Closed Loop Control Performance Monitoring

Melvin Jones

CBC 890

Closed Loop Performance Monitoring

Melvin Jones

9706060

University of Pretoria

Department of Chemical Engineering

CBC 890: Dissertation: Control

17 September 2005

Synopsis

Proportional, Integral and Derivative feedback control (PID) is a mature technology responsible for the majority of automated decision making in the process industry. Despite the high reliance on this technology, low levels of maintenance and performance measurement are the norm in the process industry.

Several analysis techniques exist for identifying oscillation, and then highlighting the root cause of the problem. Several time and frequency domain statistical techniques, as well as wavelet analysis are used to diagnose loop performance. In this study, 127 different control loops are analysed, and in depth troubleshooting is performed on a selection of 18 different control loops.

The performance of flow loop F1035 is tracked through a number of different analysis techniques, highlighting the pitfalls of using only a single analysis technique. Lower order statistics and minimum variance performance analysis show that the loop is performing well. Plotting the PV-OP relationship suggests non-linear tendencies on F1035, and this is corroborated using high order statistical analysis (bicoherence). Non-linear loop behaviour is often as a result of a slip stick cycle, a sign that valve maintenance may be required. Frequency (power spectrum) analysis shows a 43 minute dominant oscillation, suggesting a low frequency disturbance affecting loop performance.

Process units are typically exposed to cyclic behaviour occurring at several different frequencies, each having a different effect on the control of the process. By using a frequency based approach based on sinusoidal basis functions (ie Fourier analysis), these different frequencies get aggregated. This smudging of specific frequency information makes it difficult to pin-point the root cause, and makes the grouping of common oscillations difficult. In order to address the above issue, F1035 is analysed using othornormal wavelet basis functions. The results show that the period of oscillation is affected between day and night, with roughly a 2 minute oscillation prevalent at mid night, compared to a 100 minute

oscillation at mid day. Obviously the 12 hour day-night swing is also prevalent. This information is unique to this approach. Ways of visualising changes in oscillatory behaviour using the wavelet analysis are also presented.

Technical analysis of controller performance is only a small subsection of the issues that need to be considered when implementing a loop monitoring and maintenance solution. Issues such as connectivity, configuration, analysis, reporting and auditing are key in designing a workable maintenance environment for PID loop maintenance.

Several packages are available commercially to assist industry in performing loop maintenance. When evaluating which package is best suited to a specific requirement, it is important to consider several different issues. The different audiences with a vested interest in loop performance require special attention in terms of reporting requirements. Visualisation of results is often more important than the physical measure of performance. Finally, the ability of a company to benchmark itself against current best practices and performance is often perceived as a major advantage.

The results presented and discussed were generated using real industrial data. Information regarding suggested best practice when evaluating commercially available products is based largely on the author's personal experience in the large scale industrial installation of such a monitoring solution.

SYNOPSIS 3

NOMENCLATURE LIST 8

LIST OF FIGURES 11

1 INTRODUCTION 13

2 CLOSED LOOP PERFORMANCE MONITORING 15

2.1 Introduction 15

2.2 Minimum-variance 16

2.3 Modified minimum variance 18
2.3.1 Requirement 18
2.3.2 Closed-loop pole modification 19

2.4 Auto correlation 20

2.5 Filtering and Correlation analysis 21

2.6 Nonlinear techniques 22
2.6.1 Slip-stick cycling 22
2.6.2 Normality index - histograms 24
2.6.3 Cross correlation 25

2.7 Statistics 28
2.7.1 Lower order statistics 28
2.7.2 Higher order statistics (HOS) 29

2.8 Frequency analysis 30
2.8.1 Power spectrum 30
2.8.2 Bicoherence 30
2.8.3 Common oscillation detection 31

2.9 Input – Output pattern recognition 32

2.10 PID-achievable performance 33

2.11 Open loop benchmark 33

2.12 Interventions 34

3 WAVELET ANALYSIS 35

3.1 Fourier Analysis 35
3.1.1 Windowed Fourier transform (WFT) 37
3.1.2 Uncertainty 39

3.2 Why the Wavelet (Mathematically) 40

3.3 Why the Wavelet? (Motivational) 41

3.4 What is a Wavelet? 42

3.5 Approximations and Details 44

4 METHODOLOGY 47

5 RESULTS 49

5.1 Lower order statistics 49

5.2 Minimum variance 51

5.3 PV-OP pattern recognition 55

5.4 Histograms and normality 61

5.5 Cross Correlation 64

5.6 Frequency analysis 70
5.6.1 Autocorrelation analysis 70
5.6.2 Power spectrum analysis 74
5.6.3 Common oscillation detection 77
5.6.4 Bicoherence 80
5.6.5 Wavelets 81

6 DISCUSSION 89

6.1 Results 89

6.2 Industrial requirements 90

6.2.1 Connectivity 92
6.2.2 Configuration 94
6.2.3 Analysis 95
6.2.4 Reporting 96
6.2.5 Auditing and work flow structure 98

7 CONCLUSIONS 100

8 REFERENCES 105

Nomenclature List

δ - impulse signal

H_T – Hilbert space

a - wavelet scaling factor

a(t) – white noise signal

ARMA – autoregressive moving average

b – wavelet translation factor

C – complex numbers

C_i – i^{th} order cumulant

CLP – closed loop potential

c_n – fourier transform coefficients

d – process input/disturbance variable

$D(z^{-1})$ – discrete disturbance transfer function

$\delta(t)$ – impulse function

$\delta(w)$ – fourier transform of $\delta(t)$

E – expectation

e(t) – process error

e_n – Fourier modes

F(t) – fourier transform of t

f(t) – period function

$\hat{f}_T$ (w) – fourier transform of *f*

$\hat{f}_T$ (w, t) – windowed fourier transform of *f*

$\tilde{f}$ (w) – wavelet transform of *f*

g(u) – windowing function

j – resolution counter

$K(z^{-1})$ – discrete PID controller transfer function

K_c – controller gain

K_p – process gain

M_i – i^{th} order moment

MSE – mean square error

MV – minimum variance

$P(z^{-1})$ – discrete process transfer function

PCA – principle component analysis

$p_e(k)$ – first order decay pattern

PID – proportional, integral, derivative

P_j – approximation operator

Q_j – detail operator

R – real numbers

r_{uy} – cross correlation function of input/output

T_i – controller integral time

T_s – sampling interval (s)

u – process input/manipulated variable

v(t) – unmeasured disturbance

W_i- normalised eigen vector

w_n – frequency function

X(f) – fast fourier transform of f

y_{sp} – process setpoint/desired value

y_t – process output

Z – square matrix of correlation coefficients

Z^T – transform of Z matrix

λ – first order decay pole

μ_i – time series model coefficients

μ_y – mean of input signal

σ^2_a – white noise signal variance

σ^2_{MV} – minimum variance

σ^2_y – actual process variance

τ – process time constant

ψ_i– autoregressive moving average model coefficient

ψ - mother wavelet

$\|$ - Norm

Ω - width of frequency band

$\rightarrow$ Mapping

List of Figures

Figure 2- 1: Process block diagram 15

Figure 2- 2: Schematic of idealised slip-stick cycle 23

Figure 2- 3: Cross correlation function between input *u(t)* and output *y(t)*. Left: odd function (stiction oscillation), Right: even function (external oscillation) 26

Figure 2- 5: Schematic of cross correlation calculation points 27

Figure 2- 5: Graphic of decision criterion....28

Figure 4- 1: F1002 data 50

Figure 4- 2: Windowed CLP for F1002 53

Figure 4- 3: Windowed CLP for F1035 54

Figure 4- 4: PV-OP pattern for control loop F1035....55

Figure 4- 5: PV-OP pattern for control loop F1024....56

Figure 4- 6: General information for control loop F1081 57

Figure 4- 7: Zoomed PV-OP pattern for control loop F1081 58

Figure 4- 8: PV-OP pattern for control loop L1064....59

Figure 4- 9: PV-OP pattern for control loop L2051....60

Figure 4- 10: PV-OP pattern for control loop P1067 60

Figure 4- 11: Histogram of errors for F2081 61

Figure 4- 12: Histogram of errors for F1038....62

Figure 4- 13: Histogram of errors for F2024 63

Figure 4- 14: Cross correlation for L1030....65

Figure 4- 15: PV versus SP for L1030....66

Figure 4- 16: Cross correlation curve for F1021....67

Figure 4- 17: Cross correlation curve for F1038....67

Figure 4- 18: Cross correlation curve for F2024 68

Figure 4- 19: Cross correlation curve for L1043 69

Figure 4- 20: Gaussian distribution of L1043 69

Figure 4- 21: Autocorrelation of F2002 71

Figure 4- 22: Autocorrelation of L1025 72

Figure 4- 23: Autocorrelation of P1001 73

Figure 4- 24: Timed data for P1001 73

Figure 4- 25: Power spectrum analysis for control loop F1020 74

Figure 4- 26: Power spectrum analysis for control loop F1035 75

Figure 4- 27: Power spectrum analysis for control loop L2026 76

Figure 4- 28: Three dimensional common oscillation plot 77

Figure 4- 29: Flow loops three dimensional common oscillation detection 78

Figure 4- 30: Zoomed three dimentional analysis 79

Figure 4- 31: Bicoherence for control loop F1021 80

Figure 4- 32: Bicoherence for control loop F1035 81

Figure 4- 34: Time information for control loop F1035: PV, SP and OP information 82

Figure 4- 35: Full wavelet decomposition for the F1035 error signal 83

Figure 4- 36: Level 10 of control loop F1035 wavelet decomposition 84

Figure 4- 37: Signal estimation of F1035 using detail at level 10 of the wavelet decomposition 85

Figure 4- 38: Analysis of higher frequency oscillations in levels 4 and 5 of wavelet decomposition for control loop F1035 87

Figure 5- 1: SAMI process 91

1 Introduction

Basic control loops, based on PID technology, ultimately carry out most business decisions made in the chemical industry. Despite the integral role these controllers play in the production process, between 66% (Miller & Desborough, 2001) and 80% (Hugo, A.J, 2002) of all controllers in industry perform ineffectively.

Compounding this poor performance is the fact that most control engineers responsible for maintaining the basic control layer do not have the necessary tools at their disposal (Desborough, Nordh & Miller, 2001). The problem is thus two fold:

1. The performance of industrial processes is currently sub-optimal due to poorly performing basic controllers.
2. Control engineers often have no tools to measure the basic controller performance.

In a typical chemical process, the control engineer is responsible for hundreds or even thousands of individual controllers. This makes the unique analysis for each controller impossible.

Contrasted to this is the increasing acceptance that it is vital to process performance and profitability that control assets be monitored and maintained as vigorously as other plant assets. (Shah, Patwardhan, & Huang, 2001; Ingimundarson, 2001)

The objective is to evaluate predominately, but not confined too, closed loop monitoring techniques. The techniques that were evaluated are able to analyse large volumes of data.

Several different root causes exist for poor controller performance. Differentiating between hardware (valve stiction, design, process changes etc.) and software (control philosophy, tuning, algorithms etc.) and then in turn each of the subelements, is an important step in troubleshooting controller performance. (Gustafsson, & Graebe, 1998)

The purpose is to review the measures that currently exist to determine how effectively poor performance can be diagnosed. In addition, a new way to visualise and troubleshoot oscillations within control loops is presented. The method is based on the use of orthonormal wavelet basis functions.

The study is based on controller monitoring theory available in literature, results obtained from monitoring a petrochemical process, as well as the industrial experience of the author with available commercial products.

In order to clearly show the advantage presented by wavelet analysis, an extensive, though not exhaustive, review of existing closed loop performance measures is presented in Chapter 2. In this chapter several frequency (Fourier) based methods are used which forms the basis of the introduction to Wavelet decomposion, presented in Chapter 3.

A brief outline of the Methodology followed is presented in Chapter 4. The theory presented in both Chapter 2 and Chapter 3 was extensively applied to industrial data, and the results of this work is shown in Chapter 5.

