{N_f \alpha_i}{2}(\cos\phi_i + j\sin\phi_i)\delta(f - f_i) \quad (4.10)$$

where N_f is the length of Discrete Fourier transform and $\delta(\cdot)$ is the *delta* function.The power at each frequency channel f_i is therefore:

$$P(f_i) = E(X(f_i)X^*(f_i)) + w(f_i) = \frac{N_f}{4}\alpha_i^2 + \frac{\sigma^2}{N_f} \quad (4.11)$$

where $w(f_i)$ is the power of Gaussian noise $v(n)$ which is a constant $\frac{\sigma^2}{N_f}$.

The power of the fundamental and harmonic frequency channels are likely to be stronger than that of others. By the definition of the bispectrum, define the *first bi-amplitude*, B_1 as the modulus of the bispectrum at the bi-frequency (f_0, f_0) i.e.

$$\begin{aligned} B_1 &= |B(f_0, f_0)| = |E[X(f_0)X(f_0)X^*(2f_0)]| = |E[X(f_0)X(f_0)X^*(f_1)]| \\ &= |\frac{N_f\alpha_0}{2}(\cos\phi_0 + j\sin\phi_0)\frac{N_f\alpha_0}{2}(\cos\phi_0 + j\sin\phi_0)\frac{N_f\alpha_1}{2}(\cos\phi_1 + j\sin\phi_1)| \quad (4.12) \\ &= \frac{N_f^3}{8}\alpha_0^2\alpha_1 \end{aligned}$$

Note that the bispectrum of Gaussian noise is zero so that the above expression contains only the Fourier transforms of the sinusoids. Define the *second bi-amplitude* B_2 as the modulus of the bispectrum at the bi-frequency (f_0, f_1):

$$B_2 = |B(f_0, f_1)| = |E[X(f_0)X(f_1)X^*(f_2)]| = \frac{N_f^3}{8}\alpha_0\alpha_1\alpha_2 \quad (4.13)$$

It is noted that B_1 is dependent on the magnitudes of the fundamental and second harmonics and that B_2 is dependent on the magnitudes of the fundamental, second and third harmonics. Therefore the first and second bi-amplitudes are associated

with the powers at f_0, f_1, f_2 (see Equation 4.11). The *bi-amplitude ratio* is defined as the ratio of the first and the second bi-amplitude:

$$r = \frac{B_1}{B_2}. \tag{4.14}$$

From (4.12) and (4.13), $r = \frac{\alpha_0}{\alpha_2}$, which means that r is the ratio of power of the fundamental at frequency f_0 to that of the third harmonic at frequency f_2. In practice, bi-amplitude ratio r is likely to be greater than one because the fundamental should dominate. The larger the bi-amplitude ratio, the stronger the fundamental power compared to the third harmonic power.

4.3.2 The Calculation of The Bi-amplitude Ratio

The bi-amplitude ratio can be calculated automatically. To do this the first and second bi-amplitudes need to be estimated so that the bi-amplitude ratios can be calculated. The results in the procedure are given below.

(1) Test the raw data for an oscillation. Eliminate data whose variation is small relative to the quantisation imposed by the data acquisition system, because quantisation introduces harmonics into the records and makes the analysis more susceptible to noise. Then normalize the time records to zero mean and unit standard deviation.

(2) Calculate the power spectrum $P(f)$ of each of the signals and obtain the frequency that locates the maximum power (fundamental frequency) in each of their

spectra. Form a subset of signals that have a common frequency index (f_0) because they pertain to a common source of oscillations.

(3) For each signal in this subset, calculate its bispectrum. Bispectra can be estimated using the direct approach and segment averaging (Section 4.1.2). The magnitude bispectrum is an $N_f \times N_f$ matrix with two frequency indices f_1 and f_2, where N_f is the discrete Fourier Transform length of the bispectrum estimates.

(4) For each magnitude bispectrum matrix search the principle domain to obtain the maximum bi-amplitude and record its bi-frequency (f_1, f_2). If $f_1 = f_2 = f_0$, i.e. the coupling between the fundamental and the second harmonics is the strongest, then set this maximum bi-amplitude as the first bi-amplitude $B_1(f_0, f_0)$. Otherwise, based on the assumption that the closer to the source loop, the stronger the signal-to-noise ratio and hence a stronger coupling between the fundamental and second harmonics, the signal is likely to be corrupted by significant quantities of non-Gaussian noise or relates to an oscillation with a different fundamental frequency, in which case eliminate the signal from the subset because it is not possible to be close to the source loop.

(5) For each of the remaining signals obtain the second bi-amplitude $B_2(2f_0, f_0)$ from the magnitude bispectrum matrix through the bi-frequency index $(2f_0, f_0)$.

(6) Calculate the bispectrum ratio: $r = \frac{B_1}{B_2}$

(7) Eliminate those ratios that conform to one of the special situations listed in Section 4.2.4.

4.3.3 Inter-loop Variation in The Bi-amplitude Ratios

To examine inter-loop variation, first suppose that the source of the oscillation is in loop *k-1*:

$$x_{k-1}(n) = \sum_{i=0}^{M_{k-1}-1} \alpha_i^{k-1} \cos(2\pi f_i n + \phi_i) + v_{k-1}(n) \tag{4.15}$$

where f_i is the i^{th} discrete frequency, α_i^{k-1} is the amplitude of the i^{th} harmonic, ϕ_i its phase, M_{k-1} is the number of significant harmonics, $v_{k-1}(n)$ is Gaussian noise with zero mean and variance σ_{k-1}^2 and is uncorrelated with the signal. The corresponding bi-amplitude ratio is:

$$r_{k-1} = \frac{B_1}{B_2} = \frac{\alpha_0^{k-1}}{\alpha_2^{k-1}} \tag{4.16}$$

Then if the k^{th} loop interacts with the $k\text{-}1^{th}$ loop:

$$x_k(n) = \sum_{i=0}^{M_{k-1}-1} \alpha_i^{k-1} |G_k(2\pi f_i)| \cos(2\pi f_i n + \phi_i) + v_k(n) \tag{4.17}$$

where $G_k(2\pi f_i)$ is the transfer function from the $k\text{-}1^{th}$ loop to the k^{th} loop, and normally has a 'bell'-shaped frequency response plot. The corresponding bi-amplitude ratio is then:

$$r_k = \frac{B_1}{B_2} = \frac{\alpha_0^{k-1} |G_k(2\pi f_0)|}{\alpha_2^{k-1} |G_k(2\pi f_2)|} \tag{4.18}$$

According to the frequency response of G_k, $|G_k(2\pi f_0)| \lhd G_k(2\pi f_2)|$ if the fundamental and lower harmonics are located on the positive slope (low frequency case) and the harmonics are amplified, $r_k < r_{k-1}$, otherwise $r_k > r_{k-1}$.

This information forms the basis of a search algorithm that considers each loop in turn. It then hypothesises that the source of the oscillations is located in that particular loop and assesses whether the harmonic distortion observed at each target loop and indicator agrees with this hypothesis by testing for the conditions discussed in the previous section. To do this a number of the harmonics observed at each target/indicator must be compared with those observed at the source so that the slope of the log-ratio plot can be estimated in the vicinity of these harmonics. The source loop is then deemed to be that loop that conforms to these conditions the closest.

Let the bi-amplitude ratio of the source loop be r_s and the bi-amplitude in the target loop r_t. Then if the first and third harmonics in the source loop are given by $\alpha_1 \sin\omega t$ and $\alpha_3 \sin 3\omega t$, $r_s = \frac{\alpha_0}{\alpha_2}$ and $r_t = \frac{\alpha_0 r(\omega)}{\alpha_2 r(3\omega)}$. Suppose that the first and third harmonics lie on the same constant slope of the $\log_{10}\omega$.vs.$20\log_{10}$ r(ω) graph and that this slope is of gradient of g dBs/decade. Then

$$\begin{aligned} 20\log_{10}\left(\frac{r_s}{r_t}\right) &= 20\log_{10}\left(\frac{r(\omega_3)}{r(\omega_1)}\right) \\ &= 20\log_{10} r(\omega_1) + g[\log_{10} 3\omega_1 - \log_{10}\omega_1] - 20\log_{10} r(\omega_1) \quad (4.19) \\ &= g\log_{10} 3 \end{aligned}$$

A positive g implies amplified harmonics whilst a negative g implies attenuated harmonics.

The first stage of the search strategy is therefore to produce a table of bi-amplitude ratios scaled by $\log_{10}3$ and with each row pertaining to a different source loop hypothesis. Each row would then be tested against the conditions described in the

previous section. These tests are encapsulated in the flow diagram shown in Figure 4-7 and explained below, for that situation where 'bell' shapes are expected, i.e. controllers are of type (d) (see Section 3.4.1) and target loop disturbance transfer function resonant frequencies are estimated from controller integral time constants(see Section 3.4.3). Tolerances ε_1 and ε_2 are non-zero because of undulations in plateaus. They are typically set at about 2 in the current studies.

1. If the third harmonic frequency of the plant-wide oscillations is lower than all the apexes of the 'bells' (i.e. $< 2\pi/T_{Ik}$), harmonics would be amplified in all target loops and loops associated with rows in which all the *g*'s are positive can be marked as candidate source loops.

2. If the fundamental harmonic frequency of the plant-wide oscillations is higher than all the apexes of the 'Bell' (i.e. $> 2\pi/T_{Ik}$), harmonics would be attenuated in all target loops and loops associated with rows in which all the *g*'s are negative can be marked as candidate source loops.

3. If the oscillation period happens to be higher than the integral time constants in some loops and lower than those in the other loops, the situation is more complicated. Unless the controllers have been detuned or over-tuned, the procedure involves marking those loops that are associated with rows, which have positive *g*'s in target loops whose integral time constants are lower and negative g's in those loops whose integral time constants are higher.

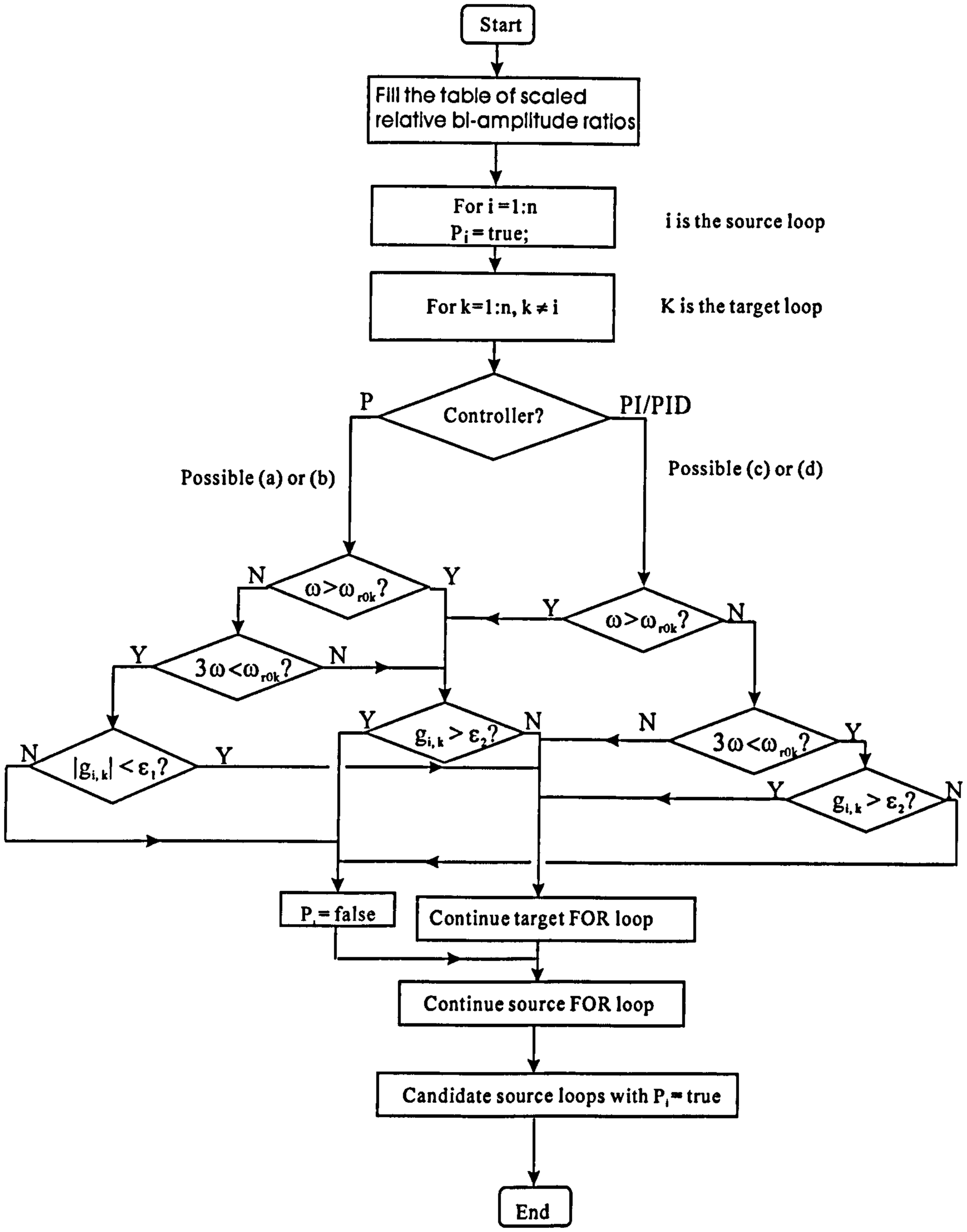


Figure 4-7: Flow chart of the isolation process

4.3.4 Fifth Harmonic Based Verification

Although the biamplitude ratio only evaluates the power ratio of the fundamental to the third harmonic, there are some other measures that can be derived from the biamplitude spectrum which may help to verify the isolated root cause candidates. For example the bi-amplitude ratios of the source loop and a target loop are very close, or the ratio of the source loop is even larger than that of a target loop because the third harmonic in a target loop is corrupted by non-Gaussian noise. The power ratio of the fundamental to the fifth harmonic could provide one such measure. This 'fifth' bi-amplitude ratio can be obtained by simply calculating another biampitude $B_3(2f_0,3f_0)$, where

$$B_3 = |B(f_1, f_2)| = |E[X(f_1)X(f_2)X^*(f_4)]| = \frac{N_f^3}{8}\alpha_1\alpha_2\alpha_4 \tag{4.20}$$

The 'fifth' bi-amplitude ratio is defined by

$$r^5 = \frac{B_2}{B_3} \tag{4.21}$$

In the case of low pass attenuation, the 'fifth' bi-amplitude ratio of signals recorded in the source loop should be the minimum amongst all those in all the loops.

4.3.5 Examples

Three examples are examined here to demonstrate the method proposed. One is a simulated plant with two interacting loops, in which one loop has a sticking valve. The other two are from industrial data. The simulation has the fundamental and lower harmonics located on the positive slope of the disturbance transfer function

frequency plot so that the source of the oscillation has the larger bi-amplitude ratio. The second pertains to an Eastman Chemical Company facility that oscillated at a relatively low fundamental frequency and the third pertains to a facility that had a relatively high frequency plant-wide oscillation. The same three examples are also used in Chapter 5 and Chapter 7 to assess other methods. Two more simulations pertaining to 4-loop plant simulations are also given, one with weak coupling, the other with tight coupling. The first is straightforward, whilst the second demonstrates the limitations of the approach.

The Simulated Plant

Time series were generated by simulating a simple distillation column model described by Seborg *et al.* (1989), which was adapted by incorporating a valve model into one of its loops. Details pertaining to the revised model are given in Figure 4-8; stiction in the valve (in Loop 1) was increased until the model oscillated. The static friction model that was used is that due to Horch & Isaksson (1998). Both loops oscillated. Note that the loops were excited by applying a set point change at time zero, whilst at the same time introducing a Gaussian distributed noise disturbance with zero mean and standard deviation $\alpha\sigma$, where σ is the standard deviation of the pure signal, and α is the noise level (α=10%). Their controlled process variables (*PVs*) were recorded, resulting in two 4096-sample time series. The normalised records and associated power spectra are shown in Figure 4-9, where the frequency axis in the right panel is normalised by the sampling frequency (10 min^{-1}). It can be seen that there is a dominant peak at f=0.025 frequency/sample frequency ($0.25 min^{-1}$ or $4 min$ period), the fundamental frequency, in each of the power spectrum plots.

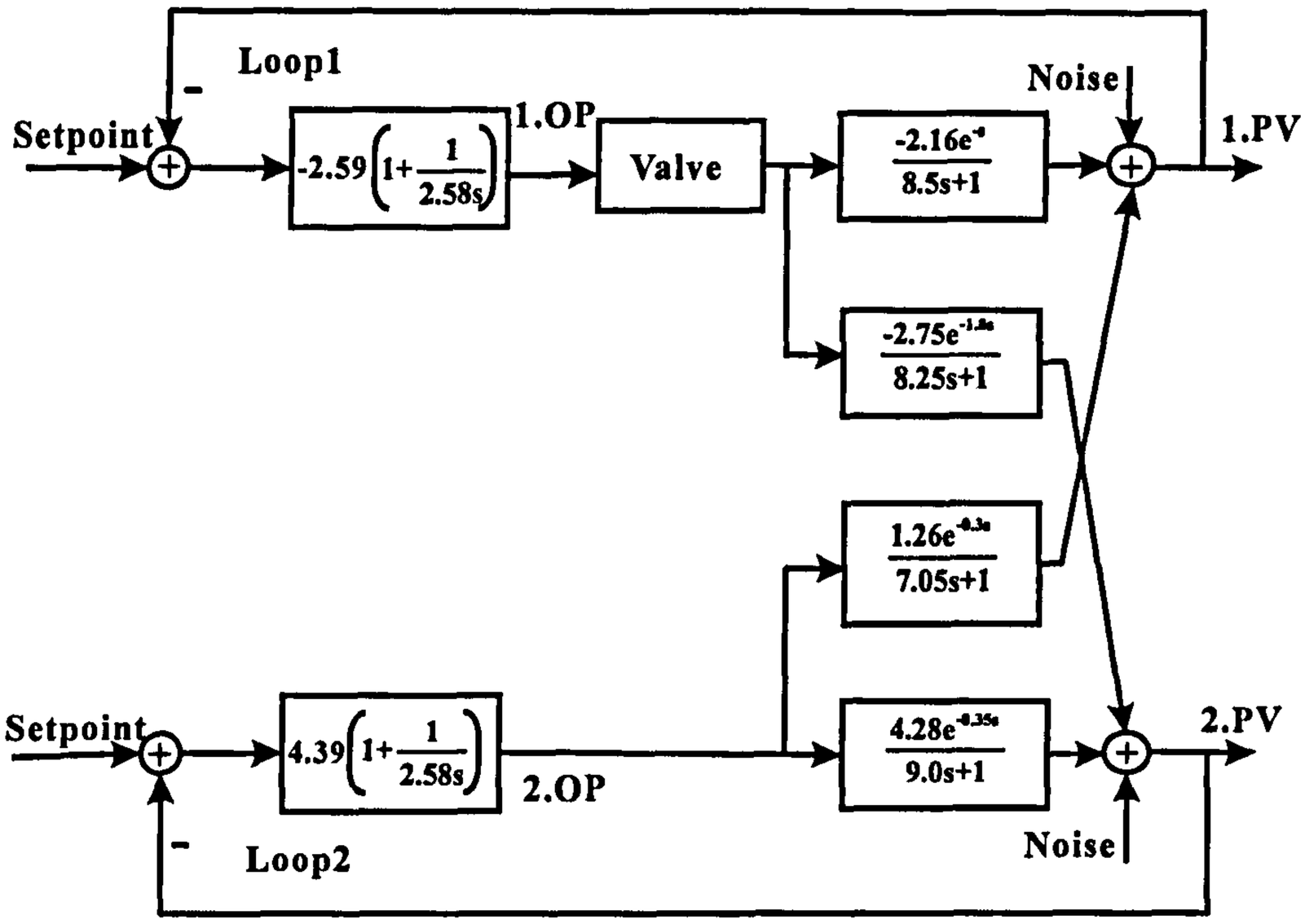


Figure 4-8: The simulated plant

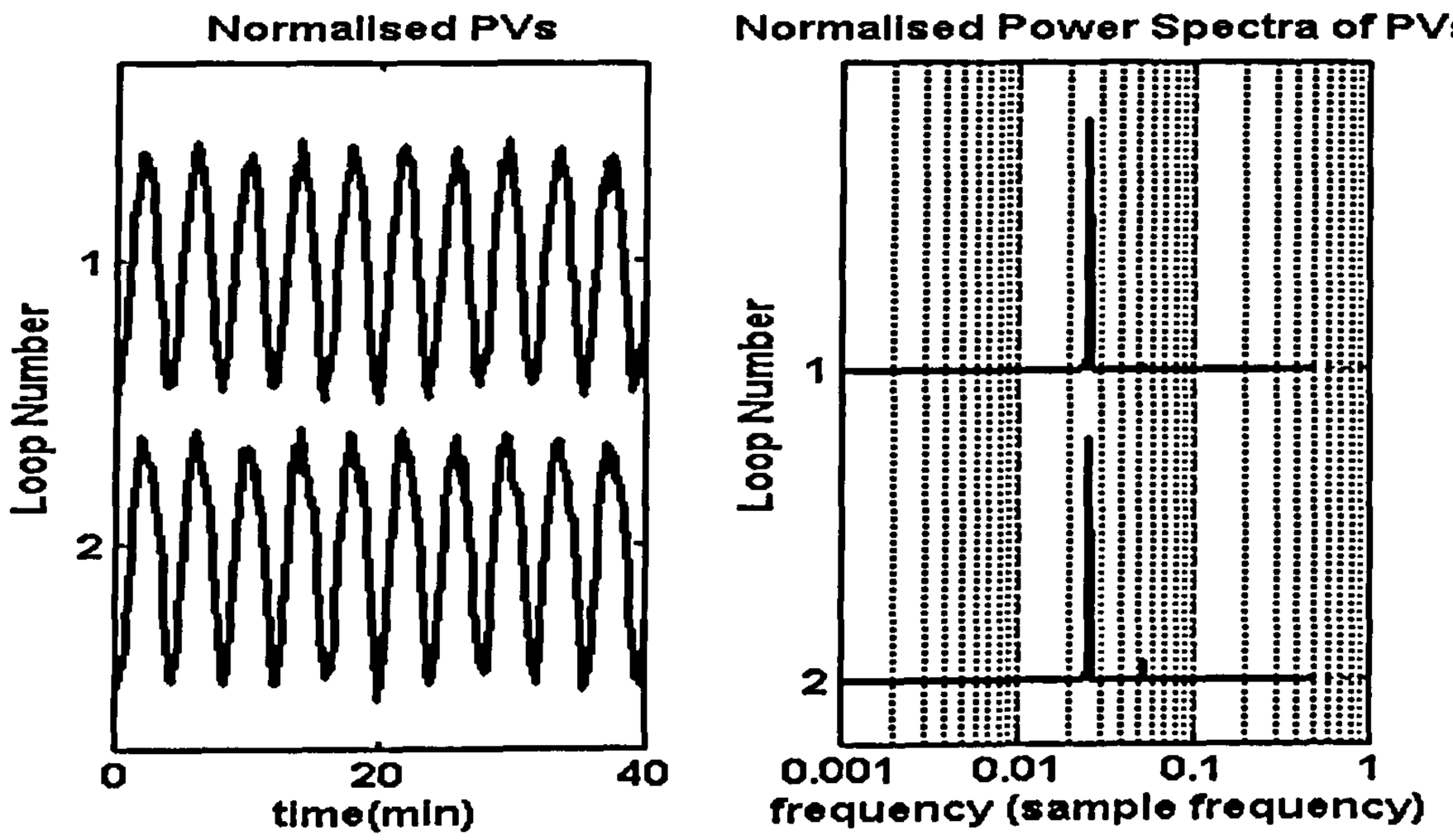


Figure 4-9: Process variable time series and power spectra for the simulation

Figure 4-10 shows the magnitude Bode diagram of transfer function $G = \frac{P_{21}}{P_{11} + C_2\left(P_{11}P_{22} - P_{21}P_{12}\right)}$, as derived on the basis of the transfer functions given in Figure 4-8. The cut-off frequency is at 4.41 rad/min (0.7 min^{-1}). Therefore the

fundamental frequency (0.25 min^{-1}) is on the positive slope so that the lower order harmonics will be amplified relative to the fundamental. Note that the diagram is flat at higher frequency channels because the simulation is probably unrealistic at high frequencies.