A wide range of different examples are presented, and these are discussed at length in Chapter 6. The examples chosen are specifically designed to highlight the advantages and draw backs inherent in each of the individual measures. A discussion of industrial monitoring is also presented.

Finally, the conclusions and recommendation based on the above research are presented in Chapter 7.

2 Closed loop performance monitoring

2.1 Introduction

Many of the techniques used in quantifying performance make use of process models. For the purpose of this discussion, all methods will be discussed in terms of linear discrete time transfer function models (Luyben, 1989):

$$y_t = P(z^{-1})z^{-k}u(t) + D(z^{-1})d(t) + v(t) \tag{2. 1}$$

where y is the process output, u is the manipulated input, d is the disturbance, v accounts for the additive effect of noise and unmeasured disturbances. $P(z^{-1})$ and $D(z^{-1})$ are discrete transfer functions for the manipulated input and disturbance respectively.

Continuing along the same lines, it is possible to describe the feedback controller:

$$\begin{aligned} u(t) &= K(z^{-1})e(t) \\ e(t) &= y_{SP}(t) - y(t) \end{aligned} \tag{2. 2}$$

Figure 2.1 shows the block diagram for the process.

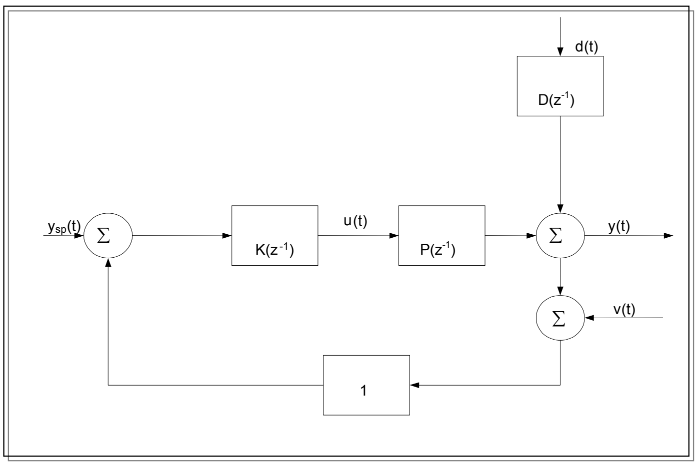

Figure 2- 1: Process block diagram (Stephanopouos, 1984: 570)

From the above, it is possible to derive the closed loop output error:

$$e(t) = \frac{y_{sp}(t) - D(z^{-1})d(t) - v(t)}{1 + P(z^{-1})K(z^{-1})} \tag{2. 3}$$

The dynamic response trend for e(t) can be solved using an Autoregressive Moving Average (ARMA) model of the form (Ljung, 2002)

$$e(t) = (1 + \psi_1 z^{-1} + \psi_2 z^{-1} + ... + \psi_n z^{-1})a(t) \tag{2. 4}$$

The moving average coefficients ψ_i are determined only from the combined effect of the open loop dynamics of the measured disturbances, d(t), the noise and unmeasured disturbances, v(t) and the set point changes, y_{SP}(t).

2.2 Minimum-variance

The development of the minimum variance (MV) index by Harris is often considered the start of feedback controller monitoring. This index or benchmark is based on a comparison between the actual performance of a controller and that of a theoretical MV controller.

A MV controller generates a controller output that will minimise the error between the set point and the output n controller executions in the future (where n is the number of controller executions within the dead time) (Tham, 1999).

In order for this controller to exist, we require a perfect model of the process, as well as a perfect model of the disturbance. Thus, one sample time after the process dead time, the controller error will be zero (except for the measurement noise) **(**Harris, Seppala & Desborough, 1989).

The performance index a.k.a the closed loop potential (CLP) is obtained by taking the ratio of the minimum achievable variance (σ_{mv}^2) to that of the actual controller variance (σ_y^2).

$$CLP = \frac{\sigma_{mv}^2}{\sigma_y^2} \tag{2. 5}$$

The above measure was first introduced by DeVries and Wu (1978), and is known as the Closed-Loop Potential (CLP). When the CLP is 1 the best achievable feedback control is being applied relative to both the process and the disturbance characteristics. When the CLP is equal to 0 there is effectively no control driving the error to zero. By evaluating the CLP in combination with the actual variation in the process it is possible to ascertain whether changes need to be made to the controller (CLP close to 0 but σ_y^2 high), or to the source of the variability (CLP close to 1 but σ_y^2 high).

In order to calculate the CLP the individual variances are required. The actual process variance , σ_y^2, is easily obtained from process data.

As mentioned, in order to design, and thus calculate the variance of a MV controller, one requires a perfect process and perfect disturbance model. Neither of these are readily available. However, it is possible to determine the closed loop response of a MV controller given only the dead time of the process, and the closed loop data.

The MV is estimated by fitting a time series model of the form of equation 2.4 to the closed loop error e(t), given that n is equal to the number of sampling intervals within the dead time. Thus

$$\sigma_{MV}^2 = (1 + \psi_1 z^{-1} + \psi_2 z^{-1} + \ldots + \psi_n z^{-1})\sigma_a^2 \qquad (2.\,6)$$

where the variance σ_a^2 is obtained from the variance in the disturbance, and the coefficients ψ_i.

Assumed in equation 2.6 is that the mean of the measured disturbance is 0. Thus, effectively the input vector is a representation of zero mean noise, with a variation of σ_a^2. Equation 2.6 gives an indication of how well the perfect controller could reject the noise, given the inherent process dead time.

The crux of the Harris approach is that a model of the process is not required. The reasoning is that since the controller cannot affect the system output before the process dead time has elapsed, the first n-1 coefficients of the impulse response of the closed loop transfer function must be equal to the first n-1 coefficients of the disturbance or the noise model. This means that the performance of a controller can be determined using only knowledge of the dead time of the system. This property is known as the invariance property.

2.3 Modified minimum variance

2.3.1 Requirement

Before discussing modified MV techniques, it is necessary to understand the requirement for such measures. There are a number of inherent limitations to the MV measure (Horch, 1998).

1. The MV technique is only appropriate if the major dynamic in the system is dead time, ie the process is dead time dominant.
2. There is no way to achieve MV control except through the application of MV controllers. If the actual variance is close or equal to the MV, then the implemented controller is in effect a MV controller.
3. When the controller structure is pre determined, as with PID controllers, the comparison to a MV controller may be harsh. It is only appropriate to use MV benchmarking if the PID controller approximates the MV controller, ie when:
    - There is negligible dead time, for example flow loops
    - The process order is low
    - The unmeasured disturbances are stationary.

   Approximately 20% of all PID control loops in the refining industry fall into this category (Qin, 1998).
4. If variability is intentionally allowed in a controller, MV comparisons will be meaningless.

The above set of limitations immediately eliminate the applicability of the MV index to any process that has:

- Any multivariable interactions
- Process non-linearities
- Physical process input/output constraints

2.3.2 Closed-loop pole modification

This method is derived specifically for processes where the system dead time is very short. In these cases there may be other process limitations which have a larger influence on the achievable performance.

By taking performance limitations due to non-minimum phase zero's and non-stable poles into account, it is possible to modify the Harris index to make a more accurate performance assessment. (Tyler and Morari, 1995) However, a major limitation exists in that this approach requires the user to identify the unstable poles and zeros.

A more useful approach would be to modify the index, but to make use of only the measured output and the process dead time, as in the original MV index. The difference is introduced in that not all the closed loop poles are assumed to be at the origin, as in MV control. Rather, one additional pole is introduced. This modified index is by necessity always greater than the Harris MV index.

It can be shown (Harris, 1998) that the placement of the additional pole is obtained by lagging the time series in equation 2.6. Thus:

$$y(t)=\left[\left(1+\psi_1 z^{-1}+\psi_2 z^{-1}+\ldots+\psi_n z^{-1}\right)+\psi_n\left(1+\mu z^{-n-1}+\mu^2 z^{-n-2}+\ldots.+\mu^i z^{-n-i}\right)\right]d(t) \quad (2.\ 7)$$

where μ^i gives the time series coefficients for the closed loop pole.

This modified index has the property of being larger, ie indication of more sluggish control, the further we place the additional pole from the origin. Intuitively this makes sense, since the location of a pole should make the achievable control worse. This achievability is then taken into account by the new measure.

The major limitation with this modified index is process knowledge. In order to sensibly place the additional closed loop pole, one requires knowledge of the slowest process time constant. Generally, there is the feeling that knowledge of the dominant time constant is not more difficult to obtain than an accurate indication of the dead time, and as such should not be dramatically more difficult to implement.

Once the dominant time constant is known, the pole is placed so as to indicate to what degree the process should be speeded up by control, ie if the process should be speeded up 3 times, the pole would be $\mu \geq e^{\left(-\frac{3T_S}{\tau}\right)}$.

2.4 Auto correlation

The auto correlation test is based on one of the properties of minimum variance. Since the auto correlations of the sampled errors are zero for all lags greater than the period of dead time (for minimum variance controllers), a comparison to minimum variance based on this condition is possible. (Harris, 1989; Huang *et. al.*, 2000)

As with the basic minimum variance testing, the rating of output error based on this “harsh” minimum variance technique is often not practical. In order to make the result more representative, a different autocorrelation pattern can be used to represent the condition of minimum variance. Consider the case where a first order decay of output is acceptable.

$$e_t = \frac{1}{1-\lambda z^{-1}} a_t \qquad \lambda = \exp\left(-\frac{T}{\tau}\right) \tag{2. 8}$$

The autocorrelation pattern can then be given as

$$p_e(k) = \lambda^k \tag{2. 9}$$

The values of $p_e(k)$ can then be compared to the autocorrelation coefficients. (Kozub, 1996)

2.5 Filtering and Correlation analysis

In the previous sections outlining the measure of minimum variance benchmarks, very little detail about the actual calculation was presented. In this section, an algorithm outlining an efficiency algorithm for the calculation of the minimum variance performance benchmark is given (Huang and Shah, 1999).

Equation 2.6 shows the minimum variance performance of the process. This can also be termed the invariant portion of the output variance, since the controller has had no effect on the process before the number of time steps of the dead time have elapsed.

In order to calculate the minimum variance, equation 2.6, it is necessary to relate the process variation to the minimum variance, which is in turn directly determined by the input variance. As mentioned, the input variation is assumed to be white noise excitation, with zero mean.

By manipulating the equation relating output process variation to input variation, equation 2.4, and using the definition of closed loop potential, equation 2.5, it is possible to show that,

$$CLP \equiv ZZ^T \tag{2. 10}$$

where Z is the vector of the cross-correlation coefficients between the input and output vectors for lags 0 to n-1 (where n is the number of intervals of dead time).

In solving the above it is an obvious requirement to find an input vector that meets the requirements of a whitened noise sequence. This signal can also be termed the "innovation sequence", and several procedures for obtaining this sequence exist. The procedures make use of time series modelling, and depending on the data, either an autoregressive (AR) or an autoregressive moving average (ARMA) model can be picked. The final test of model acceptance becomes an analysis of the "whiteness" of the residuals.

2.6 Nonlinear techniques

Minimum variance based techniques/indices are based on the assumption that the system is linear. For systems where this assumption approaches reality, the diagnosis of performance given by minimum variance type techniques is good. However, when systems have substantial process nonlinearities, the accuracy of the indication rapidly decays. (Horch, 1998)

While most processes approach linearity around the operating point (hence the suitability of PID control algorithms), one form of nonlinearity is commonly encountered in industry.

2.6.1 Slip-stick cycling

Nonlinearity arises because of the saturation of the actuator due to static friction, commonly called stiction. The static friction in the valve needs to be overcome by the force exerted by the actuator. The actuator force continues to increase until the friction exerted by the linkage, positioner and valve body are overcome. This increase in force occurs with no change in the valve position. When the energy exerted exceeds the friction forces, the valve slips to a new position, the move being disproportionately large due to the continued feedback action.

Figure 2.2 shows the effect of this slip/stick cycle.