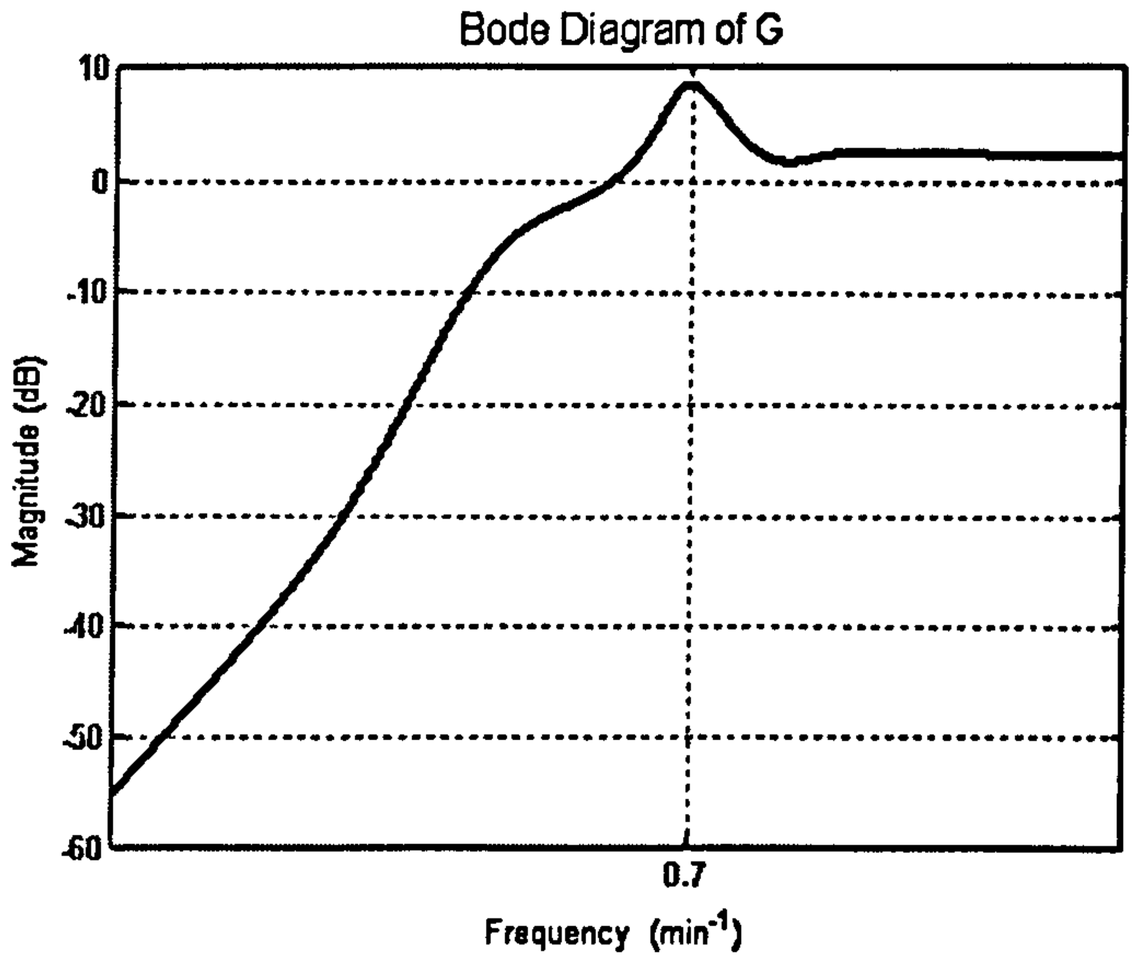


Figure 4-10: Magnitude plot of *G* in the simulated case

A larger bi-amplitude ratio is expected in Loop1 because of the harmonic amplification. Table 4-1 gives the bi-amplitude ratios that were calculated for both loops and Figure 4-11 shows the non-redundant bi-amplitude plots of both loops. Note that the maximum bi-amplitudes in both are around the bi-frequency (0.025,0.025).

Loop Number	B_1	B_2	**Ratio (*r*)**
Loop1	3.66	0.03	108.64
Loop2	8.12	0.23	35.96

Table 4-1: The biamplitude ratios for both loops

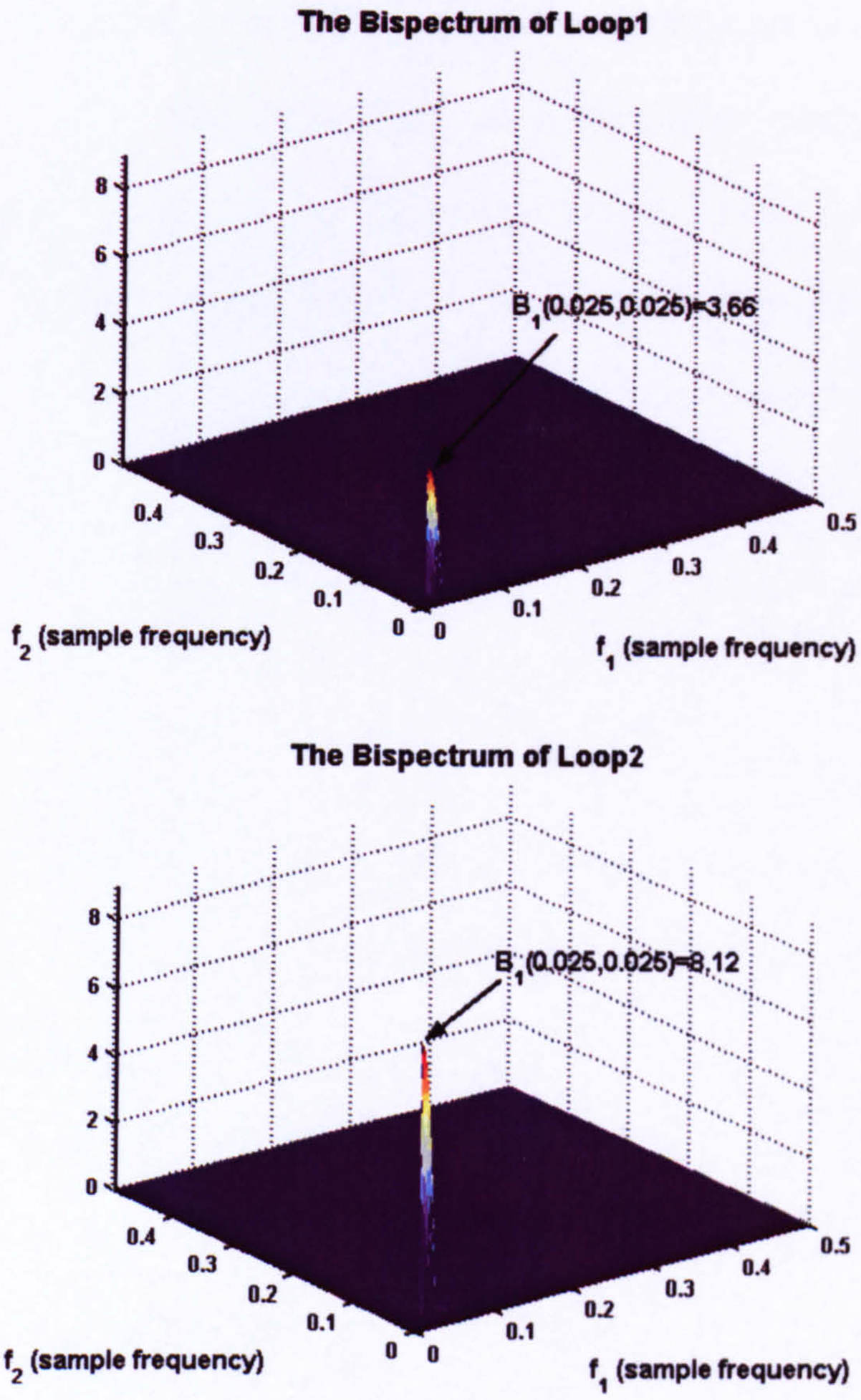


Figure 4-11: The bi-amplitude plots of both loops

Case study 1 (Low-Frequency Case)

Other researchers (Thornhill *et al.* 2002b; Thornhill *et al.* 2003a; Thornhill *et al.* 2003b; Xia & Howell 2003a) have analysed this process data. The operators have reported that a sticking valve was found in the loop with Tag22, and that this was associated with a plant-wide oscillation in the frequency range [0.001 0.01] (100-1000 samples/cycle).

Figure 4-12 shows the normalised measurements and their power spectra, for the 30 filtered *pv's*. It is clear that there is a dominant, plant-wide oscillation corresponding to a fundamental frequency at about 0.0029 on the normalised frequency axis. Note that Tags 1, 2, 9, 14, 16, 17, 21 and 30 can be eliminated from the group because their power spectra show peaks at frequencies other than at 0.0029; these loops are likely to be contaminated by excessive noise and cannot be considered here as root cause candidates. Tags 3, 4, 6, 11, 12, 15, 18, 20, 24, 26 and 27 relate to indicators.

Table 4-2 lists the scaled relative bi-amplitude ratios of the control loops selected as root cause candidates and their corresponding controller integral times. The sample frequency is 20 seconds, so the fundamental frequency is at about a 114-minute period. Almost all of the integral time constants are less than 114/3, which suggests that both the fundamental and third harmonic frequency are likely to be lower than their resonant frequencies and so amplified harmonics (positive *g*'s) are expected in the target loops. Only Tag 13 has an integral time larger than 114/3, which suggests that the third harmonic frequency may be larger than the resonant frequency of Tag13; hence it is possible that the third harmonic might be amplified (positive *g*), attenuated (negative *g*) or even remain the same (zero *g*) relative to the fundamental. The diagnostic aim can therefore be interpreted as finding a source loop in which all the target g's are positive with the exception of that pertaining Tag13. The loop with Tag22 conforms to this requirement, because the row with Tag22 at its head is all positive; a very small *g* (3.8) in the Tag13 suggests that the third harmonic is not on the 20db/decade slope but on the slope after the apex of the 'bell'. The table of scaled relative bi-amplitude ratios for loop to indicators (Table

4-3) provides supporting evidence for this decision of Tag22 as the source loop, because positive g's are to be expected in most plants.

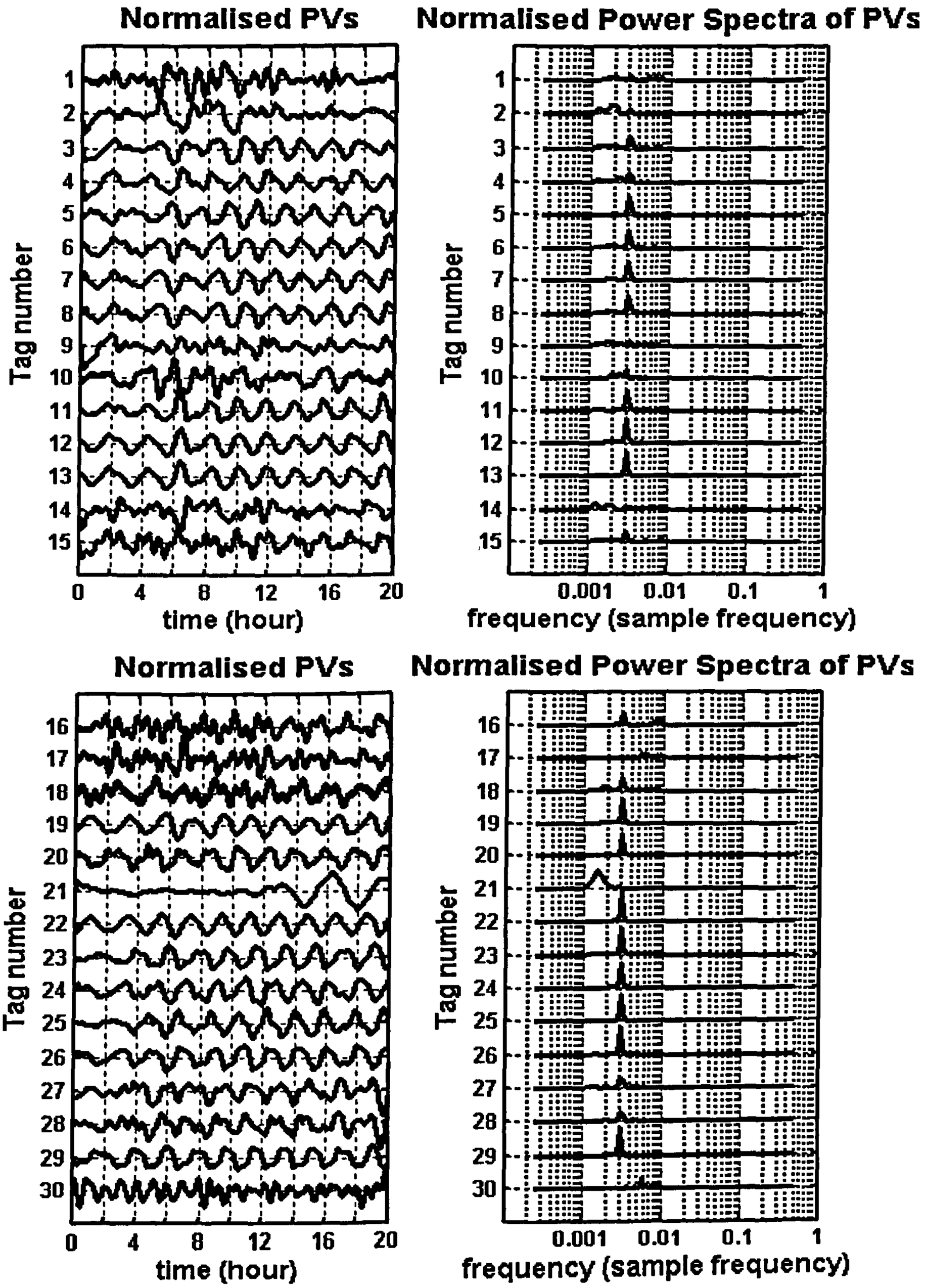


Figure 4-12: Normalized time trends of pvs and the power spectra of 30 plant instruments

Source Loop \ Target Loop		5	7	8	10	13	19	22	23	25
	T_i	20	0.3	0.3	0.5	50	1.5	15.0	0.3	20
	5	-	8.5	1.1	18.8	-19.2	-20.2	-23.0	-14.2	-10.7
	7	-8.5	-	-7.4	10.3	-27.7	-28.7	-31.5	-22.7	-19.2
	8	-1.1	7.4	-	17.8	-20.3	-21.3	-24.1	-15.3	-11.8
	10	-18.8	-10.3	-17.8	-	-38.1	-39.0	-41.8	-33.0	-29.5
	13	19.2	27.7	20.3	38.1	-	-1.0	-3.8	5.0	8.5
	19	20.2	28.7	21.3	39.0	1.0	-	-2.8	6.0	9.5
	22	23.0	31.5	24.1	41.8	3.8	2.8	-	8.8	12.3
	23	14.2	22.7	15.3	33.0	-5.0	-6.0	-8.8	-	3.5
	25	10.7	19.2	11.8	29.5	-8.5	-9.5	-12.3	-3.5	-

Table 4-2: Scaled relative bi-amplitude ratios (loop-to-loop)

Source Loop \ Indicator		3	4	6	11	12	15	18	20	24	26	27
	5	-	-	-5.9	-8.9	-	18.2	17.5	-5.3	14.6	-5.9	-0.4
	7	-	-	-	-	-	12.1	11.4	-	8.5	-	-6.5
	8	-	-	-6.0	-9.0	-	18.1	17.3	-5.5	14.4	-6.1	-0.6
	10	-	-	-	-	-	-0.4	-1.2	-	-4.1	-	-
	13	5.3	5.1	13.8	10.7	2.2	37.9	37.1	14.3	34.2	13.7	19.2
	19	5.4	5.2	13.9	10.9	2.4	38.0	37.3	14.4	34.3	13.8	19.3
	22	5.6	5.4	14.0	11.0	2.5	38.1	37.4	14.6	34.5	14.0	19.5
	23	-1.3	-1.5	7.2	4.2	-4.3	31.3	30.6	7.7	27.7	7.2	12.7
	25	-3.7	-3.9	4.7	1.7	-6.8	28.8	28.1	5.3	25.2	4.7	10.2

Table 4-3: Scaled relative bi-amplitude ratios (loop-to-indicator)

Case study 2 (High-Frequency Case)

This set of refinery data (courtesy of a SE Asian refinery) has been examined previously by others (Thornhill *et al.* 2001) who have recommended that loops with tags 33 and 34 should be considered as prime locations of the root cause. Twelve loops have been found to be involved in a plant-wide oscillation, so the 512-sample process variables pertaining to these loops were analysed. Figure 4-13 plots their time trends and spectra: all show a strong spectral peak at 0.06 min^{-1} corresponding to oscillations with a period of 16.7 *min.*

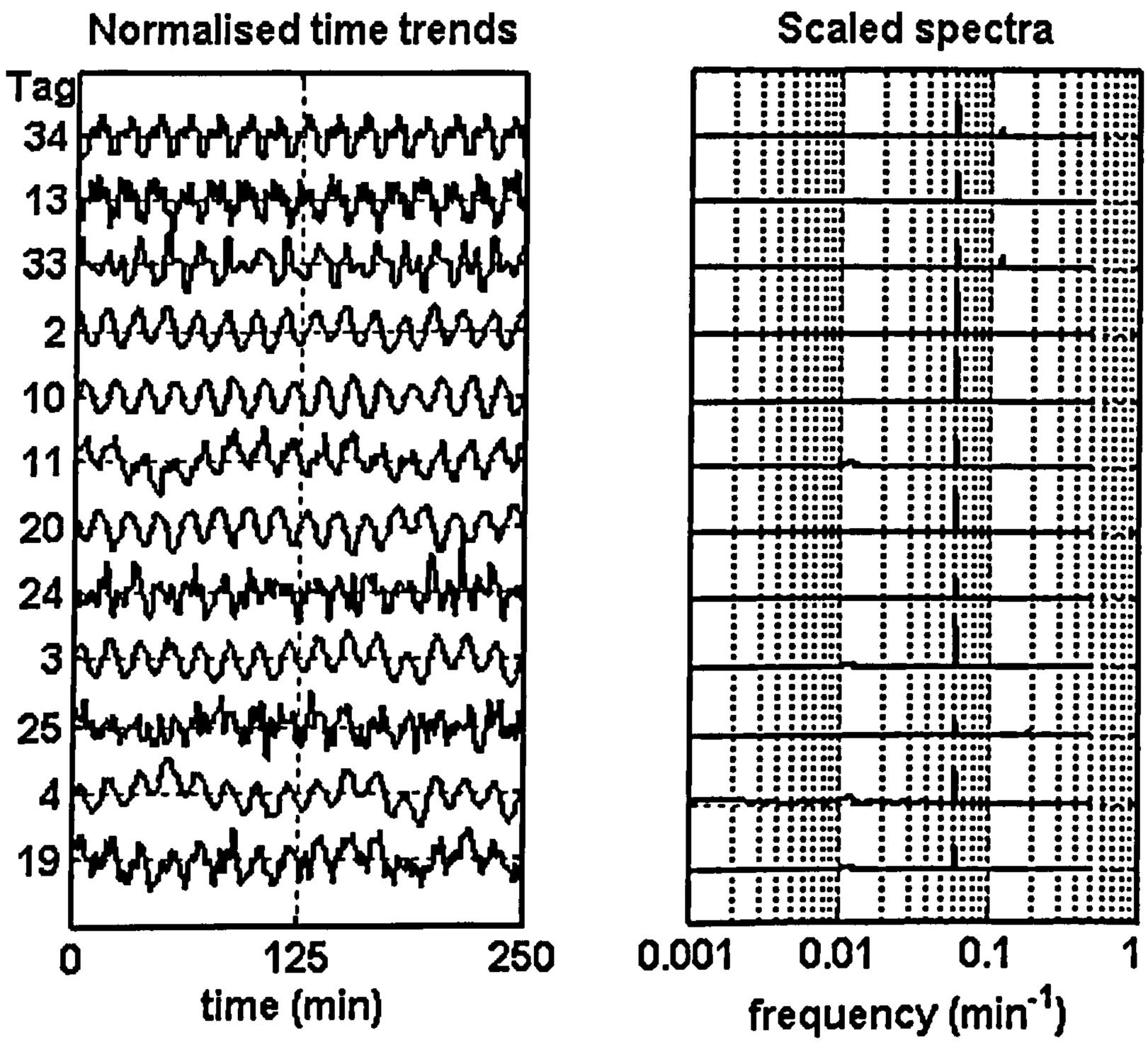


Figure 4-13: Time trends and spectra for the tags with a common fundamental frequency

Tag 33 and 24 were then eliminated from this group because their deviations were very small, which would make harmonic analysis unreliable. The 16.7-min period is known to be a high frequency so that low pass filtering was expected and therefore the source of such plant-wide oscillations should have the minimum bi-amplitude ratio. Table 4-4 gives their bi-amplitude ratios. The maximum bi-amplitude peak for the loop pertaining to Tag4 was found to be located at bi-frequency (0.058,0.0019) instead, raising the suspicion that the signal was corrupted by low-frequency noise. Tag4 was therefore eliminated from the search. Figure 4-14 shows bi-amplitude contour and surface plots for the two loops, Tag25 and Tag34. The plot of Tag34 shows distinct the first peak at bi-frequencies (0.06, 0.06); in contrast, the plot of Tag25 contains more features, the bi-amplitude peak at (0.12, 0.06) is not as evident as the peak at (0.18, 0.06) implying that the power at the third harmonic frequency f=0.18 is larger than the power at the second harmonic frequency f=0.12. Thus it can be associated with Situation 3 in Section 4.2.4, and concluded that non-Gaussian noise has enlarged the third harmonic, so that the small bi-amplitude ratio of Tag25 is not due to a strong third harmonic but to a strong content of noise around the third harmonic frequency. Thornhill, Shah & Huang (2001) have reached a similar conclusion about Tag25.

Tag	B_1	B_2	Ratio (r)
34	0.1	4.8e-3	20.92
13	0.056	1.1e-3	49.82
2	0.036	1.58e-4	229.67
10	0.01	2.12e-4	46.37
11	0.029	2.6e-3	**11.1**
20	0.036	1.42e-4	253.98
3	0.014	1.61e-5	880.68
25	0.021	2.3e-3	(8.89)
4	(0.01)	(1.1e-4)	-
19	0.028	5.28e-4	52.91

Table 4-4: Bi-amplitude ratios for the 12 loops in the high frequency case study

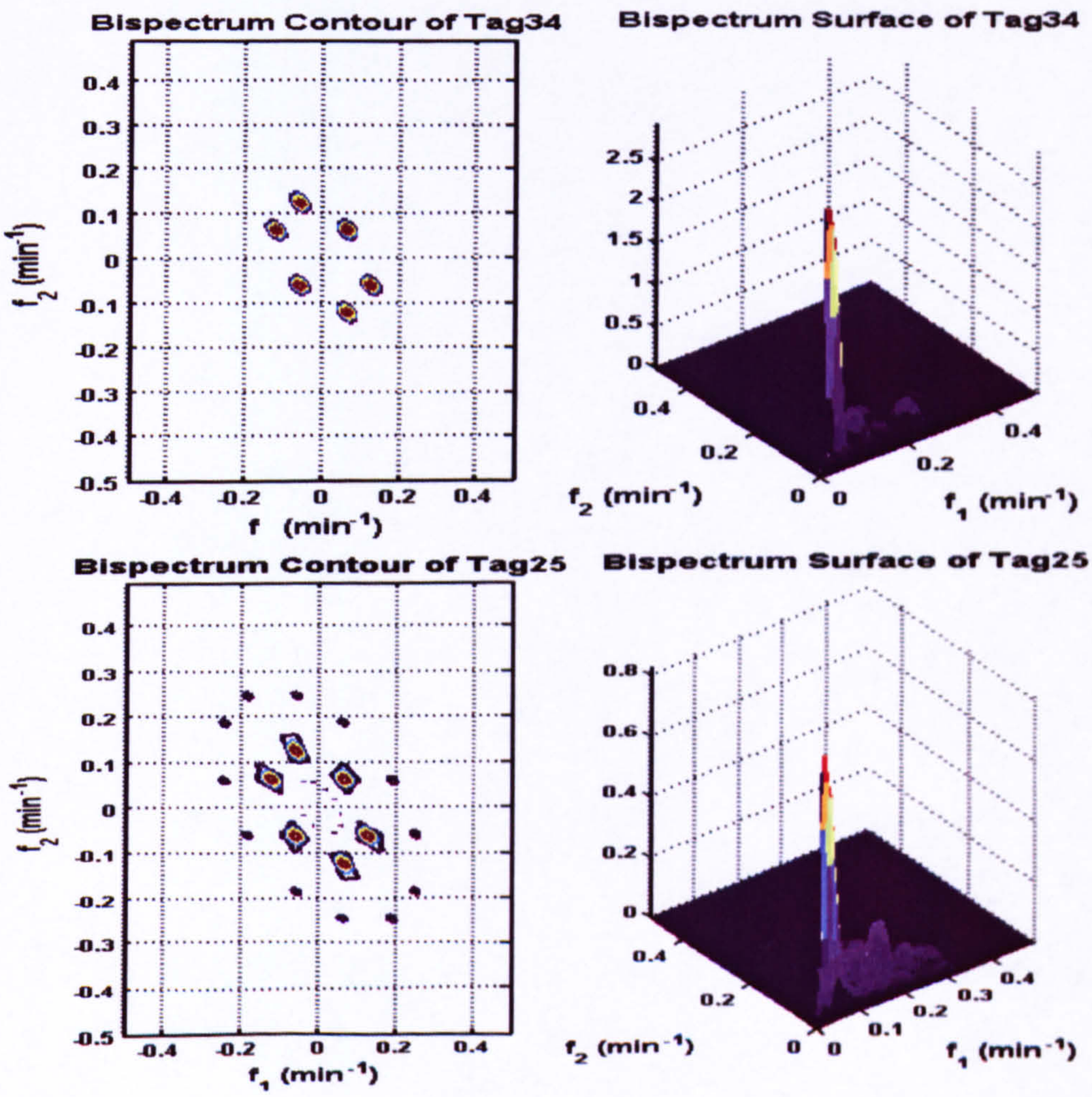


Figure 4-14: Bi-amplitude plots for Tag34 and Tag25

The results in Table 4-4 suggest that the root cause is in the vicinity of Tag11 because this has the smallest biamplitude ratio. Tag34 is the next likely candidate. Tag11 was verified by examining the ratio of the fundamental to the fifth harmonic (Table 4-5). Although the Tag11 ratio (r^5) is small, that pertaining to Tag34 is even smaller. Plant knowledge is now needed to progress the investigation further. A reason for this contradictory evidence can be obtained by looking at the plant schematic (Thornhill, Shah and Huang, 2001). Tag11 and Tag34 are in the same process unit: Tag11 is the product indicator whilst Tag34 is a flow loop. They have similar fundamental to fifth harmonic ratios (r^5) because they are physically close. But Tag34 is more likely to be associated with the root cause, because the plant-wide oscillation is very likely to have been caused by a valve nonlinearity problem.

Tag	B_2 (0.12,0.06)	B_3 (0.12,0.18)	Ratio (r^5)
34	4.8e-3	1.6e-3	**3.0**
13	1.1e-3	1.95e-4	5.64
2	1.58e-4	9.22e-7	171.37
10	2.12e-4	2.17e-6	97.70
11	2.6e-3	4.28e-4	6.07
20	1.42e-4	2.93e-7	484.64
3	1.61e-5	3.41e-8	472.14
25	2.3e-3	4.90e-4	4.69
4	(1.1e-4)	(1.31e-6)	-
19	5.28e-4	1.22e-5	43.28

Table 4-5: Fifth Bi-amplitude ratio indices for the first case study

A Simple 4-loop Example

The first example has a structure that might be obtained from a plant, which has a number of process units connected in series:

$$G(s)=\begin{bmatrix} \dfrac{2.22e^{-2.5s}}{(36s+1)(25s+1)} & 0 & 0 & 0 \\ \dfrac{-2.33e^{-5s}}{(35s+1)^2(25s+1)} & \dfrac{3.46e^{-1.01s}}{(32s+1)(25s+1)} & 0 & 0 \\ 0 & \dfrac{3.511e^{-13s}}{(12s+1)^2} & \dfrac{4.41e^{-1.01s}}{(16.2s+1)(25s+1)} & 0 \\ 0 & 0 & \dfrac{-1.25e^{-2.8s}}{(43.6s+1)(9s+1)} & \dfrac{4.78e^{-1.15s}}{(48s+1)(5s+1)} \end{bmatrix}$$

$$C_1(s)=0.385\left(1+\frac{1}{34.72s}\right),\ C_2(s)=6.19\left(1+\frac{1}{60.8s}\right),$$
$$C_3(s)=2.836\left(1+\frac{1}{60.22s}\right),\ C_4(s)=0.732\left(1+\frac{1}{60.93s}\right)$$

A stiction model (Horch & Isaksson 1998) was incorporated in **Loop1** to produce the oscillations. An oscillation period of about 55 min/cycle was chosen as this is approximately the same as the resonant periods of the loops. (As will be seen, the close proximity of the period of oscillation with that of resonance poses difficulties in the next example). The bi-amplitude ratios of the four loops are listed in Table 4-6. Application of the procedure of Figure 4-7 results in the isolation of Loop1. Although not used explicitly, the 3 log-ratio plots are given in Figure 4-15 to enable the reader to note how close the resonant periods are to the integral time constants.

		Target Loop			
		1	2	3	4
	T	34.72	60.8	60.22	60.93
Source Loop	1	-	-18.59	-27.82	-10.83
	2	18.59	-	-9.23	7.75
	3	27.82	9.23	-	16.98
	4	10.83	-7.75	-16.98	-

Table 4-6: Scaled relative bi-amplitude ratios (simple case)

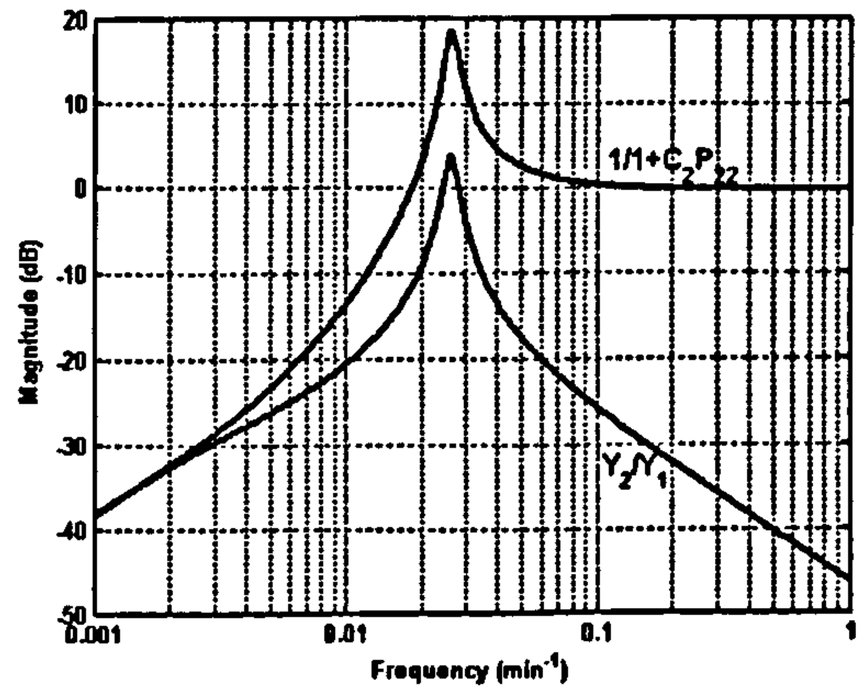


Figure 4-15: Log-ratio plots for the simple example

A Tightly Coupled 4-Loop Plant

The 4 × 4 process simulated was obtained from Lee *et al.* (1998):

$$G(s)=\begin{bmatrix} \frac{2.22e^{-2.5s}}{(36s+1)(25s+1)} & \frac{-2.94(7.9s+1)e^{-0.05s}}{(23.7s+1)^2} & \frac{0.017e^{-0.2s}}{(31.6s+1)(7s+1)} & \frac{-0.64e^{-20s}}{(29s+1)^2} \\ \frac{-2.33e^{-5s}}{(35s+1)^2} & \frac{3.46e^{-1.01s}}{32s+1} & \frac{-0.51e^{-7.5s}}{(32s+1)^2} & \frac{1.68e^{-2s}}{(28s+1)^2} \\ \frac{-1.06e^{-22s}}{(17s+1)^2} & \frac{3.511e^{-13s}}{(12s+1)^2} & \frac{4.41e^{-1.01s}}{16.2s+1} & \frac{-5.38e^{-0.5s}}{17s+1} \\ \frac{-5.73e^{-2.5s}}{(50s+1)(8s+1)} & \frac{4.32(25s+1)e^{-0.01s}}{(50s+1)(5s+1)} & \frac{-1.25e^{-2.8s}}{(43.6s+1)(9s+1)} & \frac{4.78e^{-1.15s}}{(48s+1)(5s+1)} \end{bmatrix}$$

$$C_1(s)=0.385\left(1+\frac{1}{34.72s}\right),\ C_2(s)=6.19\left(1+\frac{1}{21.8s}\right),$$

$$C_3(s)=2.836\left(1+\frac{1}{19.22s}\right),\ C_4(s)=0.732\left(1+\frac{1}{36.93s}\right)$$

Particular difficult features of this simulation are how full the 4 × 4 matrix is, and the low orders of the off-diagonal terms. This simulation was intended for control studies and was less concerned with model validity at higher frequencies where it is unlikely to be realistic. Two of the consequences here are that horizontal plateaus are to be expected in many of the log-ratio plots and the integral time constant based estimation of ω_{rok} is unlikely to give accurate results.

1. Low frequency (256 min/cycle, Loop1 is the source loop)

A stiction model (Horch and Isaksson 1998) was incorporated in **Loop1** to produce the oscillations. The bi-amplitude ratios of the four loops are listed in Table 4-7. The oscillation period (256 min/cycle) is quite high relative to the integral time

constants of the controller settings so the ratios in the source loop row should all be positive. Thus Loop1 would be marked.

		Target Loop			
		1	2	3	4
Source Loop	1	-	20.16	3.49	2.01
	2	-20.16	-	-16.67	-18.16
	3	-3.49	16.67	-	-1.48
	4	-2.01	18.16	1.48	-

Table 4-7: Scaled relative bi-amplitude ratios (tightly coupled, low-frequency case)

2. Middle frequency (25.6 min/cycle, Loop2 is the source loop)

The same stiction model was now incorporated in **Loop2** to produce a different oscillation. A 25.6 min/cycle oscillating period was chosen as this is of the same magnitude as the integral time constants of the plant, making interpretation more complicated. The bi-amplitude ratios of the four loops are listed in Table 4-8. All the scaled ratios are small relative to typical gradients of about 20, 40 or 60; this is to be expected when the period of the oscillations is so close to the integral time constants. It is now pointless invoking the procedure in Figure 4-7 as slopes cannot be discerned in this data. To examine this situation further, log-ratio plots (Figure 4-16) were generated analytically by re-arranging, then evaluating the model equations as described in Appendix C. As anticipated, their resonant peaks are not in the range 0.027 – 0.052 predicted from the integral time constants and there are plateaus at high frequencies in two of the plots. The $0.04 min^{-1}$ (25.6 min/cycle) frequency of the oscillations is higher than the resonant frequency of Loop 1 but

lower than the resonant frequencies of Loop3 and Loop4, which explains the results in Table 4-8. The procedure in Figure 4-7 successfully marked Loop2 when the resonant frequencies were estimated from these log-ratio plots.

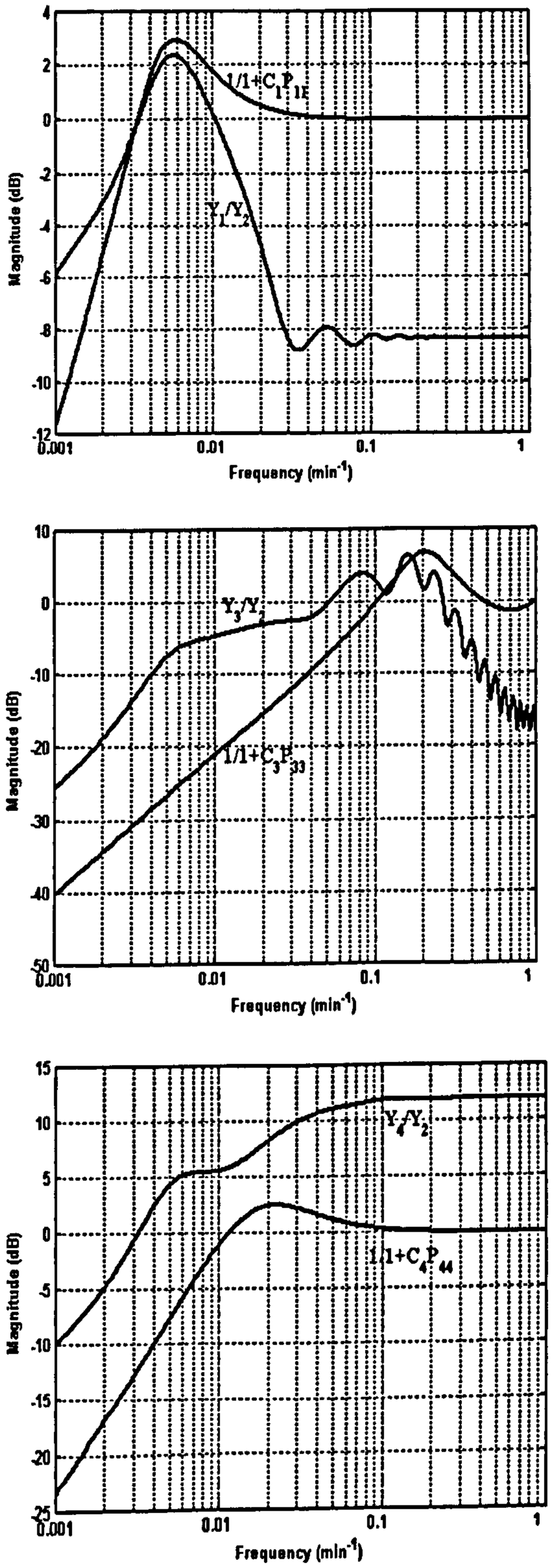


Figure 4-16: Log-ratio plots for the middle-frequency case

		Target Loop			
		1	2	3	4
Source Loop	1	-	2.55	5.89	4.86
	2	-2.55	-	3.34	2.30
	3	-5.89	-3.34	-	-1.04
	4	-4.86	-2.30	1.04	-

Table 4-8: Scaled relative bi-amplitude ratios (tightly coupled, middle-frequency case)

3. High frequency (4.1 min/cycle, Loop2 is the source loop)

Coulomb & viscous friction was incorporated in Loop2 to produce a high-frequency oscillation (4.1 min/cycle). The results (Table 4-9) show significant, negative relative bi-amplitude ratios for Loops 1, 2 and 4 as the sources. To examine this situation further, once again log-ratio plots were generated analytically from the model equations and were used to estimate the resonant frequencies. Loop 1 and Loop 4 were now eliminated on the basis of the procedure of Figure 4-7.

		Target Loop			
		1	2	3	4
Source Loop	1	-	2.09	-29.38	0.32
	2	-2.09	-	-31.46	-1.76
	3	29.38	31.46	-	29.7
	4	-0.32	1.76	-29.7	-

Table 4-9: Scaled relative bi-amplitude ratios (tightly coupled, high-frequency case)

4.3.6 Isolating Multiple Oscillations Based on The Biamplitude Ratio Index

The bispectral approach has the inherent capability of detecting and diagnosing multiple oscillations. A simple example is illustrated here to demonstrate a general procedure for isolating multiple sources of oscillations. The simulated plant (Figure 4-17) has the same process models and controllers as shown in Figure 4-8, but with independent oscillatory disturbances introduced into both loops during steady state. The oscillatory disturbance introduced into Loop1 is a sinusoidal signal with a fundamental, a second and a third harmonics:

$$s_1(i) = \sin(2\pi f_{01} i) + 0.5\sin(4\pi f_{01} i) + 0.2\sin(6\pi f_{01} i), i = 1, \cdots, 4096 \qquad (4.22)$$

where f_{01}=0.005 min^{-1} with (min^{-1}) being frequency/sampling frequency ; the disturbance introduced into Loop2 is:

$$s_2(i) = 0.8\sin(2\pi f_{02} i) + 0.4\sin(4\pi f_{02} i) + 0.1\sin(6\pi f_{02} i), i = 1, \cdots, 4096 \qquad (4.23)$$

where f_{02}=0.035 min^{-1}.

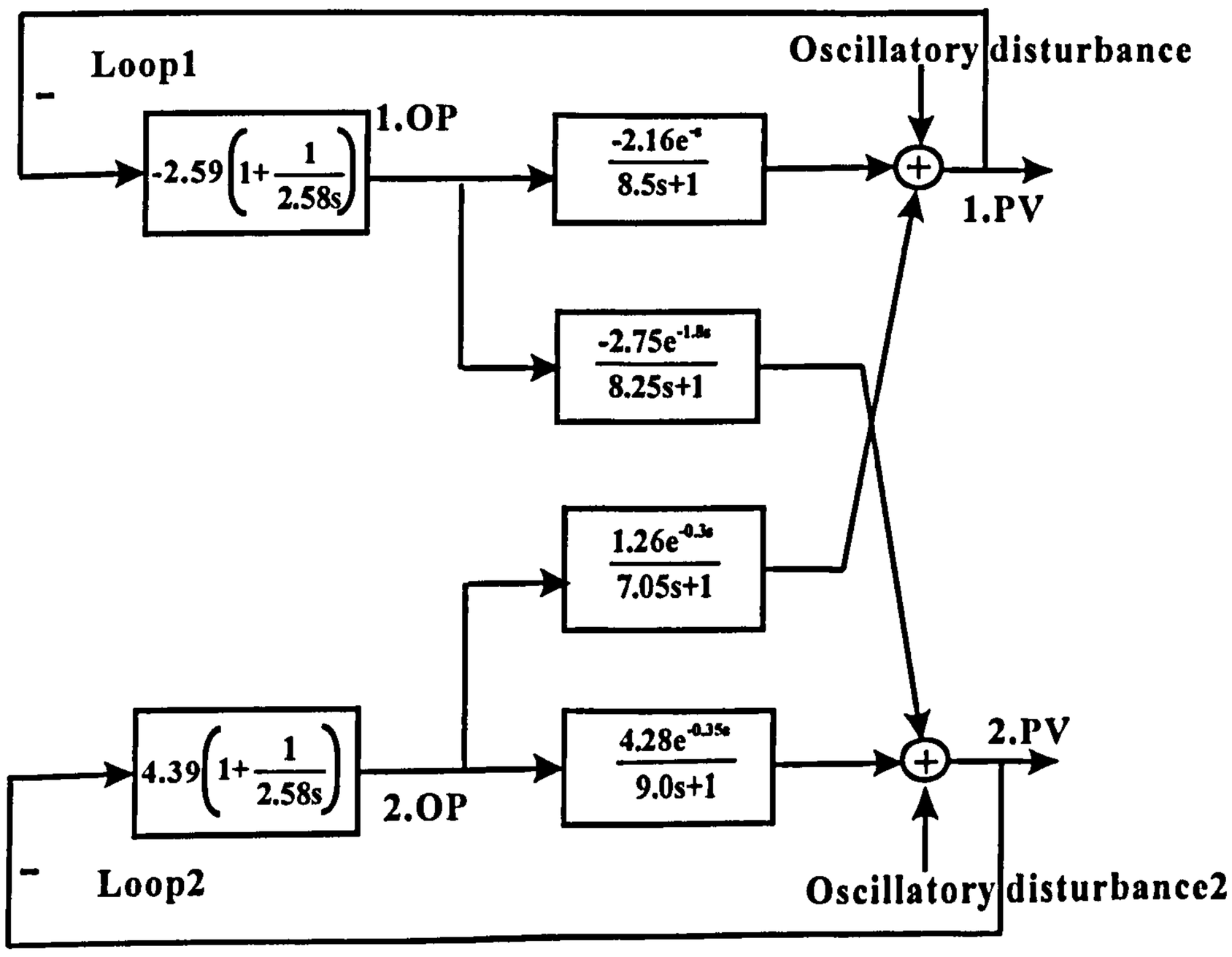


Figure 4-17. The simulated plant with multiple oscillations

These multiple oscillations are visible in the process variable time series and power spectra shown in Figure 4-18. Loop1 has dominant peaks at frequencies f_{01}, $2f_{01}$ and also a dominant peak at frequency f_{02} because of propagation, whilst Loop2 has dominant peaks at frequencies f_{02} and $2f_{02}$. Both loops have suffered from the oscillatory disturbance propagated from the other loop.

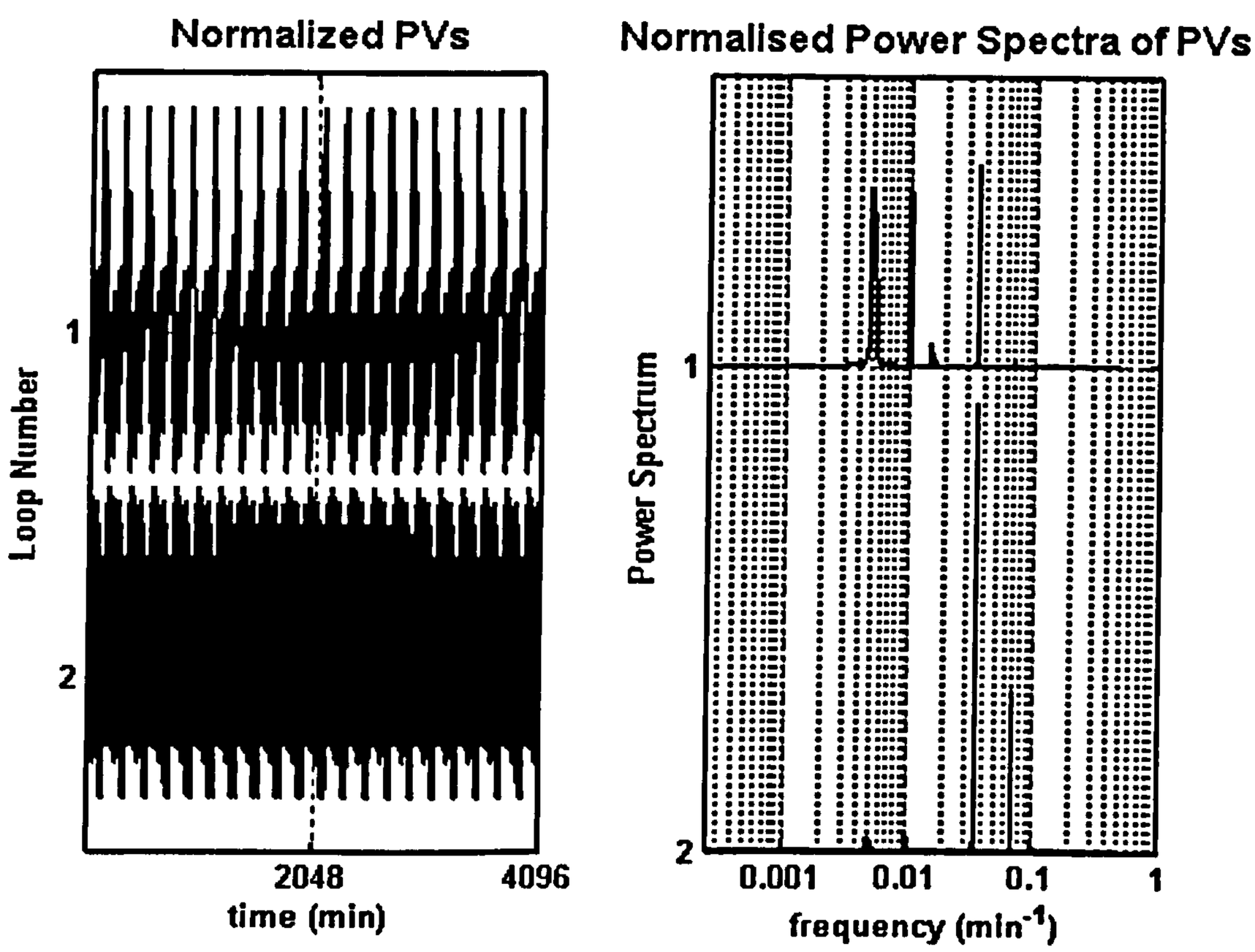


Figure 4-18: Process variable time series and power spectra for the simulation with multiple oscillations

From their power spectra it is hard to say which loop is nearer to the possible sources of the multiple oscillations. The magnitude bispectrum is now analysed. Figure 4-19 gives the contour biamplitude plots of both loops. It can be seen that the biamplitude of Loop1 shows explicit peaks at bifrequencies (f_{01}, f_{01}) and $(2f_{01}, f_{01})$ whilst the biamplitude of Loop2 only shows peaks at bifrequency $(f_{02}\, f_{02})$. It can be

concluded that Loop1 is nearer to the oscillation with fundamental f_{01}, and Loop2 is one of those loops suffered from the oscillation with fundamental f_{02}. Now Loop1 and Loop2 are categorized into parts pertaining to different oscillations. It is noted that this categorization procedure is automatic. All loops disturbed by these oscillations can be categorized into two parts in this way. In each category single-source oscillations can be assumed and the isolating procedure based on the biamplitude ratio index can be performed.