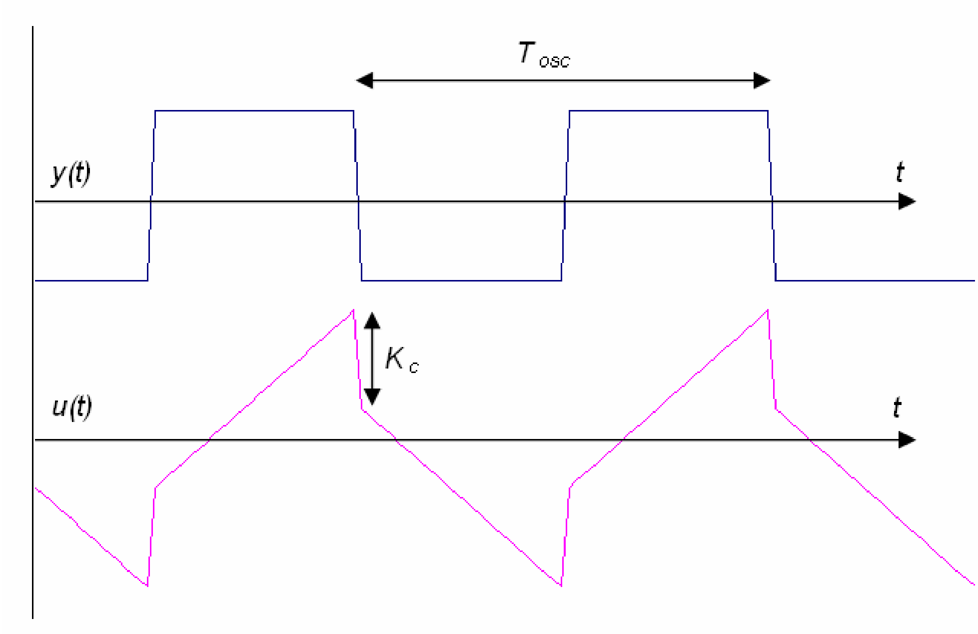

Figure 2- 2: Schematic of idealised slip-stick cycle

A detailed description of the effect of friction in valves can be found in (Hägglund, 1995)

In figure 2.2, the control algorithm was assumed to be PI control. The step in the output signal (*u(t)*) is due to the proportional action (K_c) of the controller. The period of idealised oscillation can be calculated to be,

$$T_{osc} = 4T_i\left(\frac{1}{K_pK_c} - 1\right) \qquad (2.\ 11)$$

where T_i and K are the integral time and the process/controller gains. The control signal *u(t)* will become triangular, ie the step will disappear, when the process *y(t)* shows a first order decay rather than the rectangular process response shown above.

2.6.2 Normality index - histograms

By plotting the histogram of a batch of data, it is possible to visually inspect the Gaussianity of the data. This gives an indication of whether the data is normally distributed, and gives information about whether the system is linear or not.

By normalising the data, (normalise the error rather than an absolute value), it is possible to automate the evaluation of the normality of the data. The following algorithm is suggested:

1. Calculate the mean (μ) and the standard deviation (σ) from the collected error vector.
2. Calculate the distribution of the error, and save the frequency of the error versus the bins.
3. Calculate the normal distribution *f(x)* characterised by the same information determined for the system, ie μ and σ. Minimise the sum of squared errors between function *f(x)* and the histogram, using a single multiplier *K*.
4. Based on the difference between the Gaussian curve and the histogram, use one of the following 2 measures to determine the normality of the data:
    - Calculate the mean squared error of the cumulative sum of the histogram and the Gaussian curve. If MSE > 0.01, flag oscillatory data.
    - If $\frac{|\mu_{data} - \hat{\mu}|}{\sigma_{data}} > 10\%$ then flag data as oscillatory. This calculation effectively calculates the goodness of fit between the histogram and the Gaussian curve.

By making use of this algorithm, it is possible to detect skewed or poorly shaped distributions. This check allows an easy detection of oscillation. However, to make the distinction between oscillation due to disturbances or due to nonlinearity is difficult. It is normally accepted that a histogram with two maxima (due to the rectangular oscillation of the output) is a sign of stiction. However, this is not always the case, and can be confused with certain instances of tight tuning, where two maxima are also present in the histogram (as will be demonstrated later).

2.6.3 Cross correlation

The cross correlation method is a simple way of diagnosing the source of oscillation in control loops. The method is based on the following observations:

- When the input signal, $u(t)$, and the output signal, $y(t)$, are oscillating, then the cross correlation function is also periodic.
- When stiction is prevalent in a control valve, the input and output signals are shifted by $\pi/2$, relative to each other. This results in a cross correlation function that is odd.
- When the input and output signals are oscillating 180° (π) out of phase, then the source of the oscillation can be shown to be disturbance based. This will result in a cross correlation function that is even.
- If the control loop is unstable, and oscillates with a constant amplitude due to input saturation, then the cross correlation function can also be shown to be even.

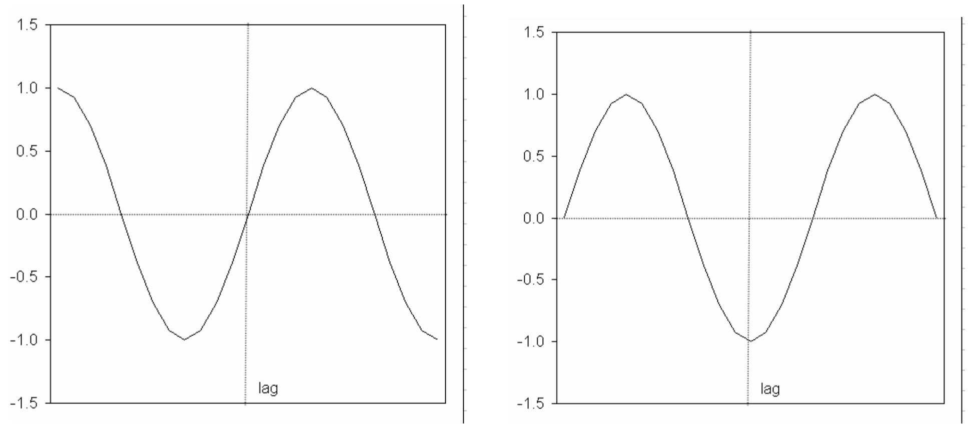

Figure 2- 3: Cross correlation function between input *u(t)* and output *y(t)*. Left: odd function (stiction oscillation), Right: even function (external oscillation)

The cross correlation function $r_{uy}(\tau)$ for signals *u(t)* and *y(t)* is defined as,

$$r_{uy}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} u(t+\tau) y(t) dt \qquad (2.\ 12)$$

The assumptions for the calculation of the cross correlation function are:

- The gain between the input signal *u(t)* and the output signal *y(t)* is positive. If the gain is actually negative, then the function needs to be inverted. However, this will have no bearing on whether the function $r_{uy}(\tau)$ is odd or even.
- The data is pre treated such that the signals are stationary and of zero mean.

The calculation of the cross correlation function $r_{uy}(\tau)$ from equation 2.12 is easily done. More challenging is the automation of whether the function is odd or even, the criterion used for the determination of the source of the oscillation. The following algorithm is proposed for this determination:

1. Calculate $r_{uy}(\tau)$
2. Calculate τ_r the first zero crossing for positive lags

3. Calculate $-\tau_l$ the first zero crossing for negative lags

4. Calculate r_0 the value of $r_{uy}(\tau)$ at lag 0

5. Calculate r_{max} the maximum value of $r_{uy}(\tau)$ in $[-\tau_l, \tau_r]$

6. Calculate $\Delta\tau = \frac{|\tau_l - \tau_r|}{\tau_l + \tau_r}$

7. Calculate $\Delta Q = \frac{|r_0 - r_{\max}|}{|r_0 + r_{\max}|}$

8. The aim is to determine the degree of phase shift. Allow a deadband of $\pi/6$. Thus, interpret a phase shift $\frac{\pi}{2} \pm \frac{\pi}{6} = \frac{\pi}{2}$ and $\pi \pm \frac{\pi}{6} = \pi$

Figure 2.4 shows the points as defined in the above calculations:

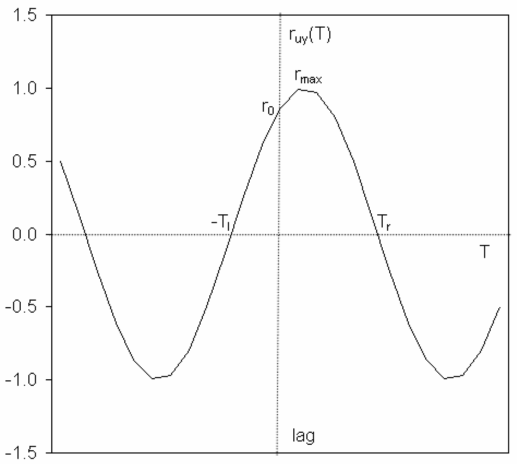

Figure 2- 4: Schematic of cross correlation calculation points

Thus, if we can graphically represent the decision criterion for whether a control loop has stiction or not. Figure 2.5 shows this representation.

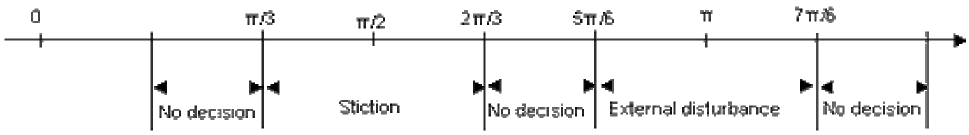

Figure 2- 5: Graphic of decision criterion

Notice that a safety band of *π/6* is placed between decisions. This is a grey area where the decision is unclear. It may be interpreted as an unusual problem area, where sensor faults may be the root cause of oscillation.

A similar technique based on the analysis of distinct peaks in the histogram of the dataset, and using segmentation and change detection can also be used. (Stenman, Forsman & Gustafsson, 2004)

2.7 Statistics

2.7.1 Lower order statistics

Most engineers are familiar and comfortable with lower order statistical measures. Lower order statistics refers to measures such as:

- Mean
- Variance
- Autocorrelation
- Power spectrum

These are examples of first and second order statistical tools, and are often a good first indication of loop performance.

The magnitude of the mean of the error function between the output signal $y(t)$ and the setpoint $y_{sp}(t)$ can be an indication of skewness in the data set.

Second order statistics (eg variance) is only applicable as a diagnosis tool when the process is linear, and the distribution thus Gaussian. Once again, the magnitude of the variance is a good indication of the size of the oscillation present in the control loop.

2.7.2 Higher order statistics (HOS)

First and second order statistics are only sufficient when processes are linear and Gaussian, ie the process can be described by a linear transfer function. However, as mentioned previously, there are many cases where the process deviates from Guassianity, and then exhibits nonlinearities. These types of processes can be conveniently studied using HOS (Choudhury, Shah & Thornhill, 2002).

The central moment of a variable is defined as:

$$M_i = E(x-\mu)^i = \frac{\sum_{t=1}^{n}(x_t-\mu)^i}{n} \tag{2. 13}$$

Thus, the second moment is equal to the variance of the data set, normalised by n rather than n-1, where n is the number of data points in the series. Higher order moments give information about skewness (3^{rd} order) and sensitivity to outliers (4^{th} order, kurtosis). The skewness of a Gaussian distribution is 0, and the kurtosis of a Gaussian distribution is 3.

In some instances, it is useful to make use of cumulants rather than moments. Cumulants are directly related to moments:

$$C_i = \ln(M_i) \tag{2. 14}$$

Cumulants have excellent noise suppressing properties, making them a useful indication of the variability inherent in the data. The following relationships hold:

$$\begin{aligned} C_1 &= M_1 \\ C_2 &= M_2 - M_1^2 \\ C_3 &= M_3 - 3M_2M_1 + 2M_1^3 \\ C_4 &= M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4 \end{aligned} \tag{2. 15}$$

2.8 Frequency analysis

While much information can be gleaned from the time domain study of process data, many useful measures require transformation to the frequency domain. The frequency domain data exposes periodicities in the signal which can aid in the understanding of the source of oscillation (Shah, Patwardhan & Huang, 2001).

2.8.1 Power spectrum

The power spectrum is a representation of the second order moment in the frequency domain (Choudhury, Shah & Thornhill, 2002). The power spectrum can be calculated by taking the Fourier Transform of the signal. The fast Fourier transform is calculated using:

$$X(f) = \sum_{n-1}^{N} x(n) w_N^{(f-1)(n-1)}$$
$$\text{Where}$$
$$w_N = e^{\frac{-2\pi i}{N}} \tag{2. 16}$$

Using this method, it is possible to determine the power of the signal at various frequencies, thereby determining the dominant oscillation periods.

2.8.2 Bicoherence

The bispectrum is the frequency domain equivalent of the third order cumulant. It is defined as:

$$B(f_1, f_2) \equiv E[X(f_1)X(f_2)X(f_1 + f_2)] \tag{2. 17}$$

where X(f) is the Fourier transform of the data series. In order to more easily interpret results (ie ensuring that results are normalised and independent of signal strength), a bicoherence is defined based on 2.16:

$$bic^2(f_1, f_2) = \frac{\left|E[B(f_1, f_2)]^2\right|}{E[\left|X(f_1)X(f_2)\right|^2]E[\left|X(f_1 + f_2\right|^2]} \tag{2. 18}$$

The bicoherence function is bounded between 0 and 1. It can be shown that the bicoherence for a linear system is constant. Thus, if the bicoherence is plotted against f_1 and f_2 in three dimensional space, a linear system approximates a plane, where the surface has features where the system displays nonlinearities. (Choudhury, Shah & Thornhill, 2002)

2.8.3 Common oscillation detection

The power spectrum has been shown to be a useful measure for dominant frequencies. However, in evaluating large numbers of signals, an automated method of detecting the common oscillations is required. The following method allows this detection by making use of principal component analysis (PCA). By taking the Fourier decompositions for several data sets, one defines a matrix:

$$A = \begin{pmatrix} P_1(f_1) & \dots & P_1(f_N) \\ \dots & \dots & \dots \\ P_m(f_1) & \dots & P_m(f_N) \end{pmatrix} \tag{2. 19}$$

where N is the number of frequency channels and m is the number of signals being evaluated. $P_i(f_j)$ gives the power of the i^{th} signal at the j^{th} frequency channel (Thornhill, Shah & Huang, 2001). A PCA decomposition of A allows the reconstruction of the matrix as a sum over m orthonormal basis functions w'_1 to w'_m. This yields,

$$A = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} w'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} w'_2 + \dots + \begin{pmatrix} t_{1,m} \\ \dots \\ t_{m,m} \end{pmatrix} w'_m \tag{2. 20}$$

The w'_l vectors are normalised eigenvectors of the m by m matrix A'A. The order is determined by the magnitude of the eigenvalues of A'A. The magnitude of the eigenvalue compared to the sum of all the eigenvalues gives an indication of the total spectral variation captured by that eigenvector.

In order to graphically represent the data captured in this manner, it is useful to truncate the decomposition to the first three principal components. This can be validated by determining the magnitude of the next principle component in the series. A good rule of thumb is to accept the truncation if the next component represents less than 5% of the sum of eigenvalues (this approach is demonstrated in the results section).

$$A = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} w_1^{'} + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} w_2^{'} + \begin{pmatrix} t_{1,3} \\ \dots \\ t_{m,3} \end{pmatrix} w_3^{'} + E \qquad (2.\ 21)$$

It is then possible to represent the spectrum in A graphically. The i^{th} vector maps to a point having the coordinates $t_{i,1}$, $t_{i,2}$ and $t_{i,3}$ in a three dimensional space. This plot is known as a scores plot.

2.9 Input – Output pattern recognition

The plot of the process variable or output *y(t)* versus the process demand or input *u(t)* gives useful information about the linearity of the system. Sharp turn-around points at the extremities of the plot show clear signs of stiction in the data. A perfectly linear relationship between the PV and OP is the control engineers dream, but is not the norm, especially in older industrial installations.

Another way of interpreting the pattern produced is to evaluate the one on one relationship/repeatability. The process variable should always have a single – unique value for each input. Due to process disturbances and other external influences, this is almost always impossible (changes in upstream pressure affect valve opening to get constant flow). However, patterns approaching such a deterministic relationship indicate a low probability of valve problems.

Unusual shapes are often encountered, indicating unknown sources of process nonlinearity.

2.10 PID-achievable performance

Roughly 97% of all regulatory or basic control in the chemical industry is done by PID feedback controllers. (Desborough & Miller, 2001) As such, a measure of "best possible" PID performance is very useful. By solving the following,

$$\sigma_{PID}^2 = \min \sigma_y^2 \tag{2. 22}$$

where,

$$\sigma_y^2 = \sum_{i=1}^{\infty} \psi_i^2 \sigma_w^2 + \sigma_w^2 \tag{2. 23}$$

and

$$y(k) = \sum_{i=1}^{\infty} \psi_i w(k-1) + w(k) \tag{2. 24}$$

Thus the minimum achievable PID variance can be calculated by estimating the closed loop transfer function based on a PID controller. This transfer function should relate the process output (y) to the noise or disturbance (w).

Unfortunately, this process is difficult to implement since the open process model is often not known (Qin, S. J., 1998). As such, this measure cannot be automated.

A different approach, based on the comparison between a user specified desired benchmark and actual loop performance, is presented by Huang, *et al.,* 1998.

2.11 Open loop benchmark

This metric is not always available in an industrial application, but does provide excellent insight into the real improvement made by a controller. The open loop variance in a process should be a bound never exceeded with control. However, some studies have found that up to 75% of controllers lead to an increase in the process variance when compared to the open loop variance. Thus, when the opportunity does arise to obtain open loop variance information, this information can be very useful (VanDoren, 1999).

While not entirely an open looping procedure, it is also possible to improve the analysis of the robustness and performance of the controller by means of injecting an excitation signal (Wan & Huang, 2002).

2.12 Interventions

Measurement of the interventions for a given control loop is a non-technical measure of performance. If regular interventions by the process personnel is required, it is an indication that there may be a problem with the loop. Interventions can be measured either by having the loop in the incorrect mode, or by the number of set point movements that are required. In configuring a system to correctly measure interventions, the effort must be expended to correctly define “normal operation” mode.

Interventions cannot be used as an exact measure, but the decrease in the number of interventions may seem to be an improvement in “customer satisfaction” with the operation of the controls.

3 Wavelet Analysis

Most published material on wavelets introduce Fourier analysis as a basis from which to approach the subject (Mitsiti, *et al*: 1996, Carrier & Stephanopoulos, 1998). Often, once this basis has been discussed, the author continues onto wavelets and the Fourier analogy is left hanging. Fourier analysis is more than just a basis from which to start discussing wavelets. It forms the essence of what wavelets analysis is, and all through the discussion on wavelets, a firm grasp of the theory of Fourier analysis is necessary.

3.1 Fourier Analysis

Let H_T be a set of functions *f*: **R** → **C**, where the functions are periodic with period T. These functions satisfy equation 3-1 (Kaiser, 1994).

$$\|f\|_T^2 \equiv \int_{t0}^{to+T} |f(t)|^2 dt < \infty \qquad 3.1$$

H_T is the Hilbert space under the inner product defined by equation 3.2

$$\langle f,g \rangle_T \equiv \int_{t0}^{t0+T} \overline{f(t)} g(t) dt \qquad 3.2$$

By considering the functions in equation 3.3, we can see that they are T-periodic, and thus a part of H_T.

$$e_n(t) = e^{2\pi \mathrm{int}/t} = e^{2\pi w_n t}, w_n \equiv \frac{T}{n} \qquad 3.3$$

Let *f* be a function, $f \in H_T$, which can be expressed as a linear combination of e_n's. Then equation 3.4 and 3.5 hold (Kaiser, 1994).

$$f = \frac{1}{T} \sum_{n=-\infty}^{\infty} c_n e_n \qquad 3.4$$

$$c_n = \langle e_n, f \rangle_T = \int_{t0}^{to+T} e^{-2\pi i w_n t} f(t)dt \qquad 3.5$$

Equation 3.4 is called the Fourier series of *f*, and c_n are the Fourier coefficients of *f*. *f*(t) represents a periodic signal, and equation 3.4 gives the decomposition of *f*. This decomposition is a linear combination of modes e_n, with frequencies w_n. Thus, if one is able to identify which frequencies in the decomposition are most desirable, one can reconstruct the signal from the decomposition, but without the unnecessary frequencies. This means that one can either remove undesirable noise, or increase the effect of desirable frequencies so that the reconstructed signal appears to be stronger in the frequency region of most concern (Kaiser, 1994).

Thus, by applying the Fourier transform, we move the signal information from the time to the frequency domain. The inverse Fourier transform moves from frequency back to time.

It is possible to show that $<e_n, f> = e_n{}^*f$. Thus 3.4 and 3.5 can be simplified to equation 3.6.

$$f = \frac{1}{T} \sum_{n=-\infty}^{\infty} e_n e_n {}^*f \qquad 3.6$$

Rather that continue to function with a discrete summable definition of the Fourier transform, we define the continuous or Fourier integral transform. This form is given in equation 3.7.

$$\hat{f}_T(w) \equiv \int_{-\infty}^{\infty} e^{-2\pi i w t} f(t)dt \qquad 3.7$$

In order to transform or reconstruct from this frequency decomposition, one requires the inverse Fourier transform. This is given in equation 3.8.

$$f(t) \equiv \int_{-\infty}^{\infty} e^{2\pi i w t} \hat{f}_T(w) dw \qquad 3.8$$

3.1.1 Windowed Fourier transform (WFT)

While the Fourier transform serves as a useful tool, it is severely limited. Most signals are not by nature periodic. This means that the direct application of the Fourier transform is meaningless. By integrating over all time, the decomposition would contain the total amplitude of all w_n's, over the entire duration of the signal, while, for the purpose of analysis, the distribution of the amplitudes is of vital importance (Kaiser, 1994).

In describing the workings of the WFT, it is useful to use the analogy of the ear. In order for the ear to process sounds, the signal passing to the ear needs to the analysed in the time and frequency domain simultaneously. In order to better define this analysis, the following assumptions are made.

The ear can only analyse sounds that have already happened, ie can only analyse $f(u)$, for $u \le t$. In addition, it is reasonable to assume that the ear has a finite memory, ie it only processes sounds that have taken place in a finite time period. Thus it only analyses $f(u)$, for $u \ge t - T$.

Thus, the WFT, $\hat{f}_T$ (w, t), can only depend on the outputs, $f(u)$, in the time interval $t - T \le u \le t$.

In order to localise the signal $f(u)$ on this time interval, it is necessary to introduce a weighting function $g(u)$. $g(u)$ localises the signal in time, and supp $g(u) \subset [-T, 0]$.

It is now possible to define a localised version of f, where $f_t \subset [t-T, t]$, and is dependent only on $f(u)$, $t - T \le u \le t$. Equation 3.9 formally defines f_t.

$$f_t \equiv \overline{g(u-t)} f(u) \qquad 3.9$$

Based on 3.9, the definition of the WFT ($\tilde{f}(w,t)$) is given by equation 3.10

$$\tilde{f}(w,t) \equiv \hat{f}_t(w) = \int_{-\infty}^{\infty} e^{-2\pi iwu} f_t(u)du \qquad 3.10$$

By combining 3.9 and 3.10, it is possible to arrive at the formal definition of the WFT, 3.11.

$$\tilde{f}(w,t) \equiv \hat{f}_t(w) = \int_{-\infty}^{\infty} e^{-2\pi iwu} f(u)\overline{g(u-t)}du \qquad 3.11$$

g(u) is an element of L^2(R). $\overline{g(u-t)}$ is the complement of *f(u)* within $f_t(u)$.

In order to simplify the definition presented in equation 3.11, we introduce an operator form of the weighting function g(u).