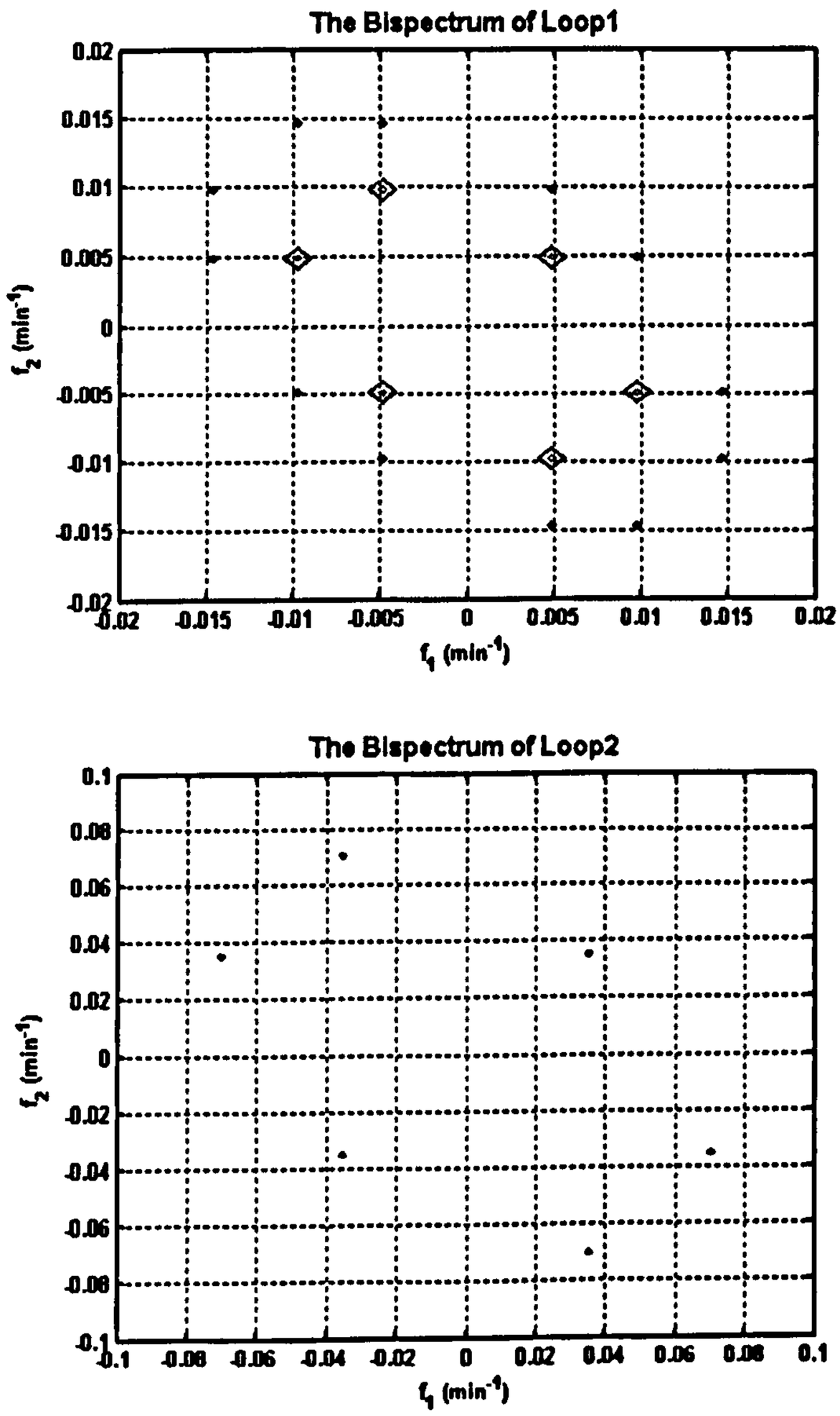


Figure 4-19: Magnitude bispectrum of both loops

Chapter 5

Root Cause Diagnosis Based on Non-linear Time Series Analysis and Spectral ICA

Having gained an understanding of harmonic propagation, harmonic evaluation methods should aid the isolation of whole-plant oscillations. In this chapter two indices based on non-linear time series analysis tools, namely *Correlation dimension* and *maximal Lyapunov exponent* are introduced to evaluate harmonic content. When combined with knowledge of harmonic propagation they are found to give indications of the root cause of non-linearity induced oscillations. The new application of the two time series analysis tools on the isolation of plant-wide oscillations is the main contribution in this chapter.

Since harmonics could be attenuated if the oscillatory frequency is larger than the cut-off frequency of the log-ratio plot, a revised interpretation of spectral ICA is required in high frequency cases.

5.1 Introduction

The correlation dimension and Lyapunov exponents are recently developed descriptions which provide quantities for the characterisation of nonlinear, deterministic and chaotic data (Kantz & Schreiber 1997). These methods have been widely used to analyse non-linear time series in mechanical systems (Boltezar *et al.*

1999; Cusumano & Moon 1995; Jaksic *et al.* 1999; Trendafilova & Van Brussel 2001; Wang & Lin 2003), medical science (Burioka *et al.* 2001; Hecox *et al.* 2003; Lei *et al.* 2001; Muller *et al.* 2003), and chemical engineering (Guo *et al.* 2003; Hay *et al.* 1995; Wang *et al.* 2003). Sometimes the correlation dimension and/or the maximal Lyapunov exponent indicate non-linear or deterministic chaotic behaviour; whilst other references conclude that a larger correlation dimension or a larger maximal Lyapunov exponent corresponds to richer dynamics (Boltezar *et al.* 1999) or more system complexity (Boltezar & Hammond 1999; Burioka *et al.* 2001; Prokoph & Veizer 1999). Few have related the correlation dimension or the maximal Lyapunov exponent of a time series to its harmonic content because chaotic time series, although usually periodic and oscillatory, do not contain distinct fundamentals and harmonics.

In this chapter the definitions and characteristics of the correlation dimension and maximal Lyapunov exponent will first be introduced. A mathematical analysis and a simple example will then be used to demonstrate their relationships with the harmonic content of an oscillating time series. Based on the understanding of harmonic propagation, these measures are found to give indications of the root cause of oscillations. Data obtained from a simulated plant and two industrial case studies will be analysed to demonstrate this approach.

5.2 Correlation Dimension and Maximal Lyapunov Exponent Based Root Cause Isolation

5.2.1 Definition and Estimation

Both techniques are based on a discrete, multi-dimensional phase space representation of the data. Given the time series $x_1, x_2, x_3, \cdots, x_N$ then the multi-dimensional phase space is formed from:

$$\mathbf{X}_i = (x_i, x_{i+T}, x_{i+2T}, \cdots x_{i+(m-1)T}), \quad i = 1, 2, \cdots, N \tag{5.1}$$

where T is the time interval and m is known as the embedding dimension; $\mathbf{X}_i$ is called an embedding vector (point) of m-dimension.

Correlation dimension

The correlation dimension provides a tool to quantify self-similarity. A larger correlation dimension corresponds to a larger degree of complexity and less self-similarity. Stochastic signals are infinite-dimensional. The most frequently used procedure to estimate the correlation dimension was introduced by Grassberger and Procaccia (1983a; 1983b). They defined the correlation sum for a collection of points $\mathbf{X}_i$ ($i=1, 2,..., N$) in some phase space to be the fraction of all possible pairs of points which are closer than a given distance ε in a particular norm:

$$C(m, \varepsilon) = \frac{2}{(N-m)(N-m-1)} \sum_{i=m}^{N} \sum_{j=i+1}^{N} \Theta(\varepsilon - \|\mathbf{X}_i - \mathbf{X}_j\|) \tag{5.2}$$

where Θ is the Heaviside step function, $\Theta(x)=0$ if $x \leq 0$ and $\Theta(x)=1$ for $x>0$. Thus Equation 5.2 counts the pairs $(\mathbf{X}_i, \mathbf{X}_j)$ whose distance is smaller than ε. To eliminate the temporal correlation, the Theiler window can be used to exclude those pairs of points that are too close in time (Kantz and Schreiber 1997).

It has been shown by Sauer & Yorke (1993) that in the limit of an infinite amount of data (i.e. $N \to \infty$) and for small ε, C scales like a power law, $C(\varepsilon) \propto \varepsilon^{D_2}$, where D_2 is known as the correlation dimension. Thus D_2 is defined by

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\partial \ln C(m, \varepsilon)}{\partial \ln(\varepsilon)} \tag{5.3}$$

Convergence to a finite correlation dimension can be checked by plotting "effective dimensions" versus scale (ε) for various embeddings (m). The easiest way to proceed is to compute (numerically) the derivative of $\ln C(m, \varepsilon)$ with respect to $\ln(\varepsilon)$, for example by fitting straight lines to the log-log plot of $C(\varepsilon)$. If the scale is large, then self-similarity should be minimal, whereas if it is sufficiently small one observes a relationship with the embedding dimension i.e. D_2 is not totally independent. This effect is due to noise. Only on the intermediate scales can one see the desired *plateau* where the results are approximately independent of m and ε i.e. where D_2 is *invariant.* The region where the scale rule holds, not just the range selected for straight line fitting, is called the *scaling range.* Takens-Theiler have developed an alternative estimate of the correlation dimension:

$$D_{TT}(\varepsilon) = \frac{C(\varepsilon)}{\int_0^{\varepsilon} \frac{C(\varepsilon')}{\varepsilon'} d\varepsilon'} \tag{5.4}$$

Maximal Lyapunov Exponent

Lyapunov exponents measure the exponential divergence (positive exponents: chaotic motion) or convergence (negative exponents: regular motion) of two initially neighbouring trajectories in a phase space. In other words, they measure the degree of unpredictability of the future. There are many different Lyapunov exponents for a dynamical system. The most important is known as the *maximal Lyapunov exponent* (λ_1) (Kantz and Schreiber 1997). Let $\mathbf{X}_i$ and $\mathbf{X}_j$ be two points in a phase space with a distance $\|\mathbf{X}_i - \mathbf{X}_j\| = \delta_0 \ll 1$ between them. Let δ_t denote the distance obtained between the two trajectories t units of time later i.e. $\delta_t = \|\mathbf{X}_{i+t} - \mathbf{X}_{j+t}\|$. Then λ_1 is defined by $\delta_t \simeq \delta_0 e^{\lambda_1 t}, \delta_t \ll 1$. A positive λ_1 means that there is an exponential divergence of these trajectories, i.e. chaos; a negative λ_1 implies the existence of a stable fixed point; if the motion settles down onto a limit cycle, λ_1 is zero.

A robust, consistent and unbiased estimator for the maximal Lyapunov exponent was proposed by Kantz and Schreiber (1997). They compute

$$S(\varepsilon, m, t) = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\frac{1}{|U(\mathbf{X}_i)|}\sum_{\mathbf{X}_j \in U(\mathbf{X}_i)} |\mathbf{X}_{i+t} - \mathbf{X}_{j+t}|\right) \tag{5.5}$$

where the reference points $\mathbf{X}_i$ are embedding vectors and $U(\mathbf{X}_i)$ is known as the neighbourhood of $\mathbf{X}_i$ with diameter ε. If $S(\varepsilon, m, t)$ exhibits a linear increase with fixed slope for all m larger than some m_0 and for a reasonable range of ε, then the slope of S is deemed to be *invariant* over this range. This slope can be taken as an estimate of the maximal exponent λ_1.

Automatic generation of the invariant estimates

Estimation of the correlation sum, correlation dimension, and the maximal Lyapunov exponent were all based on the TISEAN software package(Hegger *et al.* 1999). The automatic generation of the invariant estimates were derived as follows:

(1) Obtain the appropriate time interval T and embedding dimension m for the given time series; the mutual information method (Fraser & Swinney 1986) is used to choose T and Cao's method (Cao 1997) is used to obtain m.

(2) *Correlation dimension estimate*: apply routines *d2* and *c2t* of the TISEAN toolkit with the parameters T and m to obtain D_{TT} for a specified range of ε, i.e. the derivatives of $\ln C(m,\varepsilon)$ versus $\ln(\varepsilon)$ for 1:m+4 embeddings, which is a matrix of size length(ε) × (m+4) .

Maximal Lypunov exponent estimate: apply routine *lyap_k* to obtain $S(\varepsilon,m,t)$ versus t for 1:m+4 embeddings, which is a matrix of size length(t) × (m+4) .

(3) *Correlation dimension estimate*: calculate the *mean* and *variance* of D_{TT} for the last 4 embeddings and then the *normalized variance* (*variance/mean*), which are vectors of size (length(ε)). If the normalized variance is smaller than a given threshold, average the mean D_{TT} within the range of ε's where the normalized variances are small. Take the average value as the correlation dimension estimate. Usually the threshold for the normalized variance is given differently for various time series. A larger value is set for a signal with more noise content.

Maximal Lypunov exponent estimate: for each of the last 4 embeddings, fit a line to the calculated $S(\varepsilon,m,t)$ within the range of the first several number of

t's and calculate the slopes of the lines. The range of fitted t's is different for various time series. An automatic way is to try fitting lines for $S(\varepsilon, m, 1:num)$ with num moving from 2:t until the fitting error is minimum. Take the average value of these four slopes as the maximal Lyapunov exponent estimate.

5.2.2 Relationship Between The Invariants and Harmonic Content

Intuitively, a time series with high harmonic content will be more complex and more unpredictable than one with a low harmonic content, and hence will have a larger correlation dimension and a larger maximal Lyapunov exponent. To demonstrate this the maximal Lyapunov exponent for a sinusoidal signal with up to three harmonics is obtained analytically in Appendix D. The maximal Lyapunov exponent is found to increase with increasing harmonic content. Here a simple calculation will be used to reinforce this point. Three 4096-sample time series were analysed to examine the effect of increasing the harmonic content:

$$\begin{aligned} s1(k) &= \sin(2\pi fk) \\ s2(k) &= \sin(2\pi fk) + 0.8\sin(4\pi fk) \\ s3(k) &= \sin(2\pi fk) + 0.8\sin(4\pi fk) + 0.5\sin(6\pi fk) \\ k &= 1, \cdots, 4096 \end{aligned} \tag{5.6}$$

where f=*0.01* is the fundamental frequency. The correlation dimensions and maximal Lyapunov exponents of the three time series are listed in Table 5-1.

Time series	Correlation dimension (D_2)	Maximal Lyapunov Exponent (λ_1)
s1(k)	0.93±0.02	0.011±0.002
s2(k)	1.06±0.01	0.026±0.002
s3(k)	1.13±0.02	0.031±0.002

Table 5-1: Invariant of sinusoidal signals

It can be seen that the correlation dimension and maximal Lyapunov exponent both increase with harmonic content. This supports the claims that the correlation dimension reflects the self-similarity and the maximal Lyapunov exponent reflects the predictability. The future of time series with more harmonics should be more unpredictable and hence should have a larger invariant.

5.2.3 Invariant Based Isolation of The Source of Oscillations

Again only process variables of control loops pertaining to a plant-wide oscillation are considered and the fundamental frequency of the oscillations is determined. If the frequency is high relative to the cut-off frequency, time series that are recorded close to the root cause will have a high harmonic content, and those recorded further away from the root cause will become more sinusoidal. However the first few harmonics of an oscillation could be amplified if the fundamental is sufficiently low to locate the first few harmonics on the positive slope of the relevant log-ratio plot. In this case, and since the first few harmonics dominate, the time records further away from the source could be more complex (with large signal-to-noise ratio).

Based on this understanding of how harmonic content is affected as an oscillation propagates through a plant, the correlation dimension and maximal Lyapunov exponent can provide measures to locate the source of plant-wide oscillations. More specifically, the time series pertaining to the root cause should have the largest correlation dimension and maximal Lyapunov exponent in cases where the plant is inherently low pass in nature over the frequency range of concern, because the harmonics have been attenuated through the propagation and also because the time series become less complex and more predictable further away from the source; and the correlation dimension and maximal Lyapunov exponent pertaining to the root cause will be the smallest in low-frequency cases where there is amplification of the first few harmonics.

5.2.4 The Effect of Noise

The correlation algorithm is a tool for analysing the scaling properties of point sets in phase space (Kantz and Schreiber 1997). Deterministic signals show power law scaling of $C(\varepsilon)$. It is characteristic for such signals that the exponent in the power law is *invariant* under coordinate transformations and in particular does not depend on the embedding dimension m once m is large enough to ensure a proper reconstruction. For stochastic systems the situation is different. Delay vectors formed from a random signal are not restricted to a low-dimensional manifold but fill all available directions in phase space. A deterministic signal with low noise level will have a correlation dimension plot in which the stochastic data dominates at small length scales (ε) comparable to the amplitude of the random noise. A consequence of much noise, however, is the breakdown of the self-similarity of the

signal. As for the maximal Lyapunov exponent, for a signal with much noise, nearby trajectories should diverge diffusively rather than exponentially. Figure 5-1 gives an example of such a situation, where 4096-sample normally distributed random noise time series with mean 0, variance 1, and standard deviation 1 is inspected. No scaling range is found in the correlation dimension plot and the maximal Lyapunov exponent is close to infinite.

Noise is an omnipresent phenomenon. One way to reduce the effect of noise is to apply the non-linear noise reduction technique (Kantz and Schreiber 1997). In the root cause diagnosis procedure, the source loop usually contains a high signal-to-noise ratio so that scaling and divergence should be visible. Those loops with measurement records having no obvious scaling or divergence are possibly far away from the root cause.

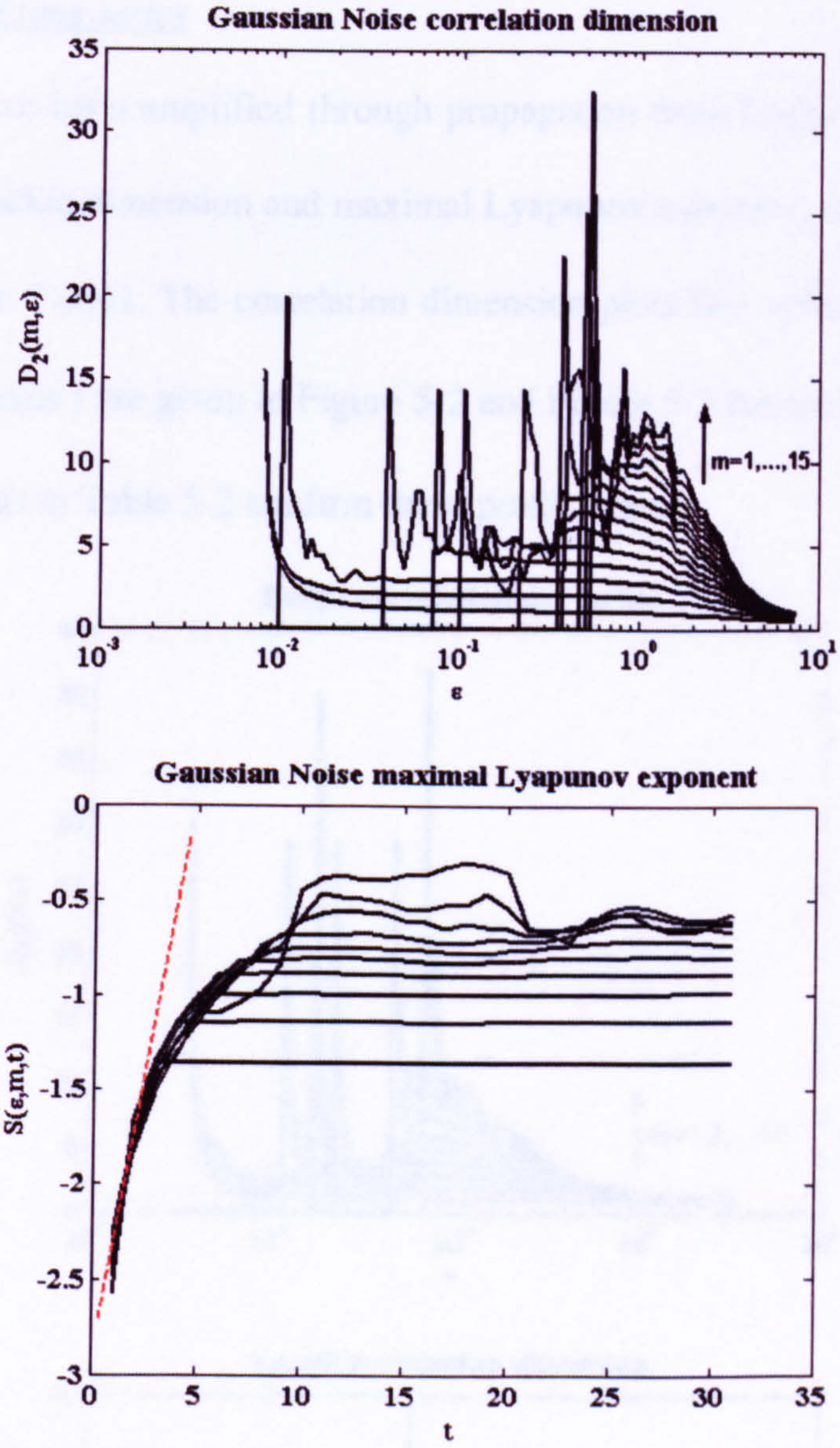


Figure 5-1: Correlation dimension and maximal Lyapunov exponent for random noise (red dashed line fitting the slope)

5.2.5 Examples

The three examples used in Section 4.3.5 are examined here again. Their harmonic propagation has already been evaluated with the bispectral analysis approach.

The Simulated time series

Harmonics have been amplified through propagation from Loop1 to Loop2, thus a smaller correlation dimension and maximal Lyapunov exponent, are expected in the source loop i.e. Loop1. The correlation dimension plots D_{TT} versus ε, and the plots of $S(\varepsilon, m, t)$ versus t are given in Figure 5-2 and Figure 5-3 respectively. The values of the invariants in Table 5-2 confirm these predictions.

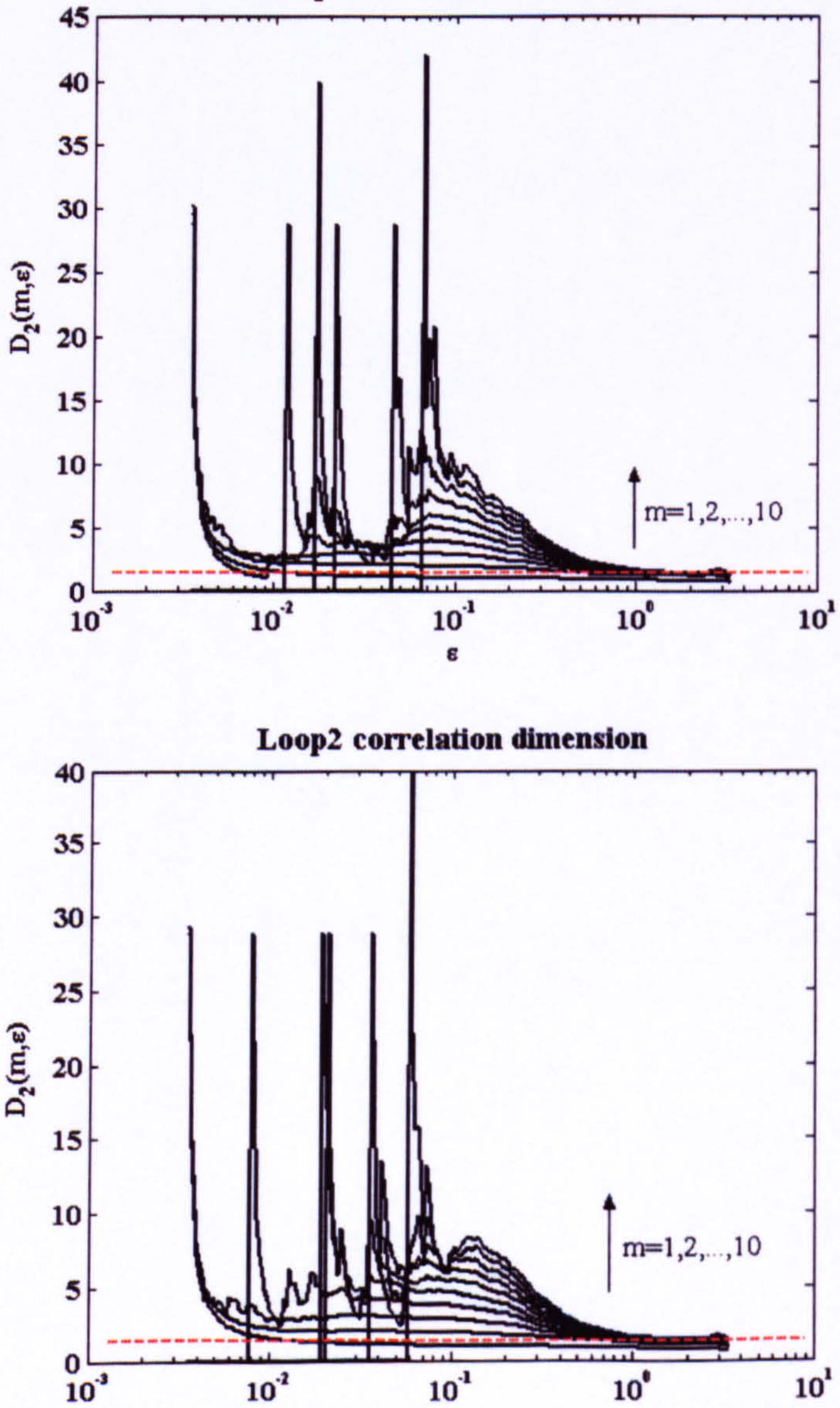


Figure 5-2: Takens estimator for correlation dimension of the two loops (red dashed line fitting the plateau)

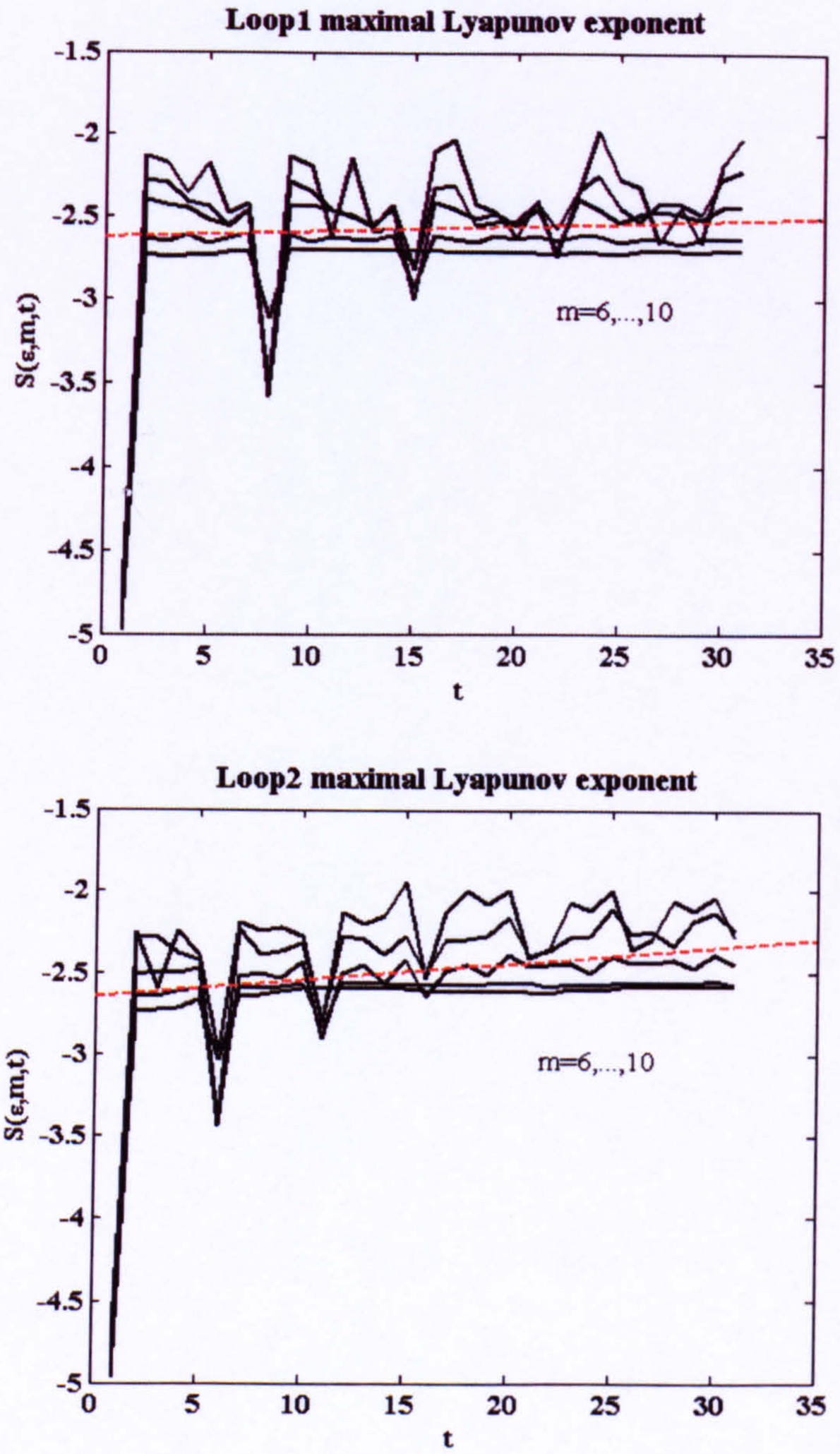


Figure 5-3: Maximal Lyapunov exponents of the two loops (red dashed line fitting the slope)

Loop Number	Correlation Dimension (D_2)	Maximal Lyapunov exponent (λ_1)
1	1.38±0.02	0.003±0.002
2	1.43±0.002	0.010±0.001

Table 5-2: Invariant estimation for the simulated case

Case study 1 (Low-Frequency Case)

The correlation dimensions and maximal Lyapunov exponents of each of the nine dominant time series are estimated. The results are listed in Table 5-3.