Let $g_{w,t}$ be defined by equation 3.12.

$$g_{w,t} \equiv e^{2\pi iwu} g(u-t) \qquad 3.12$$

Since $\| g_{w,t} \| \equiv \| g \|$, $g_{w,t}$ is also an element of L^2(R). Now it is possible to express the definition of the WFT, presented in equation 3.11, as the inner product (convolution) of $g_{w,t}$ and f. Thus, equation 3.13 can be used as an expression of the WFT.

$$\tilde{f}(w,t) = \langle g_{w,t}, f \rangle \equiv g_{w,t} * f \qquad 3.13$$

In the above discussion we have dealt with the WFT from the viewpoint of the time domain. This has shown that it is possible to localise the signals frequency content near the time t. It is possible to adopt the same argument from the viewpoint of the frequency domain.

This is done in equation 3. 14. Note that the time variable u has been replaced by a frequency variable v, and that the time window g(u - t) has been replaced by the frequency window $\hat{g}$ (v – w).

$$\tilde{f}(w,t) = \int_{-\infty}^{\infty} e^{2\pi iwu} f(v)\overline{\hat{g}(v-w)}dv \qquad 3.14$$

What equation 3.14 effectively means is the following. If one starts with a signal $f(v)$ and localises it near a frequency w, using the weighting function $\hat{g}$ $(v\text{-}w)$, then the WFT gives a local time-frequency analysis of f. This is provided that the windowing function g is reasonably well localised in time and frequency. This is to say that $\hat{g}$ (v) is small outside a small frequency band, and $g(t)$ is small outside a small time interval.

Thus the WFT of a function f yields a simultaneous localisation in time and frequency (Kaiser, 1994).

3.1.2 Uncertainty

Since the measurement of an instantaneous frequency is an invalid concept (by it's very definition), some uncertainty about the localisation of the signal in time and frequency must exist (Kaizer, 1994). This can be quantified by the Heizenburg uncertainty principle which states: Since sharp localisations in time and frequency are mutually exclusive, the precise measurement of time and frequency must be fundamentally incompatible. Thus, if $g(t)$ is small outside a time interval T, and $g(v)$ is small outside a frequency band of width Ω, then the inequality of $\Omega T > c$ must hold. The precise value of c is dependent on the size of Ω and T (Kaiser, 1994).

Thus, effectively the product of the uncertainty about the time measurement with the uncertainty of the frequency measurement, must be greater than a certain constant.

This can be better described as follows. By observing a signal over an extended period of time, the better the knowledge and accuracy about the signal's frequency becomes. However, due to the increased time interval over which the signal was observed, there is less precise time information. Thus, by improving the knowledge of the frequency content of the signal, one sacrifices time localisation.

This is exactly what the WFT does. By observing a signal *f*(t) over the window $\bar{g}$ (u-t), the value of t is no longer sharp. Rather, a unique t has been replaced by a interval of [t-T, t], with supp $g \subset$ [-T, 0]. In addition, the frequency *w* is no longer sharp, but rather a band determined by $\bar{\hat{g}}$ (*v*-*w*).

Thus, in effect, the WFT compromises between the time and frequency resolution to arrive at a better global understanding of the characteristics of the signal.

3.2 Why the Wavelet (Mathematically)

Consider a highly localised impulse, δ(t). This function is the limit of spikes of ever increasing frequency and sharpness. Equation 3.15 shows the mathematical relationship for δ(t).

$$\delta_\varepsilon(t) = \frac{1}{\pi}\frac{\varepsilon}{t^2+\varepsilon^2}; \varepsilon > 0 \qquad 3.15$$

By applying the normal Fourier transform to δ(t), one arrives at the set of equations 3.16.

$$\begin{aligned} \hat{\delta}(w) &= \int_{-\infty}^{\infty} e^{-2\pi i w t}\delta(t)dt \equiv 1 \\ \delta(t) &= \int_{-\infty}^{\infty} e^{2\pi i w t}\hat{\delta}(w)dw = \int_{-\infty}^{\infty} e^{2\pi i w t}dw \end{aligned} \qquad 3.16$$

This is the extreme case of the uncertainty principle. Since δ is sharp, T → 0. Thus the Fourier transform must have width Ω → ∞. Then, in order to obtain a spike with T ≈ ε, we must have exponentials in the frequency band 1/ε. Thus, the narrower the spike δ becomes, the more high frequency exponentials are required in the construction of δ. To obtain the sharpness required, more exponentials are needed to cancel the oscillations before and after the desired spike.

Thus, using the normal Fourier transform to construct δ, one effectively attempts to compose local variations in time, using non-local exponentials.

By developing the WFT, this problem is partially dealt with. By superimposing $g_{w,t}$ and using a coefficient function $\tilde{f}(w,t)$, one can reconstruct the signal *f*. $g_{w,t}$ can be seen as different 'notes' within *f*. Thus, when defining a sharp spike δ, with width << T, where T is the width of $g_{w,t}$, one needs to rely on the constructive and destructive interference of many different notes. Similarly, when constructing a signal with the width of $g_{w,t}$ << width of δ, one needs to rely on the constructive interference of many notes.

It may be questioned how this is different to the case of relying on the sum of many different exponentials, in the case of the ordinary Fourier transform. Since the WFT introduces a scale into the analysis of the signal, this interference is limited to the width of the window, *g*. This reduces the interaction required to achieve the required signal.

Both the ordinary Fourier transform, and the WFT, are cumbersome when trying to achieve a simultaneous time frequency localisation (indeed, it is impossible to achieve a dual localisation using the ordinary Fourier transform, and the single window width of the WFT, makes the analysis of events of different widths tedious.). Wavelet analysis provides the ideal analysis method to deal with this problem of resolution and localisation.

3.3 Why the Wavelet? (Motivational)

Firstly, the idea of a basis of a vector space needs to be reviewed. The basis of a vector space is the set of linearly independent vectors having the property that every vector in the space is a linear combination of all these vectors (O'Neil. 1995: 311).

The advantage of representing a function in terms of a basis set is that the desired information contained in a complex function can be more easily detected, extracted, or manipulated.

The Fourier transform is a method that makes use of a frequency-localised basis. The Fourier transform uses the sinusoid as a basis, each frequency being detailed by a sinusoid of different period. Since this basis is localised in frequency and is global in time, one looses all localisation of events in the time domain. This means that any non-stationary feature of a signal will not be captured. A good example of this is a burst of high frequency information.

The windowed Fourier transform attempts to deal with this by using a sliding window over time. Thus it decreases the periods over which there is a loss of locality. However, due to the uncertainty principle outlined above, this introduces a problem of its own. The window needs to be sufficiently long to contain enough information to obtain a reasonable frequency resolution. However, the window needs to be as short as possible to get a sharp time locality. Since the two concepts are mutually exclusive, a compromise must be reached.

Wavelets can be viewed as an extension to the Fourier transform. Their power lies in their ability to address the problems of a non-stationary signal by performing a local analysis.

3.4 What is a Wavelet?

A wavelet, as shown in figure 3.1, is a waveform of limited duration with a mean of zero. This can be compared to the basis of the Fourier transform, the sine wave, which is not limited in time, ie stretches from negative to positive infinity.

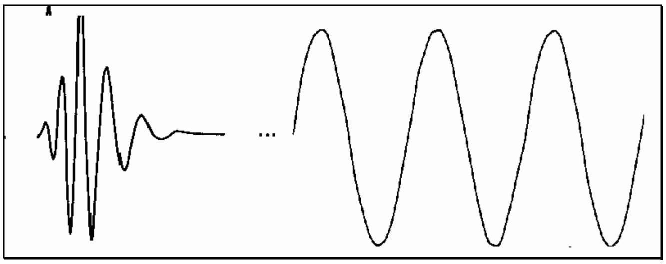

Figure 3.1: Comparison between wavelet and Fourier sine basis (Misiti et al, 1996)

The reason for the original choice of representation for the Fourier transform may be questioned. Most engineers are more familiar with the use of the sine and cosine forms of the Fourier transform. However, the transition into the wavelet transform is more easily accomplished once comfortable with the form given in equation 3.14.

Equation 3.17 gives the definition of the wavelet transform which is analogous to the Fourier transform in equation 3.14.

$$\tilde{f}(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad 3.17$$

Both the Fourier function in equation 3.14 and the wavelet function in equation 3.17 use a family of functions indexed by two labels, $g^{w,t}(s) = e^{iws}g(s-t)$ and $\psi^{a,b}(s)=|a|^{-1/2}\psi((s-b)/a)$. The functions $\psi^{a,b}(s)$ are known as wavelets, with ψ often being called the mother wavelet. In this wavelet function, *a* is known as the scaling parameter. As *a* changes the wavelet function covers different frequencies. A high absolute value of *a* corresponds to a very coarse scale evaluating low frequencies, while a finer scale (low *a*) gives a high frequency coverage. The parameter *a* is also sometimes known as the dilation operator.

The *b* parameter is the translation factor. It allows the user of the wavelets to move the time localisation centre.

The main difference between the wavelet and the WFT is the shape of the analysing window. The functions $g^{w,t}(S)$ all consist of the same envelope function g, translated to the correct point in time. Irrespective of the frequency being analysed, the window retains the same width. The wavelet function $\psi^{a,b}(s)$ is sharply in contrast to this with the time window being adapted to take the frequency into account. At high frequencies the windowing parameter is very small, allowing narrow windows capturing high frequency events. At low frequencies the window is much broader (Carrier & Stephanopoulos, 1998).

There exist many different types of wavelet transform, but all start from the basic formula given in equation 3.17. This equation is the continuous wavelet transform, and in equation 3.18, the discrete form of the equation (with counters m and n) can be seen.

$$\tilde{f}_{m,n} = a^{-\frac{m}{2}} \int_{-\infty}^{\infty} f(t)\psi(a_0^{-m}t - nb_0)dt \qquad 3.18$$

The difference between the discrete and continuous form of the wavelet transform lies in the way the scaling and translating coefficients a and b change. In the continuous case, the parameters change continuously. This creates a huge set of data, often with unnecessary redundancy (the difference between two adjacent dilations or translations is likely to be virtually nil). Therefore, the discrete transform which allows us to do calculations only at a predetermined subset of dilations and positions is computationally more efficient.

3.5 Approximations and Details

In this section the method for the construction of an orthonormal basis of wavelets is introduced, as well as how this basis can be used in the multiresolution analysis of a signal. An orthonormal basis is chosen in order to avoid any redundancy due to the wavelet decomposition.

Firstly it is necessary to recognise what the wavelet transform is doing. The signal is taken and decomposed into a detail and an approximation, at a certain resolution. This means that the signal is being decomposed into the limit of the approximations at different resolutions (Dauberchies, 1988). The detail at each resolution is given by the difference between the approximation at the current resolution, and the approximation at the previous resolution (Mallat, 1989). This concept is illustrated in figure 3.2.

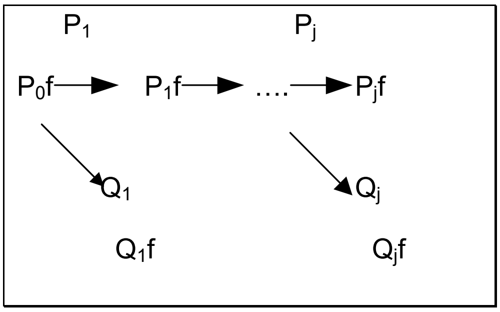

Figure 3.2: Decomposition tree (Strang, 1989)

In the decomposition tree shown in figure 3.2, let P_j be the operator that approximates the signal at a resolution *j*, while Q_j is the operator that gives the detail at that resolution. These operators have the following properties (Mallat, 1989):

- They are linear. Thus if the approximation is equal to P_1f, where *f* is the function denoting the signal, taking the approximation again at that same resolution will not alter the approximation. This means that the dot product $P_1 \bullet P_1 = P_1$.
- Of all the approximations that have been calculated, the approximation P_jf most closely resembles the signal function *f*, at the resolution *j*.
- The approximation of a signal at a coarser resolution contains all the information necessary to calculate the smaller/subsequent approximation.
- The approximation P_jf of a signal *f* is of the length 2^{-j}, when using the Haar wavelet (The Haar wavelet is the simplest family of mother wavelets. It is a second order orthonormal wavelet). However, if using a wavelet of higher

order, the size of the approximation decreases in size faster than a factor of 2.