Tag Number	Correlation Dimension (D_2)	Maximal Lyapunov exponent (λ_1)
5	-	-
7	-	-
8	-	-
10	-	-
13	2.33±0.01	0.022±0.004
19	2.29±0.05	0.007±0.0004
22	1.52±0.03	0.004±0.0001
23	-	-
25	3.86±0.01	0.029±0.007

Table 5-3: Invariant estimation for the Eastman case study

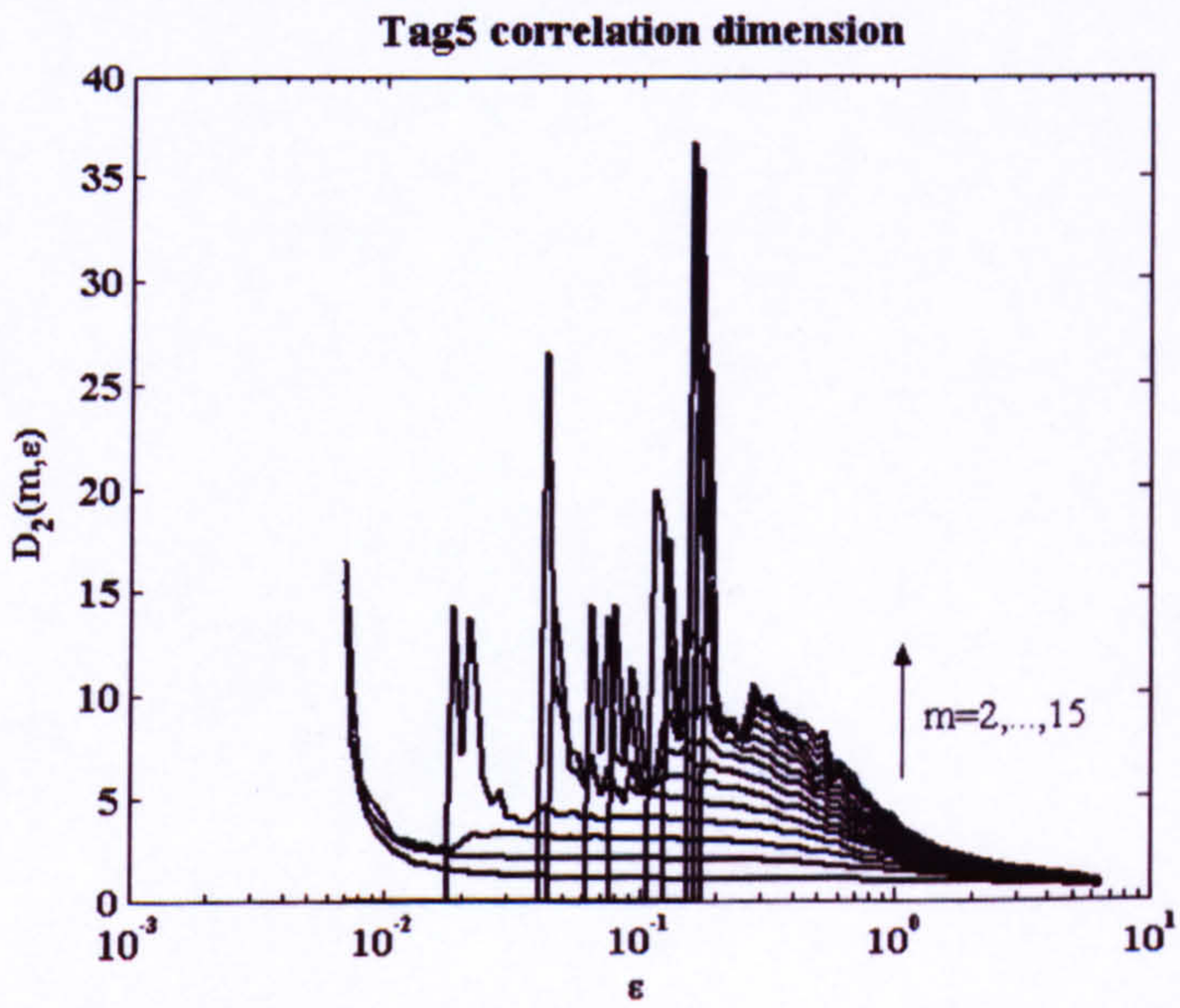


Figure 5-4(a): No typical correlation dimension found in Tag5

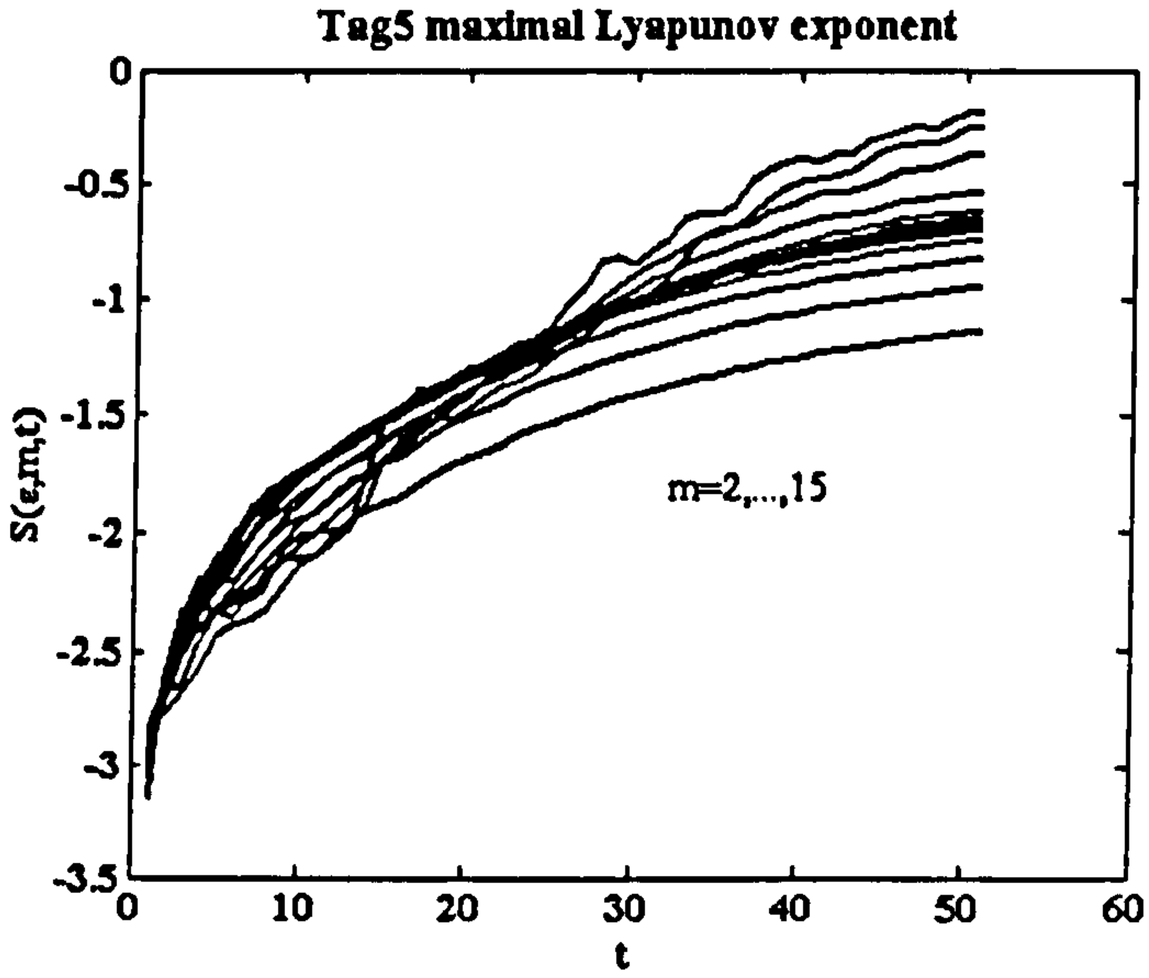


Figure 5-4(b): No exponential divergence of Tag5

Tag22 has the smallest correlation dimension and maximal Lyapunov exponent inferring that it should be the nearest to the root cause. Interestingly Figure 5-4(a) shows the correlation dimension plot of Tag5, where no typical correlation dimensions can be found. All curves from different embedding dimensions behave different and there is no common behaviour. Figure 5-4(b) shows that the time series pertaining to Tag5 exhibit no linear increase, reflecting the lack of exponential divergence of nearby trajectories. And the time series pertaining to Tag 7, 8, 10 and 23 also have no typical dimension and maximal Lypunov exponent. A possible explanation for this is that the stochastic content (for example, noise) in the time series dominates. These observations also match the propagation schematic provided by (Thornhill *et al.* 2002b), where Tag23, 5, 7 and 8 are all relatively far away from Tag22. Although Tag19 is farther away than Tag13, it has a smaller correlation dimension and maximal Lyapunov exponent. This is because the intrinsic low-pass filtering attenuates the higher harmonics and simplifies the time series pertaining to Tag19.

Case Study 2 (High-Frequency Case)

The estimated correlation dimensions and maximal Lyapunov exponents for the 12 loops are listed in Table 5-4. The 16.7-min period is known to be a high frequency so that low pass filtering was expected and therefore the source of such plant-wide oscillations should have the largest correlation dimension and maximal Lyapunov exponent. It can be seen that the time series pertaining to Tag33 have largest correlation dimensions and maximal Lyapunov exponents than those pertaining to other loops. Note that the time series pertaining to Tags 13, 11, 24, 3, 25, 4 and 19 have no typical correlation dimensions and maximal Lyapunov exponents, because these loops are suspected to be contaminated by noise.

Tag Number	Correlation Dimension (D_2)	Maximal Lyapunov exponent (λ_1)
34	1.19±0.03	0.011±0.0003
13	-	-
33	**1.94±0.02**	**0.030±0.004**
2	1.85±0.003	0.010±0.0002
10	1.58±0.09	0.011±0.001
11	-	-
20	1.74±0.05	0.019±0.001
24	-	-
3	-	-
25	-	-
4	-	-
19	-	-

Table 5-4: Invariant estimation for the SE Asian case study

5.3 A Revised Interpretation of The Spectral ICA Method

Spectral independent component analysis (*spectral ICA*) was reviewed in Chapter 2. This analysis has been proposed as an aid to isolate the root causes of multiple oscillatory sources (Xia & Howell 2003a). Its key feature is that it extracts dominant spectrum-like independent components each of which has a narrow-band peak that captures the behaviour of one of the oscillatory sources. The basis is that a unity significance index (*SI*) points to the source of an oscillation with a common fundamental frequency.

As has been mentioned in Chapter 2 the extracted independent components should be the harmonics of the oscillation. If the oscillation is amplified in target loops then the fundamental should dominate in the source loop, so that the source loop should have a unity significance index associated with the independent component pertaining to the fundamental frequency. Xia and Howell (2003a) gives examples of this case. Conversely, if the oscillation is attenuated the source shouldn't display the fundamental most strongly because it also contains harmonics. Then a smaller significance index associated with the independent component pertaining to the fundamental frequency is expected. In this case the independent components associated with the first few harmonics are likely to be extracted and the significance indices associated with these ICs must be assessed together to give an indication of the root cause. The revision on the present Spectral ICA approach is to sum the indices together, because the measurement pertaining to the root cause should have the largest harmonic content.

The following gives the spectral ICA analysis for the three case studies presented in Section 4.3.5. It will be seen that in the low or high frequency cases, the significance indices has to be interpreted differently.

5.3.1 The Simulated Time Series

Figure 5-5 shows the dominant ICs that were obtained when spectral ICA was performed on the process variables of the two loops in the simulated plant (see Figure 4-8). Values of the significance indices (i.e. the a_{ij} values) and the corresponding frequencies are listed in Table 5-5.

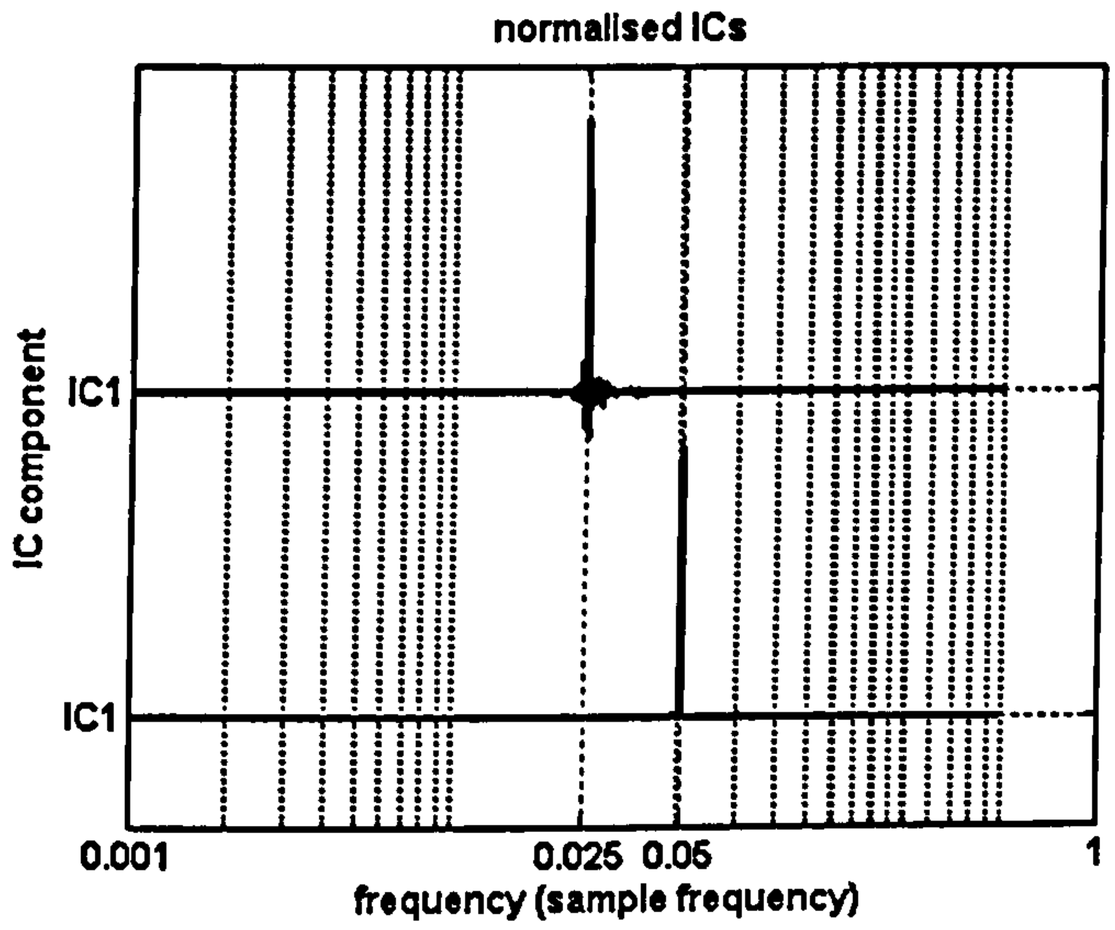


Figure 5-5: The dominant independent components for the simulated case

Significance index (*SI*)	IC1	IC2
Loop1.PV	1.00	0.17
Loop2.PV	0.97	1.00
Frequency (sample frequency)	0.025	0.05

Table 5-5: Values of significance indices and corresponding frequencies

The two dominant independent components (IC1 and IC2) are the fundamental (at frequency 0.025) and second harmonic (at frequency 0.05) respectively. Loop1 has a unity significance index for IC1 indicating its stronger link to the fundamental so that Loop1 is the root cause of the oscillation. Loop2 has a unity significance index for IC2, which represents the amplified harmonics that arise through propagation from Loop1.

5.3.2 Case study 1 (Low-Frequency Case)

In the low-frequency case study, straightforward spectral ICA failed to extract the harmonics that were expected so multi-range spectral ICA was applied instead. This meant that the ICs were extracted over a limited frequency range [0.0025 0.016] instead, which was chosen to include the three harmonics. Dominant ICs representing the fundamental, second and fourth harmonics were extracted and are shown in Figure 5-6. Table 5-6 gives the significance indices (*SI*) associated with the fundamental IC (IC1) and second harmonic IC (IC2) of the nine loops. The significance index of Tag22 associated with IC1 is the largest as expected. The significance indices associated with IC2 are all negative (and hence set to zero), thus indicating a weak association with IC2. Note also that there was little power in the second harmonic frequency channels (see Figure 4-12).

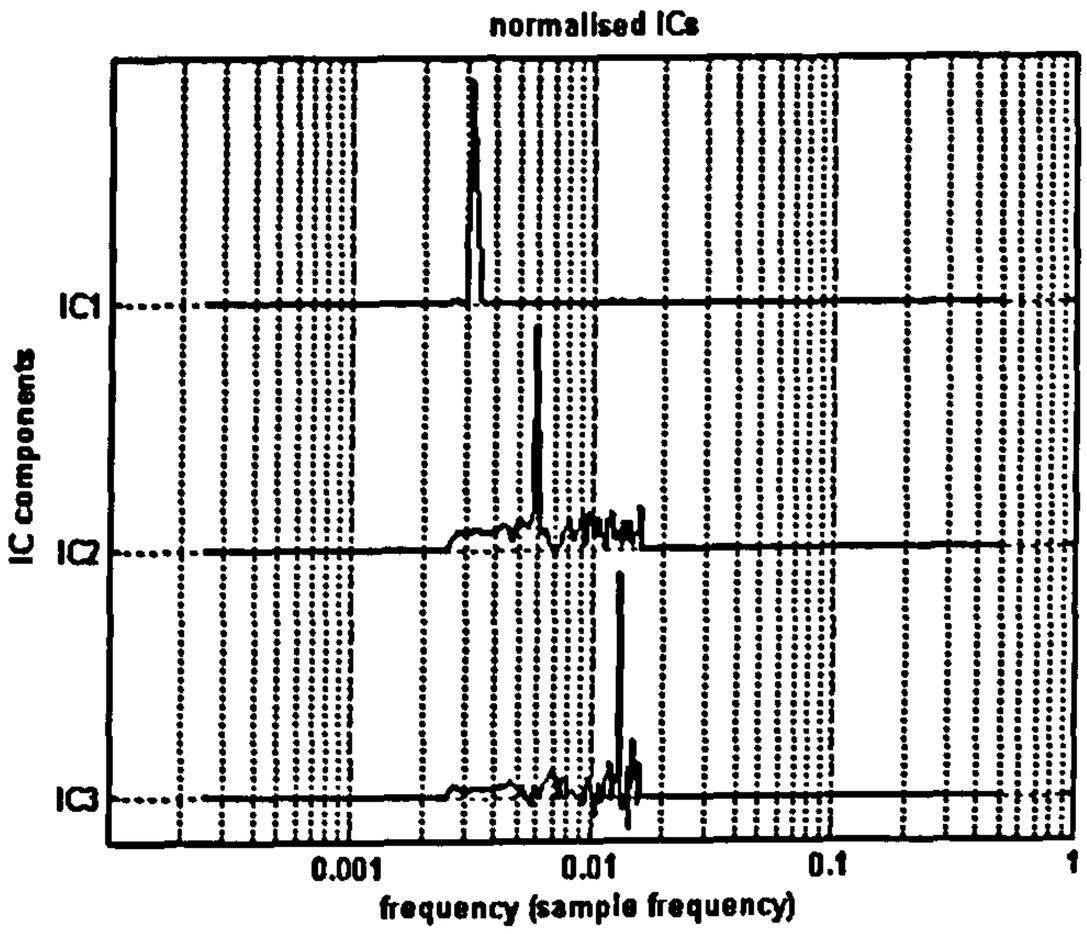


Figure 5-6: The dominant independent components for case study 1

Tag Number	IC1 (*SI*)	IC2 (*SI*)
5	0.54	0
7	0.57	0
8	0.55	0
10	0.03	0
13	0.76	0
19	0.78	0
22	**1.0**	0
23	0.80	0
25	0.70	0

Table 5-6: Values of significance indices for case study 1

5.3.3 Case study 2 (High-Frequency Case)

Narrow-band spectral ICA extracted the dominant ICs from the 12 oscillating measurements. Shown in Figure 5-7, IC1 and IC2 correspond to the fundamental and second harmonics respectively. Their values of significance indices are listed in Table 5-7. Since this is a high-frequency case, the significance indices are summed together to give an indication of the harmonic content. It can be seen that Tag34 has the largest number, which means that it is the nearest loop to the root cause. It differs from the result from the nonlinear time series analysis method which points to Tag 33 because here the weights of the fundamental and second harmonics are summed together whilst the nonlinear time series analysis tools evaluate all the harmonic contents.

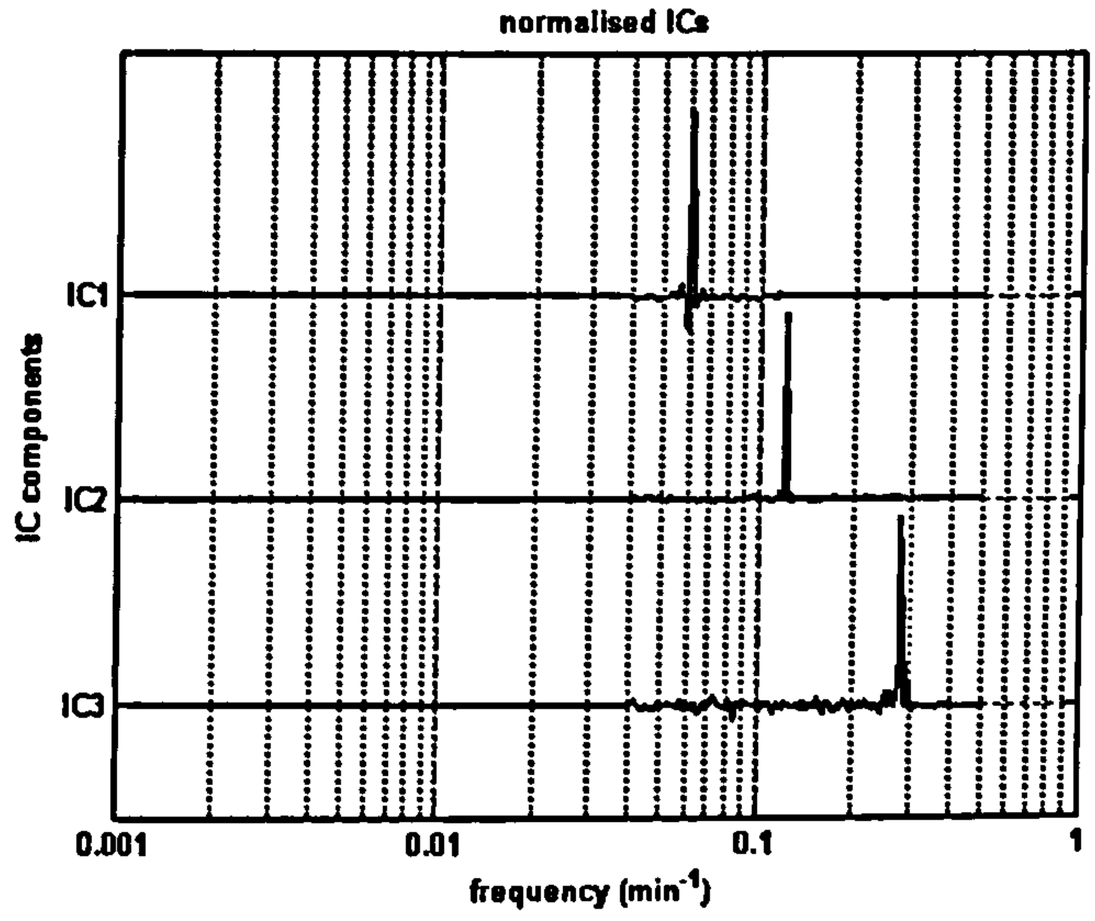


Figure 5-7: The dominant independent components of case study 2

Tag Number	IC1 (*SI*)	IC2 (*SI*)	IC1+IC2 (*SI*)
34	0.71	0.83	**1.54**
13	0.61	0.32	0.93
33	0.49	1.00	**1.49**
2	0.97	0.03	1.00
10	1.00	0	1.00
11	0.65	0.08	0.73
20	0.89	0.04	0.93
24	0.38	0.04	0.42
3	0.86	0	0.86
25	0.35	0.17	0.52
4	0.68	0	0.68
19	0.49	0.11	0.60

Table 5-7: Values of significance indices for case study 2

5.3.4 Summary

The revised version of spectral ICA uses the same idea as the present version, e.g. the significance indices and dominant ratios, but only has the different interpretation in low frequency and high frequency cases. Although it is as simple as summing up the significance indices of a few dominant ICs in high frequency cases, it is very convenient and practicable and has expanded the application of present spectral ICA approach.