- In computing the approximation of *f* at resolution *j*, some information about the function is lost.

In order to implement the multiresolution analysis, using the operators P and Q mentioned above, one first needs to construct the signal function, *f*, from the data sequence that is measured. This is done using equation 3.19.

$$\boxed{\begin{aligned} & f = \sum_n c_n^0 h_{0n}, \\ & or \\ & f(x) = \sum_n c_n^0 h(x-n) \end{aligned}} \quad \textbf{3. 19}$$

The operators P and Q can be implemented as simple filters. P is the low pass filter matrix, which yields the approximation of the original function *f*, while Q is the high pass filter matrix, giving the detail of the function (Strang, 1989)

4 Methodology

In the results section (next chapter), the techniques that have been introduced over the course of the last 2 chapters are further examined. In order to highlight advantages and shortfalls in the different techniques, extensive evaluation was performed on industrial data.

127 Control loops were chosen for evaluation. The loops represent a fairly large industrial production unit. Since the settling times of different types of control loops are dramatically different, ie flow and pressure loops tend to be tuned very quickly, while temperature loops are slower and could have significant derivative action involved, loops of all the major types were selected. The 127 loops were broken down into:

- 21 flow
- 62 level
- 34 pressure
- 10 temperature

The loops were all sampled for 24 hours, at a sampling rate of 5 seconds. 5 seconds was a practical consideration based on:

- DCS information and historian constraints
- Keeping the total data set manageable (total data set was 60Mb)
- High enough to capture significant dynamics

For each of the control loops, the PV, SP, OP and MODE of the loop was collected. This data was imported into a structured array in Matlab.

Each of the algorithms presented in the previous 2 chapters were further developed and coded into individual Matlab scripts. Since the main purpose of the dissertation is to identify and discuss results, not much attention was given to the development of the user interface with the scripts.

As mentioned, the data was structured within the Matlab environment. This made it possible to automatically differentiate between different loops, and between different loop vectors (namely PV, OP etc)

Pictorial results make up the bulk of the information presented, and were captured as static screenshots. Where tabular information was required, results were exported to Excel and manipulated before importing the static table into this document.

5 Results

Rather than concentrate on the diagnostic information of a single loop, several examples of indicative performance will be discussed. In this chapter, each of the techniques discussed in the previous sections will be investigated. This will be done by presenting the results generated by applying the techniques to several industrial examples.

5.1 Lower order statistics

Table 4.1 shows a typical example of an analysis of variance of 4 control loops. Data for each loop's measured or controlled variable (PV), set point (SP) and manipulated variable (OP) is shown.

Table 4- 1: Loop statistics

Loop Name		Average	Std Dev	Minimum	Maximum	Skewness	Kurtosis
F1002	PV	41.93	7.66	-1.54	48.54	-3.71	17.38
	SP	41.93	7.66	0.00	47.36	-3.72	17.47
	OP	75.36	10.24	0.00	87.37	-5.13	30.93
F1035	PV	3.00	0.11	2.65	3.17	-0.73	-0.32
	SP	3.00	0.00	3.00	3.00		
	OP	55.41	1.34	52.93	59.00	0.36	-0.64
P2107	PV	30.75	4.78	25.12	41.50	0.92	-0.68
	SP	30.75	4.79	25.11	41.58	0.92	-0.68
	OP	-6.86	0.00	-6.86	-6.86		
L2051B	PV	90.01	0.35	88.71	92.34	0.61	3.27
	SP	90.00	0.00	90.00	90.00		
	OP	19.30	4.87	-1.22	43.69	0.07	1.37

By analysing the total package of information, it is fairly easy to ascertain where it is necessary to expend additional effort in troubleshooting. F1002 has a relatively high standard deviation (close to 20%), and from the skewness and kurtosis data it is obvious that the loop is non-normal (a skewness of 0 and kurtosis of 3 represents a perfect Gaussian distribution). However, the extreme range over which the PV moved, as well as the range of the OP, suggests that there was a very large process upset somewhere in the data-analysed. This is confirmed in figure 4-1, which is a plot of the PV, SP, OP and error data for F1002.

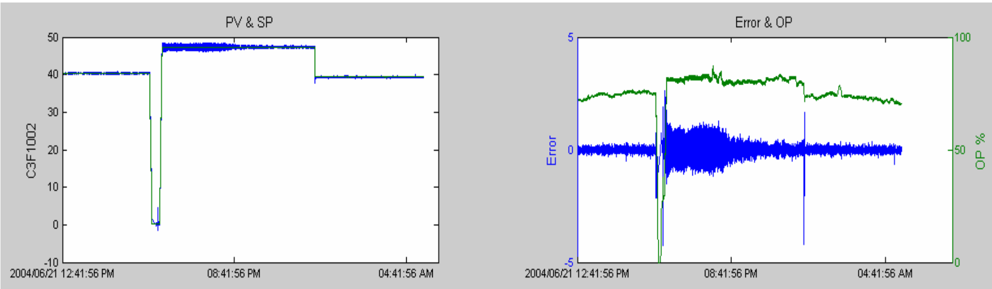

Figure 4- 1: F1002 data

Figure 4-1 also highlights one of the major flaws in using only statistical results for performance analysis. Apart form the upset which occurred, where the SP was moved to 0, there was one other SP move. Since the change in operating point dominates the statistics (standard deviation, skewness and kurtosis), it is impossible to draw a reliable conclusion from this information. By evaluating the OP information, specifically the average, one draws the conclusion that the final control element (in this case a valve), is operating fairly high, but is not really a case for concern. Other than investigating the reason for the process upset, one is unlikely to investigate the performance of this control loop based on the statistical data. One way of countering some of the miss-representation created by these changes in the operating region, is to evaluate the statistics of the error, ie the difference between the SP and the PV. Table 4-2 gives this information for loop F1002.

Table 4- 2: F1002 error statistics

Average	Std Dev	Minimum	Maximum	Skewness	Kurtosis
0.00	0.34	-4.81	2.64	-0.99	6.57

F1035 is an example of the statistics of a well performing loop. The standard deviation is extremely low (less than 5%). This is typical of a well performing flow loop. The information about the normality of the error would only be a concern should the variability in the loop increase. It is obvious from the results that there have been no changes made to the SP of this control loop.

P2107 is obviously not a normally performing pressure control loop. The variation of the OP is zero, and the analysis of the SP and PV is almost identical. This is one of the features present in most modern DCS/SCADA systems. In order to prevent an upset when a loop is switched from manual to auto, one of the options is for the SP to track the PV when the loop is in manual.

A further anomaly is the very low value of the OP. If the SP had remained constant, this may have been an indication of an over or under-pressure controller. In this case, this pressure controller is only used during startup, and remains closed on manual during normal operations.

L2051B is a condenser level controller. Since the duty, and thus the column's operation, is affected by variations, this controller has been configured to keep the level as close to set point as possible. The very low standard deviation suggests that this is exactly what has been achieved. The large variability in the OP is an indication of how hard the controller needs to work in order to achieve this tight control. The skewness and kurtosis information is a further indication of a well designed and configured controller, indicating almost perfect normality.

5.2 Minimum variance

The minimum variance performance, or the closed loop potential of a controller, can be a good indication of the performance of a control loop. However, often this measure can be a double edged sword, giving unnecessarily harsh indications of the controller performance. Table 4-3 shows the minimum variance measures for the control loops analysed in the previous section. Results are generated using a sample size of 17 000 data points, collected at 5 second intervals.

Table 4- 3: CLP for loops

Loop Name	CLP
F1002	0.27
F1035	0.79
P2107	0.03
L2051B	0.43

The performance of F1002 appears excellent from the statistical information, as well as from the visual inspection in figure 4-1. However, the CLP over the entire period shows that there is room for improvement. This highlights the very unforgiving nature of the minimum variance performance measures. Minimum variance is often not a requirement for excellent process control. On the other hand, the performance of F1035 is excellent, both from a statistical analysis and from the minimum variance performance anlaysis.

The extremely low score of P2107 is in line with expectation. The analysis technique used to determine the CLP apportions variability in the process to either the performance of the controller, or to the process. Since there is no control action being introduced in P2107, the variability in the output signal is totally unaffected by the controller.

Evaluating level loops using a minimum variance procedure should be undertaken with care. Many level loops are designed to buffer process disturbances, and thus are far from minimum variance. L2051B is designed to give tight level control, and this is reflected in the relatively high CLP.

From figure 4-1, it is clear that F1002 experienced a large change in operating region during this sample period. It is interesting to compare the performance during this upset, and this can be achieved by windowing the data into 1000 data point samples (thus there are 17 periods). Figure 4-2 shows the performance of F1002 over the entire period.

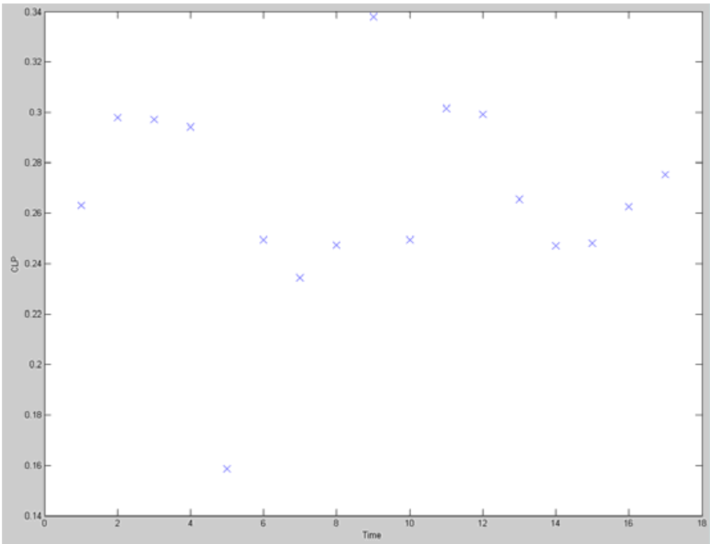

Figure 4- 2: Windowed CLP for F1002

In figure 4-2 it is clear that during the process disturbance, the CLP drops from the average around 0.3, down to 0.16. Unless rejecting SP changes of this magnitude is an important design consideration for this loop (in this case it is not), this performance decrease would not be a problem.

Figure 4-3 shows the same windowed approach applied to the F1035.

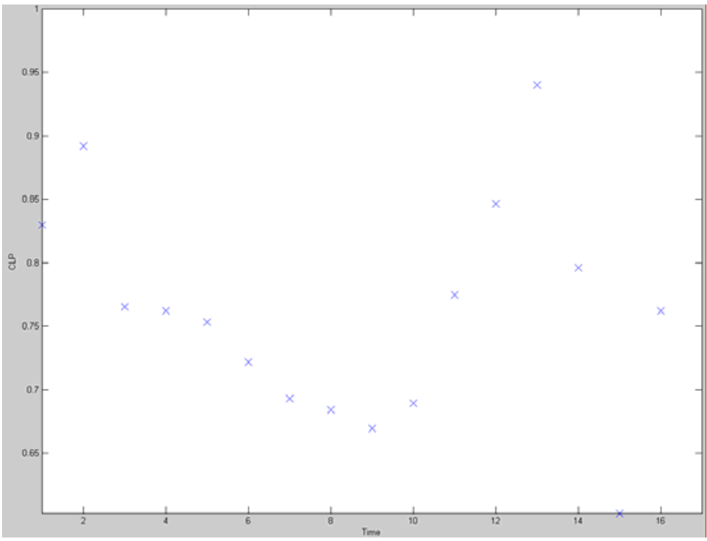

Figure 4- 3: Windowed CLP for F1035

F1035 has a much higher average CLP, and there is thus less room for improvement. However, by evaluating the CLP over time, it is obvious that there is an external cycle being introduced. F1035 tends to perform better during certain portions of the total sample. Thus, if the variability that was quantified in the previous section was deemed excessive, effort should be expended on attempting to remove this external negative influence.

One of the major draw-backs of minimum variance based performance measures is the requirement for knowledge of the dead time present in the loop. For the purposes of this analysis, the following dead times were assumed.

- Flow loops – 5 seconds
- Pressure loops – 20 seconds
- Level loops – 1 minute
- Temperature loops – 1 minute

The choice of dead times for the flow loops was limited by the collection frequency possible using the data collection method decided upon. Data was collected from the historian, and due to control and collection network bandwidth limitations, it was not possible to collect all the data used in the evaluation at an infinitely high collection frequency.

5.3 PV-OP pattern recognition

Plotting the process variable versus the process input is an easy way to visually troubleshoot problematic control loops. Specifically, the pattern gives insight into the linearity of the control relationship. If there are non-linearities, specifically due to stiction, these can be easily identified. Several examples have been presented to show different insights that can be gained from this analysis. The following control loops were analysed:

- F1035 (F1)
- F1024 (F2)
- F1081 (F3)
- L1064 (L1)
- L2051 (L2)
- P1067 (P1)

Figure 4-4 shows the PV versus OP plot for F1.

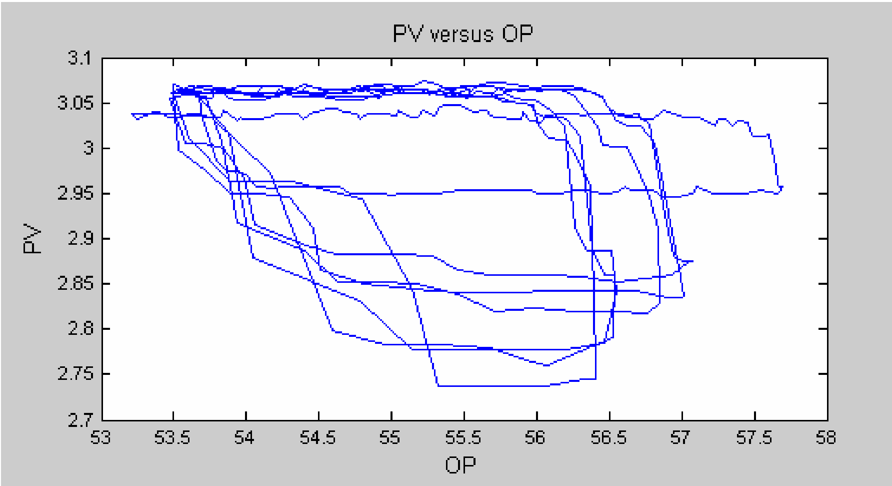

Figure 4- 4: PV-OP pattern for control loop F1

As will be shown in later examples, F1 has several irregularities in the period of data analysed. However, figure 4-4 is clipped to show the appropriate period where valve non-linearity dominates the output signal. The sharp patterns at the extremities, accompanying the valve reversals, show the tell tale signs of valve stiction. It would be suggested that before more effort is put into removing oscillations in this control loop, the valve be overhauled to correct the mechanical effects on the process.

Figure 4-5 shows the PV versus OP plot for F2.

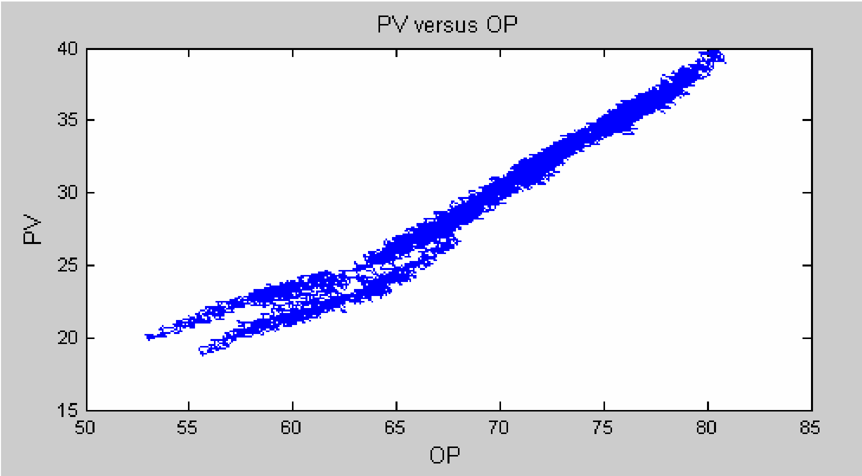

Figure 4- 5: PV-OP pattern for control loop F2

The pattern shown in figure 4-5 is unlikely to cause further investigation. F2 has a broad operating region. This is caused by a cascade master that is continuously shifting the setpoint for this control loop. However, despite the broad operating region, F2 reacts linearly and predictably. There is no indication that there is valve problems, or other non-linearities in this control loop.

Figure 4-6 shows a combined process plot and PV-OP pattern for F3.

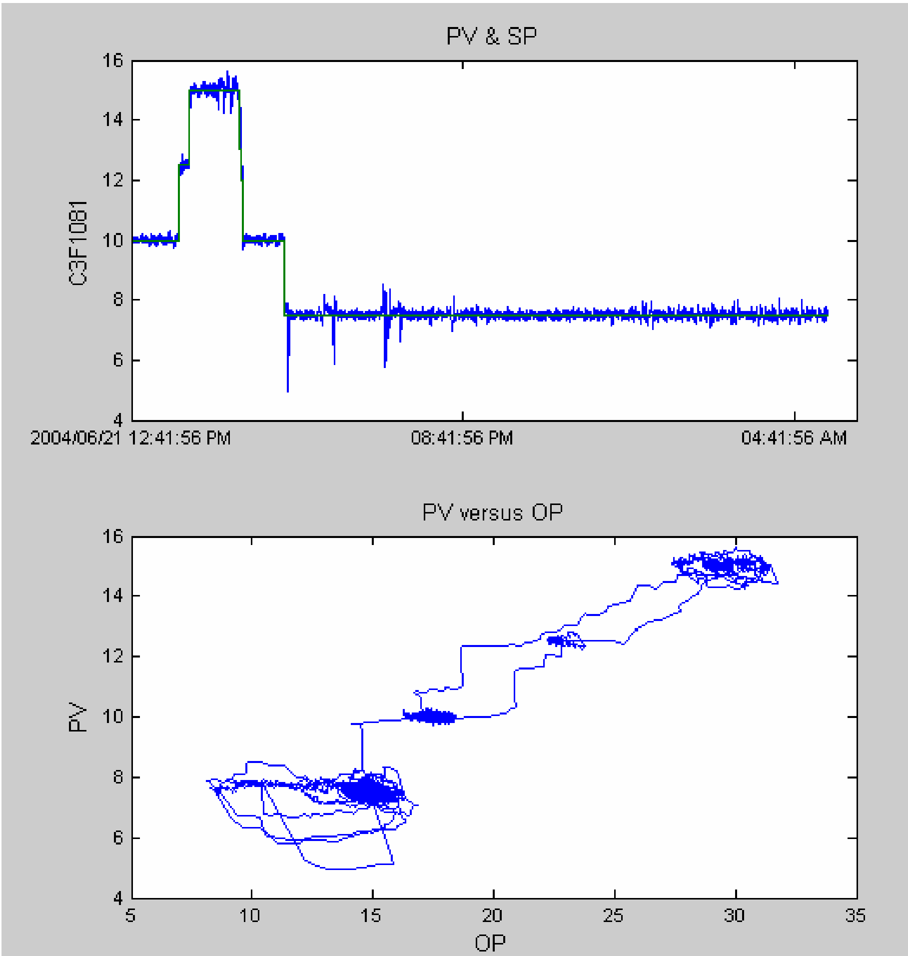

Figure 4- 6: General information for control loop F3

From figure 4-6, it is clear that there are several different operating "zones" in which F3 operated over this time period. In order to better analyse the specific non-linearity that may be present in the loop, the longest operating period (around 15% valve output) was further investigated. This information is shown in figure 4-7.

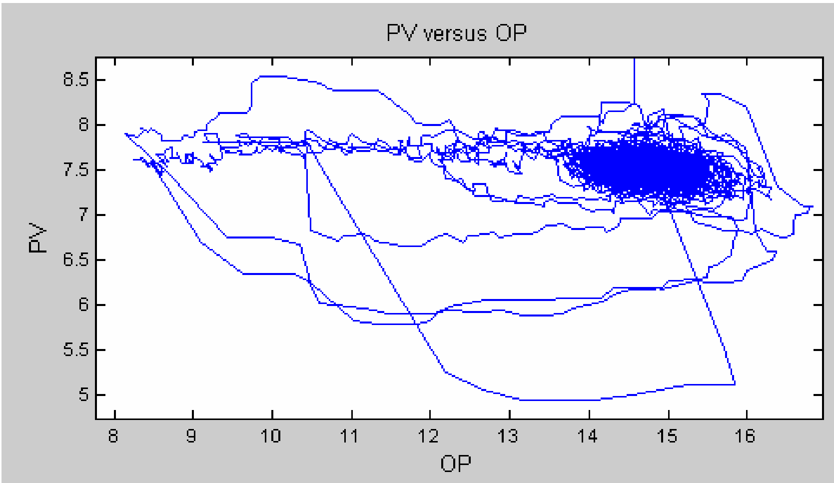

Figure 4- 7: Zoomed PV-OP pattern for control loop F3

While there is some indication of sharp edges at the extremities of the operating region, the information presented is inconclusive. Based on this information alone, it is unlikely that the control engineer would advise a valve replacement.

Figure 4-8 shows the PV versus OP plot for L1.

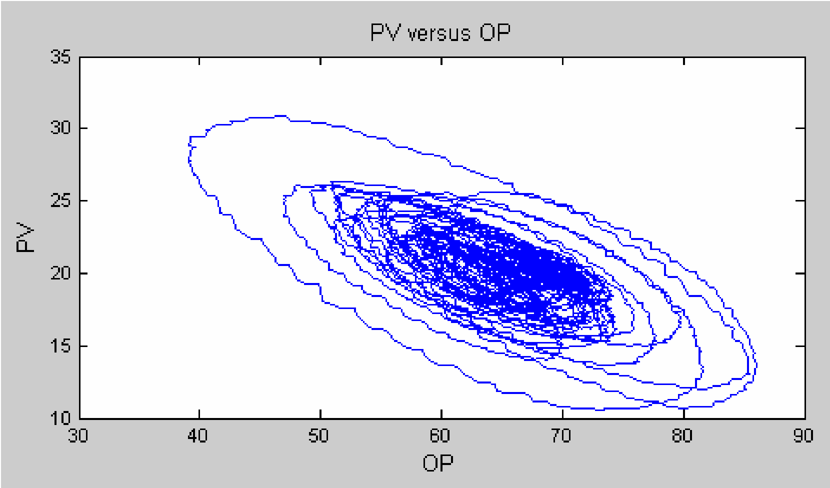

Figure 4- 8: PV-OP pattern for control loop L1

As a general rule, level loops are much harder to identify patterns indicating valve related non-linearities. Even though figure 4-8 does show a distributed pattern, the edges of the control region are regular and smooth. Thus based on figure 4-8, no action would be recommended.

Figure 4-9 shows the PV versus OP for L2.

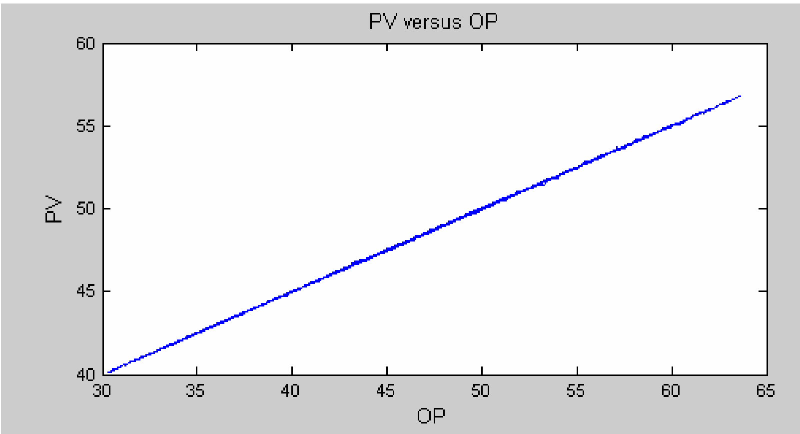

Figure 4- 9: PV-OP pattern for control loop L2

From figure 4-9 it is clear that there is an uncannily linear relationship between PV and OP. In this case, the tuning method that has been applied to L2 is the actual cause, rather than a perfect valve. There is no true error feedback into this control loop. Rather, an explicit PV-OP mapping has been used to implement the control. This is in effect a normal proportion only controller. Since there is no error based feedback, and the PV determines the OP explicitly on each execution cycle, figure 4-9 actually shows no useful information about the controller performance.