Chapter 6

Bi-spectrum Based Non-linearity Detection

Being based on higher order statistics, the bispectrum also provides a means to detect nonlinearity when oscillatory measurement records exist in process systems. The detection approach that is proposed here can discriminate between bad tuning induced and valve non-linearity induced oscillations, which are the most common reasons for limit cycles in process systems.

6.1 Introduction

The approach is based on the premise that the harmonic content of an oscillation caused by a valve non-linearity is likely to contain a significant first harmonic plus even and odd harmonics, whereas that caused by bad tuning will have little harmonic content. This is in agreement with Thornhill & Hagglund (1997) who have analysed power spectra to distinguish between limit cycle oscillations caused by non-linearities, and linear oscillations caused by, for instance, bad tuning. Their approach is based on whether the power spectrum has multiple peaks or a single peak. However, it is susceptible to noise in the time trend which can cause spurious peaks in the power spectrum plots. Bispectra are arguably superior for this type of discrimination, because white noise (Gaussian) has a bispectrum plot that is theoretically uniformly zero(Nikias & Petropulu 1993), and hence the impact of noise should be reduced.

6.2 Nonlinearity Detection Based On The Bispectrum

As has been described in Chapter 4, the bispectrum, $B(f_1, f_2)$, of a time series, $x(n)$, of finite length, N: $1 \leq n \leq N$, is defined by:

$$B(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)], \qquad (6.1)$$

where X denotes the Fourier Transform of x. If the signal $x(n)$ contains excessive noise, its bispectrum can be estimated by applying the Discrete Fourier Transform directly and then by segment averaging and frequency-domain smoothing; if it is deterministic, the bispectrum becomes (Nikias & Petropulu 1993):

$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \qquad (6.2)$$

which means that no segmentation or smoothing is needed.

Suppose that data pertaining to either a controlled variable or a controller output is composed of a mixture of harmonics and sinusoids:

$$x(n) = \sum_{i=0}^{N_s-1} \alpha_i \cos(2\pi f_i n + \phi_i) + \sum_{i=0}^{N_s-1} \sum_{k=0}^{N_q-1} \alpha_{ik}^q \cos(2\pi f_{ik}^q n + \phi_{ik}^q) \qquad (6.3)$$

where N_s is the number of sinusoids, f_i is the fundamental frequency, ϕ_i the phase angle, α_i is the magnitude of the i^{th} sinusoid, N_q is the number of harmonics, f_{ik}^q is the k^{th} harmonic frequency corresponding to the i^{th} sinusoid, α_{ik}^q the magnitude of the harmonic and ϕ_{ik}^q the phase angle. The frequencies are all distinct.

Since the frequencies are all distinct, the Fourier transform of $x(n)$ will only be nonzero at the frequencies of each of the fundamental sinusoids and their

harmonics. By definition (Equation 6.2) the bispectrum can only be nonzero if the three terms $X(f_1)$, $X(f_2)$ and $X^*(f_1+f_2)$ are all nonzero and hence it can be nonzero only if the signal has sinusoids or harmonics at frequencies: f_1, f_2, f_1+f_2. This observation can now be applied to certain specific signals.

- A signal containing a single sinusoid of frequency f_0: $X(f)$ will be zero with the exception of $X(f_0)$, so $X^*(f_0+f_0)$ will be zero, thus the bispectrum will be uniformly zero.
- A signal containing a single fundamental of frequency f_0 plus solely odd harmonics: $X^*(f_0+f_2)=0$ because the sum of any two odd harmonic frequencies will result in a frequency corresponding to an even harmonic, and hence the bispectrum will be uniformly zero.
- A signal containing a single fundamental of frequency f_0 plus even harmonics: the bispectrum will definitely have nonzero elements because, for instance $X(f_0), X(f_0)$ and $X^*(f_0+f_0)$ are all non-zero.
- A signal containing a single fundamental of frequency f_0 plus both odd and even harmonics: the even harmonics will now contribute to form nonzero elements of the bispectrum.
- A signal containing three or more fundamental sinusoids and solely their odd harmonics: is likely to have only zero elements in the bispectrum, unless an unusual situation arises where frequencies from two different sinusoidal series add up to a frequency pertaining to one of the harmonics of the third sinusoidal series.

If a process oscillation is due to bad tuning, it should oscillate at predominantly one resonant frequency. Thus its bispectrum should match the first case above; in other words, there should be a zero bispectrum plot. However, non-linearity induced oscillations introduce even harmonics to the system, so the magnitude bispectrum plot of the process variables should have some peaks. Therefore, plotting the bispectrum of an oscillating measurement record should give an indication of whether the oscillations have been caused by bad tuning or by valve nonlinearity.

6.3 Illustrative Examples

6.3.1 Simulated Data

Three test cases are first described that are based on the simple, single input single output (SISO) system given in Figure 6-1. The loop was excited by simultaneously applying a set point change at time zero and introducing a Gaussian noise disturbance of 0.01% variance. The second order process is given by the following transfer function:

$$\frac{2}{(3s+1)(10s+1)}e^{-3s} \tag{6.4}$$

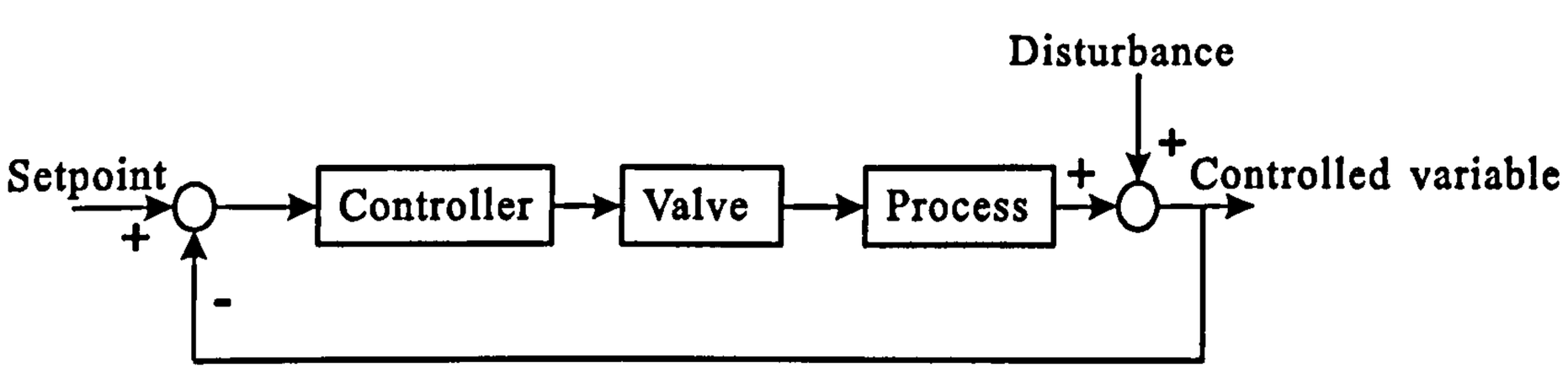


Figure 6-1: Block diagram of a simple SISO process

The controller has PI action with a gain Kc=0.5 and integral action $\tau_I = 12.3$. The simulation was performed for 10000 seconds and the sample time was 1 second. To avoid any initial transient response, only the 8192 steady state data points from 1000 to 9191 of the controlled variable were used. The FFT length of 256 was used to calculate the bispectrum, the number of samples per record was set to 256. A window length of 21 (J=10) was chosen to ensure an acceptable variance.

The normal condition

Data pertaining to normal operation was generated from a simulation that contained a linear valve model. Figure 6-2 shows the results obtained when the simulation output was analysed: the time series, power spectrum, magnitude bispectrum (non-redundant area) plot of the controlled variable (PV).

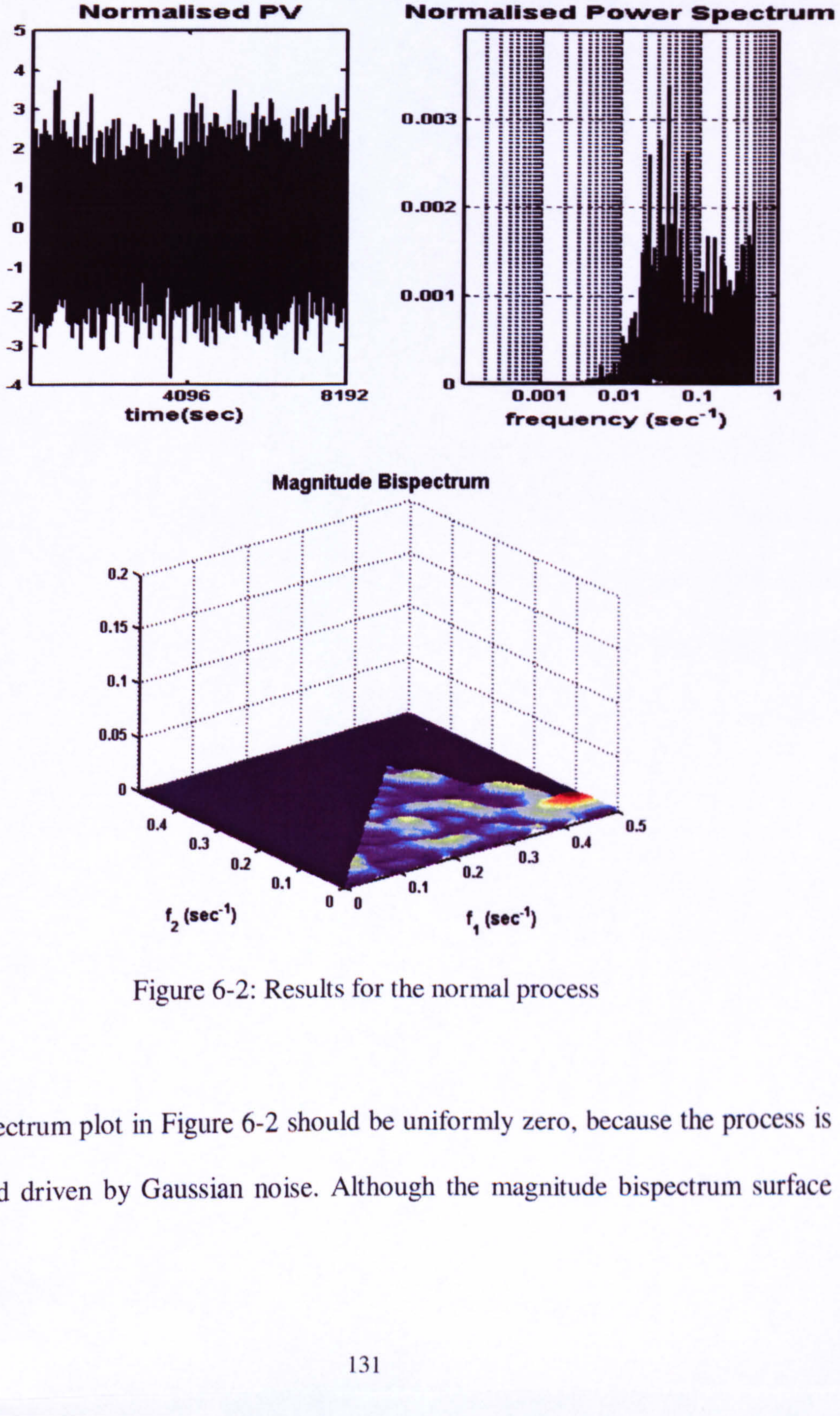


Figure 6-2: Results for the normal process

The bispectrum plot in Figure 6-2 should be uniformly zero, because the process is linear and driven by Gaussian noise. Although the magnitude bispectrum surface

appears to be flat, there are some small peaks due to the variance of the bispectrum estimate. The highest peak is 0.01 and so a detection threshold is set to be slightly larger at 0.02. Note that this threshold would be different if a smoothing window length of other than 21 was chosen.

Bad tuning

The same model was used but now the controller gain was increased until the system oscillated. Figure 6-3 shows the results obtained when the simulation output was analysed: the time series, power spectrum, magnitude bispectrum (non-redundant area) plot of the controlled variable.

The power spectrum plot shows a high peak at frequency 0.047 containing 90% of the total energy that indicates the system is oscillating at the resonant frequency. The bispectrum plot is flat. This is consistent with the prior analysis that a zero bispectrum plot would be expected for a signal containing a single frequency component. Note that the highest magnitude of the other peaks was 0.0007, which was well below the threshold of uniformly zero plot (i.e. 0.02) and hence all the other peaks were neglected.

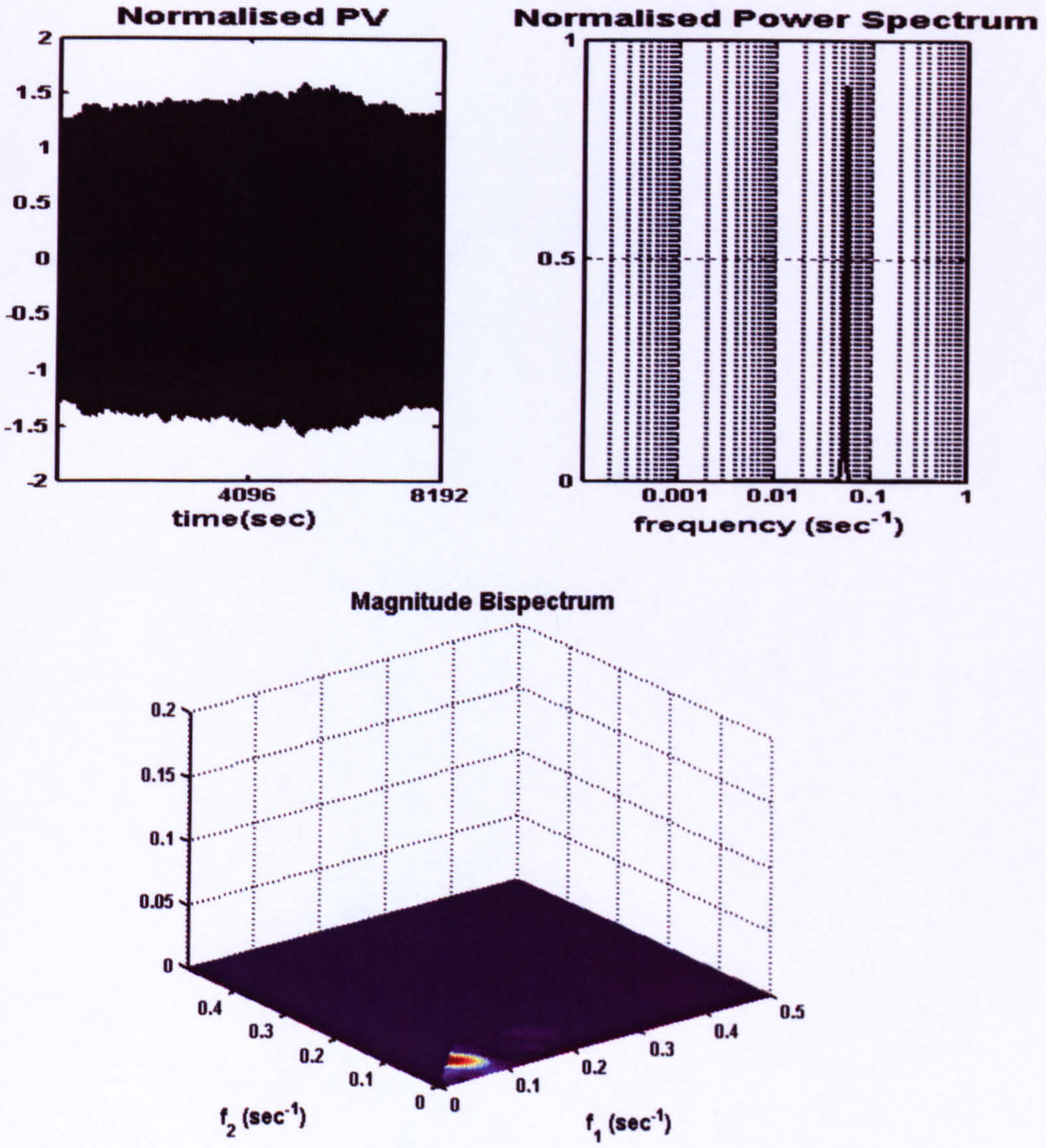


Figure 6-3: Results for bad tuning

Valve non-linearity

The effect of incorporating valve stiction into the valve component model of the simulation was examined (Figure 6-4).

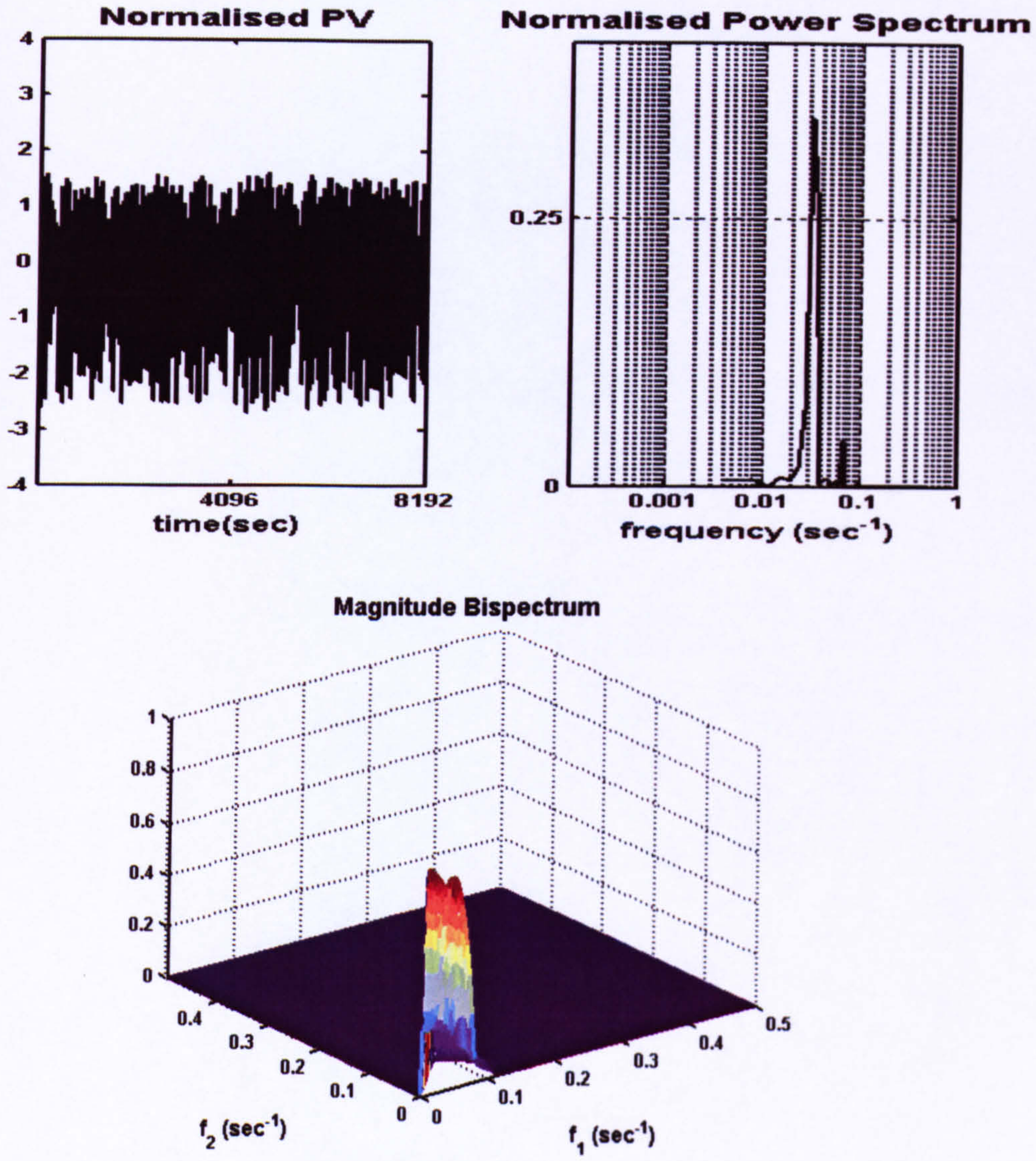


Figure 6-4: Results for valve stiction

Note that there is a peak of 0.82 in the non-redundant area of the magnitude bispectrum surface plot at bifrequency (0.031, 0.031). Also the power spectrum plot has a peak at frequency 0.031. This means that the signal oscillates at the fundamental frequency and should have a harmonic at a frequency of 2*0.031. Note the presence of this harmonic in the power spectrum plot and also its small magnitude. This supports the view that this approach can detect valve stiction, because valve stiction results in even harmonics in the process signal.

6.3.2 Industrial Data

The process variable of the loop pertaining to Tag 22 was analysed again (see Case Study 1, Section 4.3.5). Previous work (Thornhill *et al.* 2002b; Thornhill *et al.* 2003b) had detected the presence of an oscillation and a sticking valve was found in the loop with Tag 22. Figure 6-5 shows the results obtained for Tag 22: time series, power spectrum, magnitude bispectrum (non-redundant area) plot.

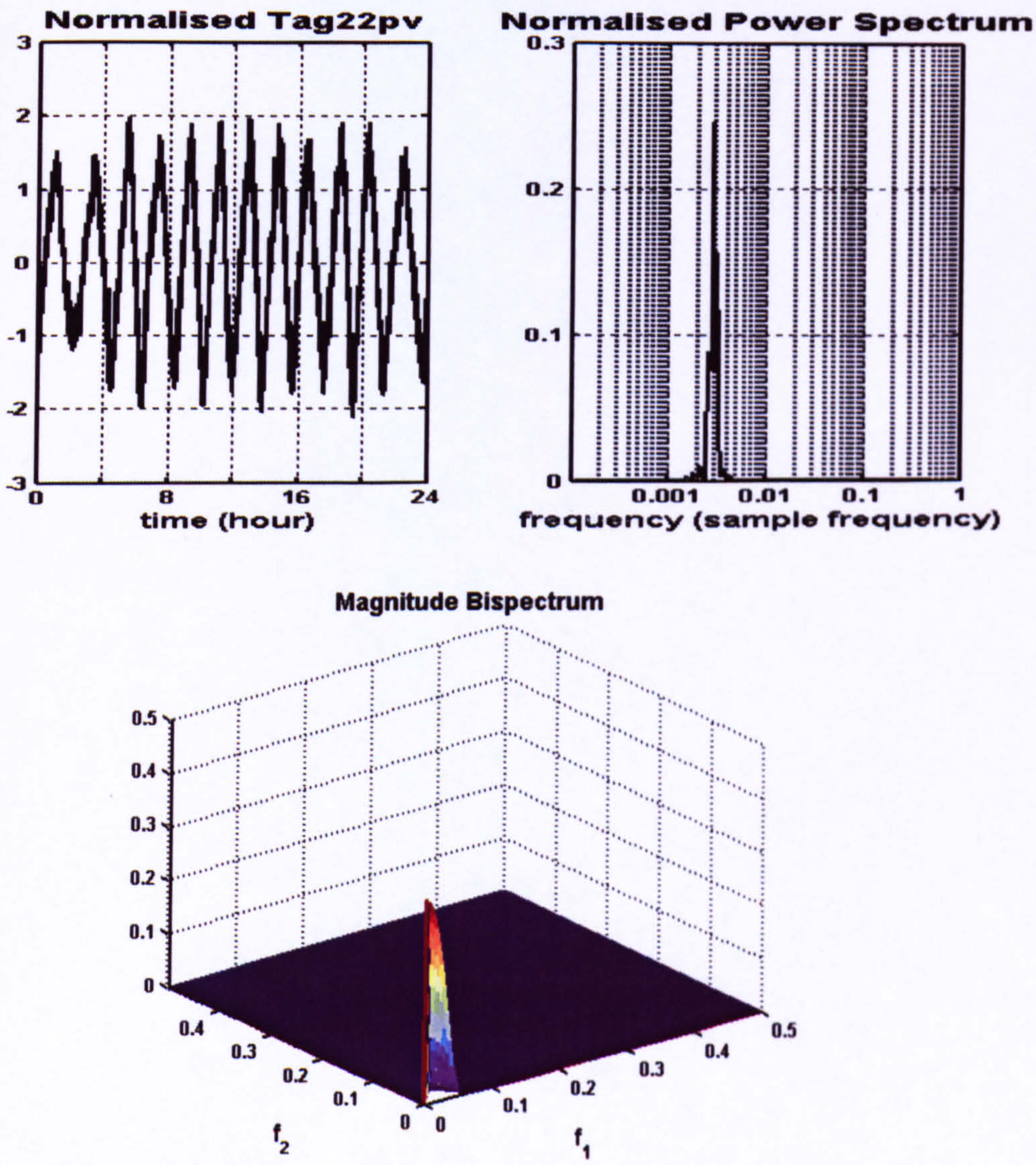


Figure 6-5: Results for the process variable of Tag22

The power spectrum plot shows that the signal has a fundamental frequency of 0.0029. The peak of the bispectrum plot is at the corresponding bi-frequency (0.0029, 0.0029). Although the fundamental of the power spectrum is relatively wide, the second and third harmonic contents support the view that the signal contains even harmonics. Therefore it can be concluded that a valve problem is more likely than bad-tuning to have caused such an oscillation.

6.4 Comparison With Previous Non-linearity Detection Methods

As has been reviewed in Chapter 2, Choudhury *et al.*(2004) have made use of higher order statistics to detect and diagnose poor control-loop performance caused by actuator nonlinearities. Their work concentrates on *bicoherence* based indices for detecting and quantifying non-Gaussianity and nonlinearity that may be present in regulated systems. The bicoherence method itself can't locate the source of a system non-linearity; it is also a nonlinearity detection method and therefore is no more powerful than the bispectrum method proposed in this chapter. A comparison between the bicoherence-based method and the bispectrum-based method is made below.

- Confidence thresholds

The bicoherence-based method obtains the NGI and NLI indices on the basis that the squared bicoherence at each frequency is a chi-squared (χ^2) distributed variable with 2 degrees of freedom. The estimated bicoherence is compared to the critical value calculated from the central χ^2 distribution table for a significance level α. In practice for α=0.05, an NGI value of less than 0.001 was assumed to be zero, and an NLI value less than 0.01 was assumed to be zero(Shoukat Choudhury *et al.* 2004).

The bispectrum-based method assumes that a signal is linear if its maximum magnitude bispectrum is less than 0.02 when 21 smoothing window is applied.

- Estimate sensitivity

  Bicoherence is a normalized bispectrum. By normalizing, the bicoherence is sensitive to division by a small number, as this can artificially increase calculated values. It is sometimes difficult to specify at which bifrequency the quadratic phase coupling happens because if strong coupling exists, the bicoherence may be nearly unity at all of the bifrequency channels.

  Magnitude bispectrum, however, can display explicit peaks at frequencies where phase coupling happens. A big smoothing window can be applied to alleviate the dependence of the bispectrum estimate on the second order spectral properties.

- Calculation efficiency

  The bispectrum estimate is more time efficient than the bicoherence because it doesn't involve normalisation. Furthermore the bispectrum-based method, unlike the bicoherence-based method, doesn't have to look up the χ^2 distribution table in order to estimate the Gaussianity of the signal.

Chapter 7

Conclusions And Recommendations

7.1 Conclusions

Although there has been considerable commercial and academic interest in methods for process monitoring and fault isolation (Chapter 2), they are usually based on a perception of the inherent process dynamics of a plant. Methods for the isolation of the root cause of an oscillation that are based on harmonic attenuation include the D- factor and Non-linearity index. In Spectral ICA the independent components align with harmonics also. Although reliant on the frequency response characteristics of harmonic propagation, all make assumptions about the attenuation of a plant, thus restricting their application. A major contribution of this PhD research is its detailed harmonic propagation analysis (Chapter 3) which leads to a model that enables these techniques to be applied to oscillations of any frequency.

A new technique, which draws on this model more explicitly, is a bispectral analysis method, which has the ability to extract the harmonic content of oscillating signals (Chapter 4). A bi-amplitude ratio index provides a measure of the power ratio of the fundamental to the third harmonic and a procedure is proposed to localize the source loop of propagated oscillations based on this measure and on the propagation model developed. Unlike the non-linearity test approach it can deal with Gaussian noise efficiently. Chapter 5 considers nonlinear time series tools as alternative measures that can be input into the same procedure, where revision to

the Spectral ICA method is also proposed. Bispectral analysis as a nonlinearity detection tool is discussed in Chapter 6 and its comparison with other nonlinearity detection methods performed. This bispectrum based nonlinearity detection method is found to be more computationally efficient and more visibly reflects the harmonic coupling in nonlinear signals.

The various root cause isolation methods, the biamplitude ratio index (Chapter 4), the correlation dimension, the maximal Lyapunov exponent (Chapter 5), the enhanced spectral ICA (Chapter 5), and the non-linearity index (as proposed by Thornhill *et al.* (2003b) and discussed in Chapter 2), will be compared in the next sub-section. The comparison is based on their different performances on a simulated case and two sets of industrial data. The results show that appropriate use of the techniques developed in this work overcomes the problems (*Criticisms 9*, *10* and *11* in Chapter2) in other techniques.

7.2 Comparison

- *Time domain versus frequency domain based method*

The correlation dimension, maximal Lyapunov exponent and Non-linearity test are time domain based methods, whilst the bi-amplitude ratio index and spectral ICA are in the frequency domain.

- *Applicability to low or high frequency cases*

The correlation dimension, the maximal Lyapunov exponent and the bi-amplitude ratio index are all measures that, when augmented by knowledge of the harmonic propagation, are applicable in both cases. The spectral ICA method proposed by (Xia and Howell 2003a) is only applicable for low frequency cases, because it assumed that the source would contain the strongest fundamental harmonic (the unity significance index to the fundamental IC).The enhanced version of this method, as presented in Chapter 5, shows that it can be applied to high frequency cases by summing the significance indices to the harmonic ICs together. As proposed by Thornhill *et al.* (2003b) the non-linearity index method is based on the premise that process dynamics filter non-linearity during oscillation propagation. So it is intended to be applied to high-frequency cases where the transfer functions, of the controllers plus processes, filter harmonics. Table 7-1 shows the Non-linearity indices (N's) for the three examples presented in the previous chapters. It can be seen that the N test points to Loop2 for the simulated case (low-frequency oscillation) and to Tag23 for the low-frequency industrial case, neither of which are the true root causes. Both the bi-amplitude ratio index (Section 4.3.5) and Spectral ICA (Section 5.3) located the right root causes, i.e. Loop 1 for the simulation and

Tag22 for the low-frequency industrial case. For the high-frequency industrial case, however, the N test points to Tag34, which agrees with both the bi-amplitude ratio index (Section 4.3.5) and Spectral ICA analysis (Section 5.3).

Simulated example	Loop1	Loop2
N	0.27	1.89

Case study 1 (low frequency)	Tag22	Tag13	Tag23	Tag5	Tag25	Tag19
N	1.29	0.88	1.44	0.53	0.86	0.84

Case study 2 (high frequency)	34	13	33	2	10	11	20	24	3	25	4	19
N	9.37	3.43	1.17	1.83	0.67	0.81	0.79	0.40	0.27	0.54	0.30	0.61

Table 7-1: The Non-linearity indices for the examples

- *Effect of noise*

The process variables recorded from process plants inevitably contain noise. The isolation methods should ameliorate the effect of noise as much as possible. Theoretically the bispectrum of a Gaussian noise is zero (Nikias & Patropulu 1993), so the bi-amplitude ratio index is less affected by noise. However the estimation variance of the bispectrum of noise depends on the choice of segment length, FFT length and frequency-domain smoothing window length. Comparable results for different loops were obtained by choosing the same parameters for all of the oscillating signals of interest. Spectral ICA extracts independent non-Gaussian sources by maximizing the kurtosis of the time series, which is relative to both the

fourth-order central moment and the 2nd-order central moment (Xia 2003), so the noise effect is minimized. The correlation dimension and maximal Lyapunov exponent methods do not give meaningful values when time series are corrupted by much noise, because noisy components hide the determinism or predictability of the time series. The non-linearity test inherently quantifies the signal-to-noise ratio of the measurement records because it compares the predictability of the time records with the corresponding surrogates. A larger-than-unity *N* value indicates non-linearity (Thornhill *et al.* 2003b). Since well-established oscillations or oscillations near the root cause usually have a large signal-to-noise ratio, the source loop or the loops near the source usually have an *N* value larger than unity.

- *Isolation of multiple oscillations*

Multi-range spectral ICA has been used to isolate multiple oscillations (Xia 2003) since the extracted independent components pertain to different fundamental frequencies. The spectral ICA method first partitions the power spectra of the time series into different frequency ranges with frequency-domain filtering and then applies the spectral ICA method on each range. Independent components at different frequency channels are extracted and the sources to different oscillations are then isolated. However the bi-amplitude ratio index has an inherent capability of detecting and diagnosing multiple-source oscillations without frequency-domain filtering. The time-domain based methods, i.e., the correlation dimension, maximal Lyapunov exponent and the Non-linearity index cannot be used to isolate multiple oscillations.

- *Which is better?*

From the application of these five methods to the three examples, all methods worked well for the low frequency simulated case and industrial case except the Non-linearity test. The inapplicability of the *N* test for the low-frequency case has been analysed. For the high frequency industrial case, the results differ slightly from each other: Correlation Dimension and maximal Lypunov exponent methods isolated Tag33 (Section 5.2.5) whilst the other methods pointed to Tag34. However, Tags 34, 33 and 13 are all highlighted by all the methods. In this kind of case, process understanding would be of benefit. The bi-amplitude ratio index and revised Spectral ICA reached the same diagnosis in all the examples because they both inherently extract harmonic distribution of an oscillatory signal. They are also applicable to multiple oscillations. In low frequency cases both give reasonable diagnoses by looking for the minimum bi-amplitude index or maximum significance index of the fundamental IC. However in the case of low pass attenuation the bispectral method has a verification procedure with the fifth bi-amplitude ratio whilst the Spectral ICA method may not extract the independent component of up to the fifth harmonic. For this reason the Spectral ICA is less applicable than the bispectral method.

7.3 Practical Steps for The Isolation of Oscillations

Based on the analysis in this chapter a diagnostic procedure is recommended here. It is first important to detect whether the oscillations emanate from one or more sources as this will affect the choice of methods that can be applied.

1) Identify those loops associated with the same oscillating source (i.e. with a common fundamental frequency). Thornhill *et al* (2003a), Xia (2003) have proposed practical methods to detect multiple oscillations.

2) Assess whether the common fundamental, plus associated second and third harmonics, are likely to be on the positive or negative slope of each of the relevant log-ratio plots.

3) Apply the various methods. Although the bispectral method is better in most cases, the other isolating methods also give meaningful results. They can be performed to verify the results from the bispectral method. For example, the Non-linearity test and nonlinear time series tools will eliminate the loops with low signal-to-noise ratios; the Spectral ICA verifies the harmonic distribution identified by the bi-amplitude ratios.

4) Apply process understanding, if possible, to confirm the identified root cause; see for instance Thornhill *et al.* (2002b) and Thornhill *et al.* (2003b) who have integrated process understanding with data-driven analysis.

References

Åström, K. J., (1991). Assessment of achievable performance of simple feedback loops, *Int.J.Adap.Control and Sig.Proc.*, vol. 5, pp. 3-19.

Bittanti, S., M. Campi, and S. M. Savaresi, (1997). Unbiased Estimation of a Sinusoid in Colored Noise via Adapted Notch Filters, *Automatica*, vol. 33, no. 2, pp. 209-215.

Boltezar, M. and J. K. Hammond, (1999). Experimental study of the vibrational behaviour of a coupled non-linear mechanical system, *Mechanical Systems and Signal Processing*, vol. 13, no. 3, pp. 375-394.

Boltezar, M., N. Jaksic, I. Simonovski, and A. Kuhelj, (1999). Dynamical behaviour of the planar non-linear mechanical system -- Part II: Experiment, *Journal of Sound and Vibration*, vol. 226, no. 5, pp. 941-953.

Burioka, N., G. Cornelissen, F. Halberg, and D. T. Kaplan, (2001). Relationship between correlation dimension and indices of linear analysis in both respiratory movement and electroencephalogram, *Clinical Neurophysiology*, vol. 112, no. 7, pp. 1147-1153.

Cao, L., (1997). Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43-50.

Chen, D. and D. E. Seborg, (2003). Design of decentralized PI control systems based on Nyquist stability analysis, *Journal of Process Control*, vol. 13, no. 1, pp. 27-39.

Choudhury, M. A. A. S., N. F. Thornhill, and S. L. Shah, (2004). A data-driven model for valve stiction, IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM), Hong Kong.

Collis, W. B., P. R. White, and J. K. Hammond, (1998). Higher-Order Spectra: The Bispectrum and Trispectrum, *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 375-394.

Computing & Control Engineering, (2005). Special section: Control loop assessment and diagnosis, vol. 16, no. 4, pp.18-46.

Cusumano, J. P. and F. C. Moon, (1995). Chaotic non-planar vibrations of the thin elastica: Part I: Experimental observation of planar instability, *Journal of Sound and Vibration*, vol. 179, no. 2, pp. 185-208.

Deibert, R., (1994). Model based fault detection of valves in flow control loops, In Preprints of the IFAC Symposium on fault detection, supervision and safety for technical processes: Safeprocess 94, Espoo, Finland, pp. 445-450.

Desborough, L. and T. J. Harris, (1992). Performance assessment measures for univariate feedback control, *Can.J.Chem.Eng.*, vol. 70, pp. 1186-1197.

Desborough, L., P. Nordh, and R. Miller, (2001). Process out of control, *Industrial computing*, August, pp. 52-55.

Dobson, J. P. and N. F. Thornhill, (2002). Analysis of plant-wide disturbances using conditional entropy, In: Proceedings 2002 IEEE International Conference On Control Applications, Glasgow, Scotland, pp. 687-688.

Ender, D. B., (1993). Process control performance: not as good as you think, *Control Engineering*, vol. 40, no. 10, pp. 180-190.

EnTech, (1998). Control valve dynamic specification (version 3.0). Retrieved from http://www.entechcontrol.com/.

Eugene Parker, J. R., H. A. Ware, D. P. Wipf, W. R. Tompkins, B. R. Clark, E. C. Larson, and H. Vincent Poor, (2000). Fault diagnostics using statistical change detection in the bispectral domain, *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 561-570.

Fackrell, J. W. A., (1996). *Bispectral analysis of the speech signals*, PhD thesis, The University of Edinburgh, Edinburgh.

Fackrell, J. W. A., P. R. White, J. K. Hammond, R. J. Pinnington, and A. T. Parsons, (1995). The interpretation of the bispectra of vibration signals--: I. Theory, *Mechanical Systems and Signal Processing*, vol. 9, no. 3, pp. 257-266.

Forsman, K. and A. Stratin, (1999). A new criterion for detecting oscillations in control loops, Europeon Control Conference, CP8-3, Karlsruhe, Germany.

Fraser, A. M. and H. L. Swinney, (1986). Independent coordinates for strange attractors from mutual information, *Phys.Rev.A*, vol. 33, no. 2, pp. 1134-1140.

Gerry, J. and M. Ruel, (2001). How to measure and combat valve stiction on line, In: Proceedings of ISA 2001, Instrumentation, Systems and Automated Society, Houston, TX.

Grassberger, P. and I. Procaccia, (1983a). Characterisation of strange attractors, *Physical Review Letters*, vol. 50, pp. 346-349.

Grassberger, P. and I. Procaccia, (1983b). Measuring the strangeness of strange attractors, *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189-208.

Guo, Q., G. Yue, T. Suda, and J. Sato, (2003). Flow characteristics in a bubbling fluidized bed at elevated temperature, *Chemical Engineering and Processing*, vol. 42, no. 6, pp. 439-447.

Hägglund, T., (1995). A control-loop performance monitor, *Control Engineering Practice*, vol. 3, no. 11, pp. 1543-1551.

Hägglund, T., (2002). A friction compensator for pneumatic control valves, *Journal of Process Control*, vol. 12, no. 8, pp. 897-904.

Harris, T. J. and C. T. Seppala, (2002). Recent developments in controller performance monitoring and assessment techniques, J.B.Rawlings, Babatunde Ogunnaike, & John Eaton, eds., Tuscon, Arizona, A.I.ChE Symposium Series: vol. 98, pp. 208-222.

Hay, J. M., B. H. Nelson, C. L. Briens, and M. A. Bergougnou, (1995). The calculation of the characteristics of a chaotic attractor in a gas-solid fluidized bed, *Chemical Engineering Science*, vol. 50, no. 3, pp. 373-380.

Hecox, K., S. Nayak, K. Gin, A. McGee, and W. van Drongelen, (2003). Linear and non-linear measures of the human neonatal EEG, *Neurocomputing*, vol. 52-54, pp. 779-786.

Hegger, R., H. Kantz, and T. Schreiber, (1999). Practical implementation of nonlinear time series methods: the TISEAN package, *Chaos*, vol. 9, pp. 413-435.

Hinich, M. J. and M. A. Wolinsky, (1988). A test for aliasing using bispectral analysis, *Journal of the American Statistical Association*, vol. 83, pp. 499-502.

Horch A., (2000). *Condition monitoring of control loops*, Royal Institute of Technology, Stockholm, Sweden.

Horch, A., (1999). A simple method for detection of stiction in control valves, *Control Engineering Practice*, vol. 7, no. 10, pp. 1221-1231.

Horch, A. and A. J. Isaksson, (1998). A method for detection of stiction in control valves, In: Proceedings of the IFAC workshop on on-line fault detection and supervision in the chemical process industry, Session 4B, Lyon, France.

Horch, A. and A. J. Isaksson, (1999). A modified index for control performance assessment, *Journal of Process Control*, vol. 9, no. 6, pp. 475-483.

Huang, H. P., J. C. Jeng, C. H. Chiang, and W. Pan, (2003). A direct method for multi-loop PI/PID controller design, *Journal of Process Control*, vol. 13, no. 8, pp. 769-786.

Jaksic, N., M. Boltezar, I. Simonovski, and A. Kuhelj, (1999). Dynamical behaviour of the planar non-linear mechanical system -- Part I: Theoretical modelling, *Journal of Sound and Vibration*, vol. 226, no. 5, pp. 923-940.

Kano, M., H. Maruta, H. Kugemoto, and K. Shimizu, (2004). Practical model and detection algorithm for valve stiction, S. L. Shah & J. F. MacGregor, eds., In: Proceedings of the Seventh IFAC-DYCOPS Symposium, Cambridge, Massachusetts, USA.

Kantz, H. and T. Schreiber, (1997). *Nonlinear time series analysis*, Cambridge University Press, Cambridge.

Kozub, D. J., (1997). Controller performance monitoring and diagnosis, In: Proceedings of Fifth International Conference on Chemical Process Control, CACHE Corp, pp. 83-96.

Lee, J., W. Cho, and T. F. Edgar, (1998). Multiloop PI controller tuning for interacting multivariable processes, *Computers & Chemical Engineering*, vol. 22, no. 11, pp. 1711-1723.

Lei, M., Z. Wang, and Z. Feng, (2001). Detecting nonlinearity of action surface EMG signal, *Physics Letters A*, vol. 290, no. 5-6, pp. 297-303.

Li, Z., Z. Wu, Y. He, and C. Fulei, (2005). Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating

machinery, *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 329-339.

Martin, G. D., L. E. Turpin, and R. P. Cline, (1991). Estimating control function benefits, *Hydrocarbon Processing* (June) pp. 68-73.

McMillan, G. K., (1995). Improve control valve response, *Chem.Eng.Prog.* (June) pp. 77-84.

Miao, T. and D. Seborg, (1999). Automatic detection of excessively oscillatory feedback control loops, In: Proceedings of the 1999 IEEE International Conference On Control Applications, Hawaii, USA.

Muller, V., W. Lutzenberger, H. Preißl, F. Pulvermuller, and N. Birbaumer, (2003). Complexity of visual stimuli and non-linear EEG dynamics in humans, *Cognitive Brain Research*, vol. 16, no. 1, pp. 104-110.

Nikias, C. L. and A. P. Petropulu, (1993). *Higher-Order Spectra: A nonlinear signal processing frame work.* Prentice Hall, New Jersey.

Ogawa, S., (1998). A data analysis and graphical presentation system for control loop performance assessment, In TAPPI Process Control Conference 98, Vancouver, Canada.

Olsson, H. and K. J. Astrom, (2001). Friction generated limit cycles, *IEEE Transactions on Control Systems Technology*, vol. 9, no. 4, pp. 629-636.

Olsson, H., K. J. Astrom, C. C. d. Wit, M. Gafvert, and P. Lischinsky, (1998). Friction model and friction compensation, *European J.Contr.*, vol. 3, pp. 176-195.

Paulonis, M. A. and J. W. Cox, (2003). A practical approach for large-scale controller performance assessment, diagnosis, and improvement, *Journal of Process Control*, vol. 13, no. 2, pp. 155-168.

Prokoph, A. and J. Veizer, (1999). Trends, cycles and nonstationarities in isotope signals of phanerozoic seawater, *Chemical Geology*, vol. 161, no. 1-3, pp. 225-240.

Qin, S. J., (1998). Control performance monitoring -- a review and assessment, *Computers & Chemical Engineering*, vol. 23, no. 2, pp. 173-186.

Rengaswamy, R., T. Hägglund, and V. Venkatasubramanian, (2001). A qualitative shape analysis formalism for monitoring control loop performance, *Engineering Applications of Artificial Intelligence*, vol. 14, no. 1, pp. 23-33.

Rossi, M. and C. Scali, (2005). A comparison of techniques for automatic detection of stiction: simulation and application to industrial data, *Journal of Process Control*, vol. 15, no. 5, pp. 505-514.

Ruel, M., (2000). Stiction: The hidden menace, *Control Magazine*, vol. 13, no. 11.

Ruel, M. (2002), Valve diagnosis identifies process problems, In: Proceedings of ISA Western Regional Conf. and Exhibition, Las Vegas, NV.

Sauer, T. and J. Yorke, (1993). How many delay coordinates do you need?, *Int.J.Bifurcation and Chaos*, vol. 3, p. 737.

Seborg, D. E., T. F. Edgar, and D. A. Mellichamp, (1989). *Process Dynamics and Control*. John Wiley & Sons, Inc., New York.

Shinskey, F. G., (1990). How good are our controllers in absolute performance and robustness? *Measurement and Control*, vol. 23, pp. 114-121.

Shinskey, F. G., (2000). The three faces of control valves, *Control Engineering*, vol. 47, no. 7, p. 83.

Shoukat Choudhury, M. A. A., S. Shah, and N. Thornhill, (2004). Diagnosis of poor control-loop performance using higher-order statistics, *Automatica*, vol. 40, no. 10, pp. 1719-1728.

Shoukat Choudhury, M. A. A., N. F. Thornhill, and S. L. Shah, (2005). Modelling valve stiction, *Control Engineering Practice*, vol. 13, no. 5, pp. 641-658.

Singhal, A. and T. I. Salsbury, (2005). A simple method for detecting valve stiction in oscillating control loops, *Journal of Process Control*, vol. 15, no. 4, pp. 371-382.

Stenman, A., F. Gustafsson, and K. Forsman, (2003). A segmentation based approach for detection of stiction in control valves, *Int.J.Adap.Control and Sig.Proc.*, vol. 17, no. 7-9, pp. 625-634.

Stephanopoulos, G., (1984). *Chemical process control: an introduction to theory and practice*. Prentice-Hall International Editions.

Taha, O., G. A. Dumont, and M. S. Davies, (1996). Detection and diagnosis of oscillations in control loops, In: Preceedings of the 35th IEEE conference on Decision and Control, Kobe, Japan, pp. 2432-2437.

Tangirala, A. K., S. L. Shah, and N. F. Thornhill, (2005). PSCMAP: A new tool for plant-wide oscillation detection, *Journal of Process Control*, vol. 15, no. 8, pp. 931-941.

Thornhill, N. F. and T. Hägglund, (1997). Detection and diagnosis of oscillation in control loops, *Control Engineering Practice*, vol. 5, no. 10, pp. 1343-1354.

Thornhill, N. F., B. Huang, and H. Zhang, (2003a). Detection of multiple oscillations in control loops, *Journal of Process Control*, vol. 13, no. 1, pp. 91-100.

Thornhill, N. F., S. L. Shah, and B. Huang, (2001). Detection of distributed oscillations and root-cause diagnosis, In: Proceedings of the CHEMFAS-4, Cheju, Korea, pp. 167-172.

Thornhill, N. F., S. L. Shah, B. Huang, and A. Vishnubhotla, (2002a). Spectral principal component analysis of dynamic process data, *Control Engineering Practice*, vol. 10, no. 8, pp. 833-846.

Thornhill, N. F., C. Xia, J. Howell, J. Cox, and M. Paulonis, (2002b). Analysis of plant-wide disturbances through data-driven techniques and process understanding, In: Proceedings of the IFAC World Congress, Barcelona, Spain.

Thornhill, N. F., J. W. Cox, and M. A. Paulonis, (2003b). Diagnosis of plant-wide oscillation through data-driven analysis and process understanding, *Control Engineering Practice*, vol. 11, no. 12, pp. 1481-1490.

Thornhill, N. F., M. A. A. Shoukat Choudhury, and S. L. Shah, (2004). The impact of compression on data-driven process analyses, *Journal of Process Control,* vol. 14, no. 4, pp. 389-398.

Trendafilova, I. and H. Van Brussel, (2001). Non-linear dynamics tools for the motion analysis and condition monitoring of robot joints, *Mechanical Systems and Signal Processing*, vol. 15, no. 6, pp. 1141-1164.

Tugnait, J. K. and Y. Zhou, (2000). On closed-loop system identification using polyspectral analysis given noisy input-output time-domain data, *Automatica*, vol. 36, no. 12, pp. 1795-1808.

Wallén, A., (1997). Valve diagnositics and automatic tuning, In: Proceedings of 1997 American Control Conference, Albuquerque, New Mexico, pp. 2930-2934.

Wang, J., T. Chen, and B. Huang, (2004). Closed-loop identification via output fast sampling, *Journal of Process Control*, vol. 14, no. 5, pp. 555-570.

Wang, S. F., R. Mosdorf, and M. Shoji, (2003). Nonlinear analysis on fluctuation feature of two-phase flow through a T-junction, *International Journal of Heat and Mass Transfer*, vol. 46, no. 9, pp. 1519-1528.

Wang, W. J. and R. M. Lin, (2003). The application of pseudo-phase portrait in machine condition monitoring, *Journal of Sound and Vibration*, vol. 259, no. 1, pp. 1-16.

Wilton, S. R., (2000). Control valves and process variability, *ISA Transactions*, vol. 39, no. 2, pp. 265-271.

Xia, C. and J. Howell, (2003a). Isolating multiple sources of plant-wide oscillations via independent component analysis, In: Proceedings of the 5th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (Safeprocess'03), Washington DC, USA.

Xia, C., (2003). *Control loop measurement based isolation of faults and disturbances in process plants,* PhD thesis, The University of Glasgow, UK.

Xia, C. and J. Howell, (2003b). Loop status monitoring and fault localisation, *Journal of Process Control*, vol. 13, no. 7, pp. 679-691.

Xia, C. and J. Howell, (2005). Isolating multiple sources of plant-wide oscillations via independent component analysis, *Control Engineering Practice*, vol. 13, no. 8, pp. 1027-1035.

Zang, X. and J. Howell, (2003). Discrimination between bad tuning and non-linearity induced oscillations through bispectral analysis, In: Proceedings of the 2003 SICE Annual Conference, Fukui, Japan.

Zang, X. and J. Howell, (2004a). Comparison of methods to identify the root cause of plant-wide oscillations, In: Proceedings of the 2004 IEEE CCA/ISIC/CACSD, Taipei, Taiwan.

Zang, X. and J. Howell, (2004b). Correlation dimension and Lyapunov exponents based isolation of plant-wide oscillations, S. L. Shah & J. F. MacGregor, eds. In: Proceedings of the Seventh IFAC-DYCOPS Symposium, Cambridge, Massachusetts, USA.

Zang, X. and J. Howell, (2005a). Isolating the root cause of propagated oscillations in process plants, *International Journal of Adaptive Control and Signal Processing*, vol. 19, no. 4, pp. 247-265.

Zang, X. and J. Howell, (2005b). Isolating the source of whole plant oscillations through biamplitude ratio analysis, *Submitted to Control Engineering Practice.*

Zang, X. and J. Howell, (2005c). The propagation of whole plant oscillations through a chemical process plant, *Submitted to Journal of Process Control.*

Zhang, Y., Q. G. Wang, and K. J. Astrom, (2002). Dominant pole placement for multi-loop control systems, *Automatica*, vol. 38, no. 7, pp. 1213-1220.

Zhu, Z. K., Z. H. Feng, and F. R. Kong, (2005). Cyclostationarity analysis for gearbox condition monitoring: Approaches and effectiveness, *Mechanical Systems and Signal Processing*, vol. 19, no. 3, pp. 467-482.

Appendix A

Fourier Series Analysis of The Output of A Valve Stiction Model

The output of the data-driven stiction model with a sine wave input with angular frequency of 1 rad·s^{-1} and period 2π is (Choudhury *et al.* 2004):

$$y(t)=\begin{cases} k\left(X_m \sin(t)-\dfrac{s-j}{2}\right) & 0\le t\le \dfrac{\pi}{2} \\ k\left(X_m-\dfrac{s-j}{2}\right) & \dfrac{\pi}{2}\le t\le \pi-\phi \\ k\left(X_m \sin(t)+\dfrac{s-j}{2}\right) & \pi-\phi\le t\le \dfrac{3\pi}{2} \\ k\left(-X_m+\dfrac{s-j}{2}\right) & \dfrac{3\pi}{2}\le t\le 2\pi-\phi \\ k\left(X_m \sin(t)-\dfrac{s-j}{2}\right) & 2\pi-\phi\le t\le 2\pi \end{cases} \qquad \text{(A.1)}$$

where X_m is the amplitude of the input sine wave, s is the stick band, j is the slip-jump, $\phi=\sin^{-1}\left(\dfrac{X_m-s}{X_m}\right)$ and k is the slope of the input-output characteristic in the moving phase (k=1 is assumed for a valve). Then the complex Fourier series expansion is:

$$\frac{1}{2\pi}\int_{t=0}^{2\pi} y(t)e^{-int}dt,\quad n=0,1,2.... \qquad \text{(A.2)}$$

where, after substitution of $\sin(t)=\dfrac{1}{2i}\left(e^{it}-e^{-it}\right)$:

$$\int_{t=0}^{2\pi} y(t)e^{-int}dt = \int_{t=0}^{\pi/2} k\left(\frac{X_m}{2i}\left(e^{it}-e^{-it}\right)-\frac{s-j}{2}\right)e^{-int}dt + \int_{t=\pi/2}^{\pi-\phi} k\left(X_m-\frac{s-j}{2}\right)e^{-int}dt$$
$$+\int_{t=\pi-\phi}^{3\pi/2} k\left(\frac{X_m}{2i}\left(e^{it}-e^{-it}\right)+\frac{s-j}{2}\right)e^{-int}dt + \int_{t=3\pi/2}^{2\pi-\phi} k\left(-X_m+\frac{s-j}{2}\right)e^{-int}dt$$
$$+\int_{t=2\pi-\phi}^{2\pi} k\left(\frac{X_m}{2i}\left(e^{it}-e^{-it}\right)-\frac{s-j}{2}\right)e^{-int}dt \qquad \text{(A.3)}$$

Writing it compactly:

$$\int_{t=0}^{2\pi} y(t)e^{-int}dt = T_1+T_2+T_3+T_4+T_5 \qquad \text{(A.4)}$$

where $T_1 = \int_{t=0}^{\pi/2} k\left(\frac{X_m}{2i}\left(e^{it}-e^{-it}\right)-\frac{s-j}{2}\right)e^{-int}dt$, and so on.

When n=0, $\frac{1}{2\pi}\int_{t=0}^{2\pi} y(t)dt = 0$; when n≠0, evaluation term by term gives:

$$T_1 = k\left(\frac{X_m}{1-n^2}+\frac{X_m n}{1-n^2}\cos\frac{1+n}{2}\pi+\frac{s-j}{2n}\cos\frac{1+n}{2}\pi\right)$$
$$-ik\left(\frac{X_m n}{1-n^2}\sin\frac{1+n}{2}\pi+\frac{s-j}{2n}\sin\frac{1+n}{2}\pi-\frac{s-j}{2n}\right)$$

$$T_2 = \frac{k}{n}\left(X_m-\frac{s-j}{2}\right)\left(\sin n(\pi-\phi)-\sin\frac{n\pi}{2}\right)+\frac{ik}{n}\left(X_m-\frac{s-j}{2}\right)\left(\cos n(\pi-\phi)-\cos\frac{n\pi}{2}\right)$$

$$T_3 = k\frac{(X_m + X_m n)\cos\big((1+n)(\pi-\phi)+2\phi\big)+(X_m - X_m n)\cos\big((1+n)(\pi-\phi)\big)-2X_m n\cos\left(\frac{n-1}{2}\pi\right)}{2\left(1-n^2\right)}$$

$$+k\frac{(s-j)\left(\sin\big((1+n)(\pi-\phi)+\phi\big)-\sin\left(\frac{n}{2}\pi\right)\right)}{2n}$$

$$-ik\frac{(X_m + X_m n)\sin\big((1+n)(\pi-\phi)+2\phi\big)+(X_m - X_m n)\sin\big((1+n)(\pi-\phi)\big)+2X_m n\sin\left(\frac{n-1}{2}\pi\right)}{2\left(1-n^2\right)}$$

$$+ik\frac{(s-j)\left(\cos\big((1+n)(\pi-\phi)+\phi\big)+\cos\left(\frac{n}{2}\pi\right)\right)}{2n}$$

$$T_4 = \frac{k}{n}\left(X_m - \frac{s-j}{2}\right)\left(-\sin\frac{n\pi}{2}+\sin n\phi\right)+\frac{ik}{n}\left(X_m - \frac{s-j}{2}\right)\left(\cos\frac{n\pi}{2}-\cos n\phi\right)$$

$$T_5 = k\frac{(X_m + X_m n)\cos\big((n-1)\phi\big)+(X_m - X_m n)\cos\big((n+1)\phi\big)-2X_m}{2\left(1-n^2\right)}-k\frac{(s-j)\sin(n\phi)}{2n}$$

$$+ik\frac{(X_m + X_m n)\sin\big((n-1)\phi\big)+(X_m - X_m n)\sin\big((n+1)\phi\big)}{2\left(1-n^2\right)}+ik\frac{(s-j)\big(\cos(n\phi)-1\big)}{2n}$$

Collect terms and evaluate the even and odd harmonics separately:

When n=2, 4, …, $\frac{1}{2\pi}\int_{t=0}^{2\pi} y(t)e^{-int}dt = 0$, so the even harmonics are zero.

When n=1,3, …,

$$\frac{1}{2\pi}\int_{t=0}^{2\pi} y(t)e^{-int}dt = \frac{k}{\pi}\left(\frac{X_m\left(\sin n\phi-\sin\frac{n\pi}{2}\right)}{n\left(1-n^2\right)}-\frac{s\sin n\phi}{n\left(1-n^2\right)}+\frac{j\sin n\phi}{n}+\frac{\cos n\phi\sqrt{2X_m s-s^2}}{1-n^2}\right)$$

$$+\frac{ki}{\pi}\frac{\cos n\phi\left(-X_m+s-j\right)}{n}$$

, so odd harmonics are present when a sinusoidal input is added to the valve stiction model.

Let j=0, the model becomes a dead-band model and the output to a sine wave has no even harmonics.

Appendix B

Harmonic Content of Triangular Waves

Without generality a triangular wave with amplitude of 1, period of 2π is given in Figure B-1, where, $-\frac{\pi}{2} < a < \frac{\pi}{2}$.

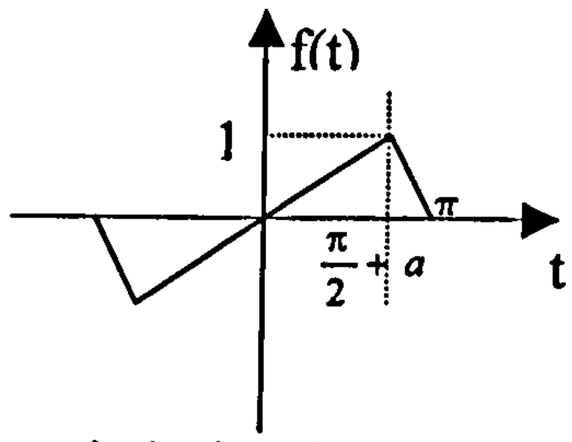


Figure B-1: A period of a triangular wave

$$f(t)=\begin{cases} -\dfrac{1}{\frac{\pi}{2}-a}t-\dfrac{\pi}{\frac{\pi}{2}-a} & -\pi \le t \le -\left(\dfrac{\pi}{2}+a\right) \\ \dfrac{1}{\frac{\pi}{2}+a}t & -\left(\dfrac{\pi}{2}+a\right) \le t \le \dfrac{\pi}{2}+a \\ -\dfrac{1}{\frac{\pi}{2}-a}t+\dfrac{\pi}{\frac{\pi}{2}-a} & \dfrac{\pi}{2}+a \le t \le \pi \end{cases} \qquad \text{(B.1)}$$

The Fourier series expansion is:

$$f(t)=\frac{1}{2}a_0+\sum_{n=1}^{\infty}a_n\cos nt+\sum_{n=1}^{\infty}b_n\sin nt, \qquad \text{(B.2)}$$

where, $a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)dt=0$, $a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\cos ntdt=0$, and

$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\sin ntdt=\left(\frac{4}{n^2(\pi+2a)}+\frac{4}{n^2(\pi-2a)}\right)\sin\left(n\left(a+\frac{\pi}{2}\right)\right).$$

- if a=0, $b_n = \frac{8}{n^2\pi}\sin\frac{n\pi}{2}$, so $f(t)$ contains only odd harmonics since only when n=1,3,5,..., $b_n \neq 0$;

- if $a \neq 0$, although there are n-1 values of a i.e. $a = \frac{1,\cdots n-1}{n}\pi - \frac{\pi}{2}, n \geq 2$ for b_n=0, b_n is most likely to be nonzero for both even and odd n's.

Appendix C

Oscillation Propagation of A 4×4 Plant

The plant is given by:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{C.1}$$

Suppose Loop1 is the source loop, i.e. there is a nonlinear valve in Loop1, the other 3 loops have the following linear relationship between controller outputs and controlled variables:

$$u_2 = -C_2 y_2,\ u_3 = -C_3 y_3,\ u_4 = -C_4 y_4 \tag{C.2}$$

Equation (C.1) could be rearranged to:

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} &= \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 + \begin{bmatrix} P_{12} & P_{13} & P_{14} \\ P_{22} & P_{23} & P_{24} \\ P_{32} & P_{33} & P_{34} \\ P_{42} & P_{43} & P_{44} \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix} \\ &= \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 + \begin{bmatrix} P_{12} & P_{13} & P_{14} \\ P_{22} & P_{23} & P_{24} \\ P_{32} & P_{33} & P_{34} \\ P_{42} & P_{43} & P_{44} \end{bmatrix} \begin{bmatrix} -C_2 & 0 & 0 \\ 0 & -C_3 & 0 \\ 0 & 0 & -C_4 \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix} \\ &= \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 - \begin{bmatrix} 0 & C_2P_{12} & C_3P_{13} & C_4P_{14} \\ 0 & C_2P_{22} & C_3P_{23} & C_4P_{24} \\ 0 & C_2P_{32} & C_3P_{33} & C_4P_{34} \\ 0 & C_2P_{42} & C_3P_{43} & C_4P_{44} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \end{aligned} \tag{C.3}$$

$$\text{Then} \begin{bmatrix} 1 & C_2P_{12} & C_3P_{13} & C_4P_{14} \\ 0 & 1+C_2P_{22} & C_3P_{23} & C_4P_{24} \\ 0 & C_2P_{32} & 1+C_3P_{33} & C_4P_{34} \\ 0 & C_2P_{42} & C_3P_{43} & 1+C_4P_{44} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 . \tag{C.4}$$

Therefore,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & C_2P_{12} & C_3P_{13} & C_4P_{14} \\ 0 & 1+C_2P_{22} & C_3P_{23} & C_4P_{24} \\ 0 & C_2P_{32} & 1+C_3P_{33} & C_4P_{34} \\ 0 & C_2P_{42} & C_3P_{43} & 1+C_4P_{44} \end{bmatrix}^{-1} \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 = M^{-1} \begin{bmatrix} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \end{bmatrix} u_1 \qquad \text{(C.5)}$$

Because most practical processes are open-loop stable (Chen & Seborg 2003), it is assumed that the system is open-loop stable. Then the closed-loop system has column diagonal dominance for all loops at all frequencies in order to make sure that the closed-loop system is stable (Chen & Seborg 2003). The column diagonal dominance is expressed as

$$\left|1+P_{ll}(j\omega)C_l(j\omega)\right| > \sum_{k=1,k\neq l}^{n} \left|P_{kl}(j\omega)C_l(j\omega)\right| \quad \forall l,\omega.$$

Therefore in (C.5), M is a column diagonal dominance matrix, whose inverse matrix M^{-1} can be approximated as diag{1, $\frac{1}{1+C_2P_{22}}$, $\frac{1}{1+C_3P_{33}}$, $\frac{1}{1+C_4P_{44}}$}. The ratio r(ω) of the magnitudes of the sinusoidal components of the source y_1 to target loops y_2, y_3, y_4 at any frequency ω is then:

$$r(\omega) = \frac{|y_k|}{|y_s|}\Big|_\omega \approx \left| \frac{P_{ks}(j\omega)}{P_{ss}(j\omega)\left[1+C_k(j\omega)P_{kk}(j\omega)\right]} \right| \qquad \text{(C.6)}$$

where s=1, k=2, 3, 4.

Equation (C.6) is the same as that derived from a 2×2 plant (Equation (3.9)). The resonant frequency of r(ω) is also determined dominantly by the disturbance transfer function $\frac{1}{1+C_kP_{kk}}$ as analysed before.

Appendix D

Relationship of Maximal Lyapunov Exponent and Harmonic Content

Without loss of generality a signal with fundamental frequency of $1/2\pi$ is considered. To calculate the maximal Lyapunov exponent, only a-period time range is enough.

1. Suppose that the signal containing a fundamental harmonic is described by:

$$x(t) = \cos t, t = 0, \cdots, 2\pi$$

This time signal can be projected into a two-dimensional state space:

$$X(t) = \begin{cases} x(t) = \cos t \\ \dot{x}(t) = -\sin t \end{cases}, t = 0, \cdots 2\pi$$

Select a point in the state space $X(t_0) = \{x(t_0), \dot{x}(t_0)\}$, and select all neighbours with distance smaller than ε, denoted as $X(t_i) = \{x(t_i), \dot{x}(t_i)\}, i = 1, ..., n$, where n is the number of the neighbours. Let $t_0 - t_i = \delta$, then the distance between the reference point and any of the neighbours is:

$$D(t_0, \delta) = \sqrt{(x(t_0) - x(t_0 - \delta))^2 + (\dot{x}(t_0) - \dot{x}(t_0 - \delta))^2} = \sqrt{2 - 2\cos\delta}$$

It can be seen that the distance is independent of t. So function $S(\Delta t)$ is a constant which is independent of Δt. Therefore the maximal Lyapunov exponent is $\lambda_1 = 0$

2. Suppose that the signal containing a fundamental and second harmonic is described by:

$x(t)=\cos t+\alpha\cos 2t, t=0,\cdots,2\pi$, where α is the amplitude of the second harmonic.

This time signal can be projected into a three-dimensional state space:

$$X(t)=\begin{cases} x(t)=\cos t+\alpha\cos 2t \\ \dot{x}(t)=-\sin t-2\alpha\sin 2t\ , t=0,\cdots 2\pi \\ \ddot{x}(t)=-\cos t-4\alpha\cos 2t \end{cases}$$

In order to obtain the maximal Lypunov exponent, the following things have to be done.

Select a point in the state space $X(t_0)=\{x(t_0),\dot{x}(t_0),\ddot{x}(t_0)\}$, and select all neighbours with distance smaller than ε, denoted as $X(t_i)=\{x(t_i),\dot{x}(t_i),\ddot{x}(t_i)\}, i=1,...,n$, where n is the number of the neighbours. Let $t_0-t_i=\delta_i^{t_0}$, then the distance between the reference point and any of the neighbours is:

$$D(t_0,\delta_i^{t_0})=\sqrt{(x(t_0)-x(t_0-\delta_i^{t_0}))^2+(\dot{x}(t_0)-\dot{x}(t_0-\delta_i^{t_0}))^2+(\ddot{x}(t_0)-\ddot{x}(t_0-\delta_i^{t_0}))^2} \quad \text{(D.1)}$$

Since the distance is smaller than ε, the time range between the reference point and the neighbours $\delta_i^{t_0}$ should be very small so that $\cos\delta_i^{t_0}\simeq 1$ and $\sin\delta_i^{t_0}\simeq\delta_i^{t_0}$. Then Equation (D.1) can be simplified as

$$D(t_0,\delta_i^{t_0})\simeq\left|\delta_i^{t_0}\right|\sqrt{\sin^2 t_0+12\alpha\sin t_0\sin 2t_0+52\alpha^2\sin^2 2t_0}\ ,$$

So the average distance of all neighbours to the reference point $X(t_0)$ is

$$D(t_0)=\frac{\sum_{i=1}^{n}\delta_i^{t_0}}{n}\sqrt{\sin^2 t_0+12\alpha\sin t_0\sin 2t_0+52\alpha^2\sin^2 2t_0}$$

The average distance at time $t_0+\Delta t$ is then

$$D(t_0+\Delta t)=\frac{\sum_{i=1}^{n}\delta_i^{t_0}}{n}\sqrt{\sin^2(t_0+\Delta t)+12\alpha(\sin t_0+\Delta t)\sin(2t_0+2\Delta t)+52\alpha^2\sin^2(2t_0+2\Delta t)}$$

.

Repeat this for many values of t and obtain

$$D(t+\Delta t)=\frac{\sum_{i=1}^{n}\delta_i^{t}}{n}\sqrt{\sin^2(t+\Delta t)+12\alpha\sin(t+\Delta t)\sin(2t+2\Delta t)+52\alpha^2\sin^2(2t+2\Delta t)},t=0,\cdots 2\pi$$

.

Since the average distances for different t's should be all smaller than ε, one of the logarithm of the average distances can represent the average over all the t's, for example at t=0:

$$S(\Delta t)=E(\ln D(t+\Delta t))=\ln D(0+\Delta t)=\ln(\frac{\sum_{i=1}^{n}\delta_i^{0}}{n}\sqrt{\sin^2\Delta t+12\alpha\sin\Delta t\sin 2\Delta t+52\alpha^2\sin^2 2\Delta t}+D_0)$$

,

where D_0 is the initial distance, i.e. the distance at $\Delta t=0$.

Since the maximal Lyapunov exponent happens at $\Delta t\ll 1$, where $\sin\Delta t\simeq\Delta t$ and

$\sin 2\Delta t\simeq 2\Delta t$, $S(\Delta t)\simeq\ln(\frac{\sum_{i=1}^{n}\delta_i^{0}}{n}\sqrt{\Delta t^2+24\alpha\Delta t^2+208\alpha^2\Delta t^2}+D_0)=\ln(C\Delta t+D_0)$,

where $C=\frac{\sum_{i=1}^{n}\delta_i^{0}}{n}\sqrt{1+24\alpha+208\alpha^2}$. And at $\Delta t\ll 1$,

$C\Delta t+D_0=D_0(\frac{C}{D_0}\Delta t+1)\simeq D_0e^{\frac{C}{D_0}\Delta t}$. Therefore,

$S(\Delta t)=\ln D_0+\frac{C}{D_0}\Delta t=\ln D_0+C_1\sqrt{1+24\alpha+208\alpha^2}\Delta t$, where $C_1=\frac{\sum_{i=1}^{n}\delta_i^{0}}{nD_0}$.

So $\lambda_1=\frac{S(\Delta t)}{\Delta t}=C_1\sqrt{1+24\alpha+208\alpha^2}$.

3. Suppose that a signal containing a fundamental, second and third harmonic is described by:

$x(t)=\cos t+\alpha\cos 2t+\beta\cos 3t, t=0,\cdots,2\pi$, where β is the amplitude of the third harmonic.

This time series can be projected into a three-dimensional state space:

$$X(t)=\begin{cases} x(t)=\cos t+\alpha\cos 2t+\beta\cos 3t \\ \dot{x}(t)=-\sin t-2\alpha\sin 2t-3\beta\sin 3t \\ \ddot{x}(t)=-\cos t-4\alpha\cos 2t-9\beta\cos 3t \end{cases}, t=0,\cdots 2\pi$$

Repeat the same procedures to calculate $S(\Delta t)$ in order to obtain the maximal Lypunov exponent as in part 2, we have

$$D(t+\Delta t)=$$

$$\frac{\sum_{i=1}^{n}\delta_i^t}{n}\sqrt{\begin{array}{l}\sin^2(t+\Delta t)+52\alpha^2\sin^2(2t+2\Delta t)+657\beta^2\sin^2(3t+3\Delta t)+12\alpha\sin(t+\Delta t)\sin(2t+2\Delta t)\\ +42\beta\sin(t+\Delta t)\sin(3t+3\Delta t)+372\alpha\beta\sin(2t+2\Delta t)\sin(3t+3\Delta t)\end{array}}$$

and

$$S(\Delta t)=\ln(\frac{\sum_{i=1}^{n}\delta_i^0}{n}\sqrt{\begin{array}{l}\sin^2\Delta t+12\alpha\sin\Delta t\sin 2\Delta t+52\alpha^2\sin^2 2\Delta t+657\beta^2\sin^2 3\Delta t\\ +42\beta\sin\Delta t\sin 3\Delta t+372\alpha\beta\sin 2\Delta t\sin 3\Delta t\end{array}}+D_0)$$

$$\simeq \ln D_0+C_1\sqrt{1+24\alpha+208\alpha^2+5913\beta^2+126\beta+2232\alpha\beta}\,\Delta t$$

,

where D_0 is still the initial distance, and $C_1=\frac{\sum_{i=1}^{n}\delta_i^0}{nD_0}$.

Therefore $\lambda_1=C_1\sqrt{1+24\alpha+208\alpha^2+5913\beta^2+126\beta+2232\alpha\beta}$