Figure 4-10 shows the PV versus OP plot for P1.

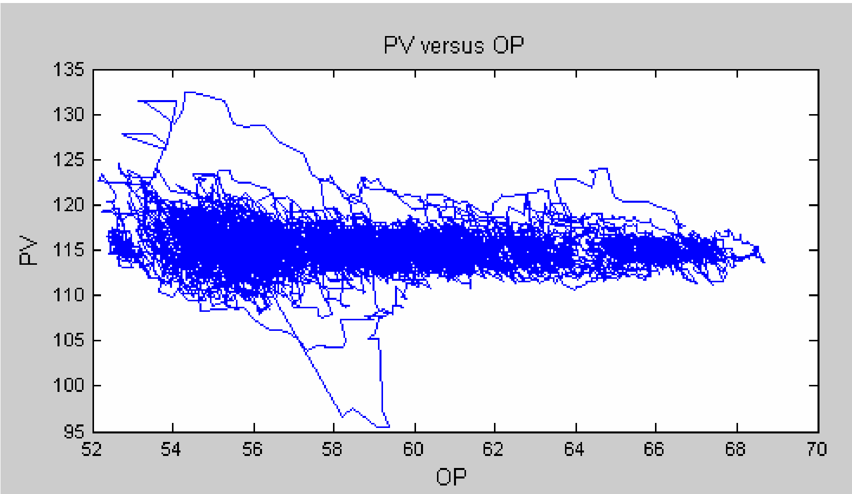

Figure 4- 10: PV-OP pattern for control loop P1

This pattern tells us nothing of the non-linearity in the control loop. However, notice the broad range on the OP, with the almost negligible move in the PV. P1 is an example where the control valve has been sized to minimise the pressure drop, and in so doing, has left almost no control. The only thing really keeping this control valve close to setpoint is the stable operations up and down stream. This control loop would be incapable of handling major process upsets.

5.4 Histograms and normality

Histograms are a graphical method for analysing how linear the performance of the control algorithm is. PID is obviously a linear control technique, but since the process and the final control element may not be, the response to linear outputs may be poorly suited to controlling the system well. Once again, 3 different control loops are chosen to highlight the interpretation and use of histograms.

- F2081
- F2024
- F1038

Figure 4-11 shows the histogram of a well performing control loop, F2081.

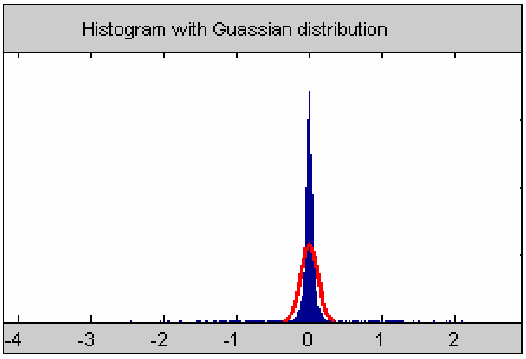

Figure 4- 11: Histogram of errors for F2081

It is clear that the error is normally distributed around zero. The average error is zero (5 significant figures). It is also clear that the error is tightly distributed around the mean, graphically indicating a low standard deviation.

Using the approach outlined in earlier sections, it is possible to determine the degree of normality. Using the equation, $\frac{|\mu_{data} - \hat{\mu}|}{\sigma_{data}} > 10\%$, one finds the degree of normality to be unacceptably high (21.68%).

Since the spread of error is even more tightly spread around the mean than would be expected from the Gaussian curve (see figure 4-11), one is unlikely to give this loop much attention just based on the normality investigation.

Figure 4-12shows the error distribution and gaussian curve for controller F1038.

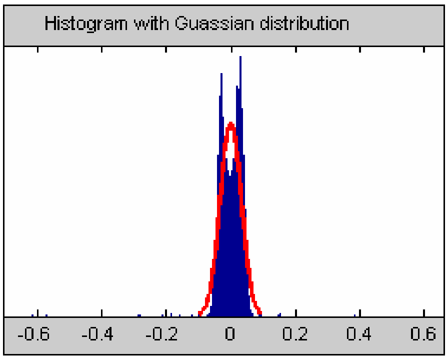

Figure 4- 12: Histogram of errors for F1038

It is immediately clear that there is a major non-linearity present in this control loop. The pattern present in figure 4-12 is typically caused by a slip-stick cycle in the control valve. The error is distributed around a zero error mean, with 2 excessive peaks either side of the mean.

Using the same equation presented above, the non-normality of F1038 is lower than F2081, but still significant (14.69%).

The approach that would be taken when encountering a problem with non-linearity, such as F1038, differs depending mainly on the following factors:

- Actual variability resulting from non-linearity or slip-stick cycle. If the influence of the variability is insignificant, no action will be taken.
- Criticality of the control loop. Control loops with significant slip-stick may have a rapid decay of control performance over the short term. This diagnosis would definitely feed into the work scheduled for the next plant maintenance cycle. In the case of extremely sensitive/critical control valves, operations may decide to schedule maintenance more aggressively to prevent possible production, quality or safety incidents.
- Ease of repair. Critical control loops may have by-passes installed that allow the control valve to be overhauled without requiring production loss. This is the ideal scenario for the maintenance personnel responsible for the performance of this loop.

Figure 4-13 shows the error distribution and gaussian curve for controller F2024.

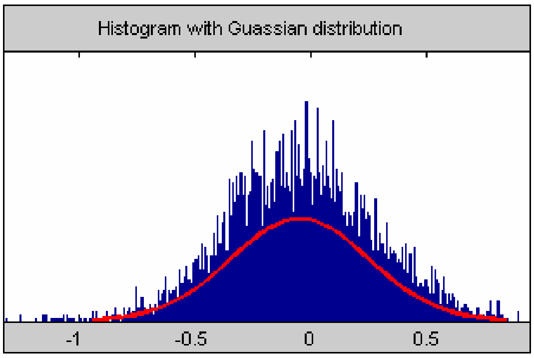

Figure 4- 13: Histogram of errors for F2024

F2024 would not be flagged attention based on the non-linearity result of this test (1.7%). The broad distribution gives an indication of a relatively high standard deviation. Table 4-4 gives the performance statistics of F2024, and immediately places everything in perspective.

Table 4- 4: Performance statistics of F2024

Loop Name	CLP	Average	Std Dev
F2024	0.7	13.62	0.83

While the standard deviation is relatively high for a flow controller (6% of mean), the CLP is high. This indicates there may be little room for improvement in this control loop.

5.5 Cross Correlation

As was the case with histograms, cross correlations are a technique for highlighting non-linearity in a control loop. The technique specifically identifies patterns prevalent in the closed loop data of loops suffering slip-stick cycle. 5 control loops will be used to highlight the usefulness of cross correlation analysis. The following control loops were chosen:

- L1030
- F1021
- F1038
- F2024
- L1043

ΔT and ΔQ tests were introduced in Chapter 2. These tests were applied to each of the loops investigated. Table 4-5 shows these results.

Table 4- 5: Selected test results

Loop Name	ΔT	ΔQ	ΔT test	ΔQ test
L1030	0.02	0.00	No Stiction	No Stiction
F1021	0.00	0.00	No Stiction	No Stiction
F1035	0.02	0.23	No Stiction	No Decision
F2024	0.03	0.53	No Stiction	Stiction
L1043	0.40	0.37	No Decision	Stiction

From all the loops analysed, the ΔQ test is much more critical of loop non-linearities. Of the 127 loops analysed, 16 loops were determined to have stiction related non-linearity. Of these, the ΔQ test failed to identify 4, while the ΔT test failed to identify 8. As is highlighted in the following detailed discussions, the flag raised by the tests is only an indication. The test should be viewed in conjunction with other loop analyses.

The cross correlation curves were obtained using 5 second process data (roughly 17 000 data points). Each curve has been zoomed into to show a representative portion of the correlation information. In each case, the x-axis represents the number of positive/negative lags, and the y-axis the magnitude of the correlation coefficient.

Figure 4-14 shows the information for control loop L1030.

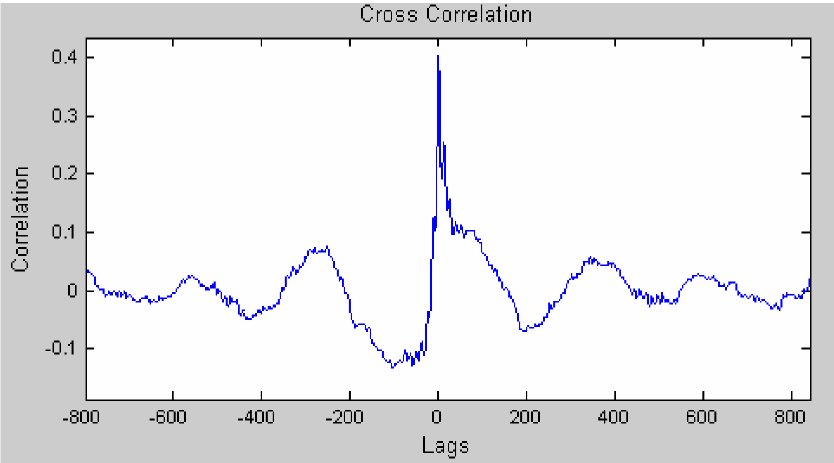

Figure 4- 14: Cross correlation for L1030

The cross correlation information for L1030 is an almost perfectly even function. Both tests confirm that the loop has negligible non-linearity. Figure 4-14 does indicate that the cross correlation coefficient does not decay immediately beyond the initial highly correlated information. There appears to be a persistent swing in the correlation information that only gradually disappears. By investigating the process data for L1030, shown in figure 4-15, the persistent oscillation is easily observed.

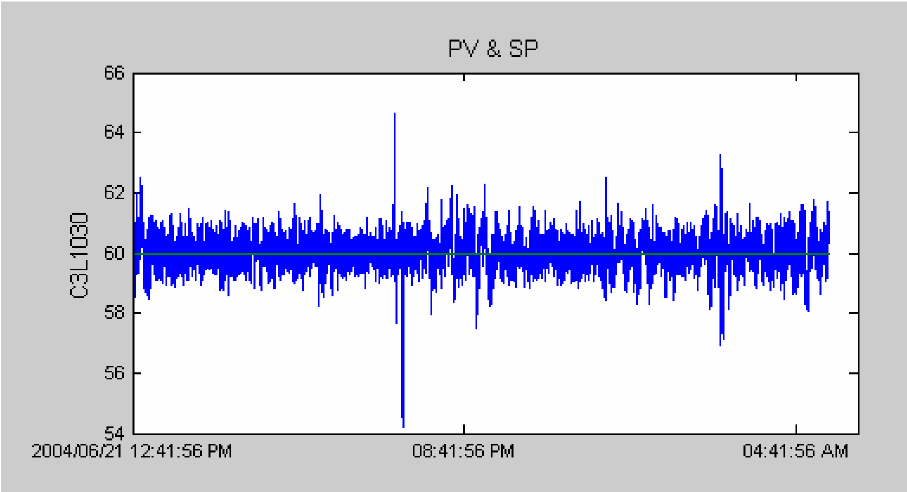

Figure 4- 15: PV versus SP for L1030

The perfectly even nature of this oscillation suggests that this loop is too aggressively tuned.

Figure 4-16 shows the correlation information for F1021.

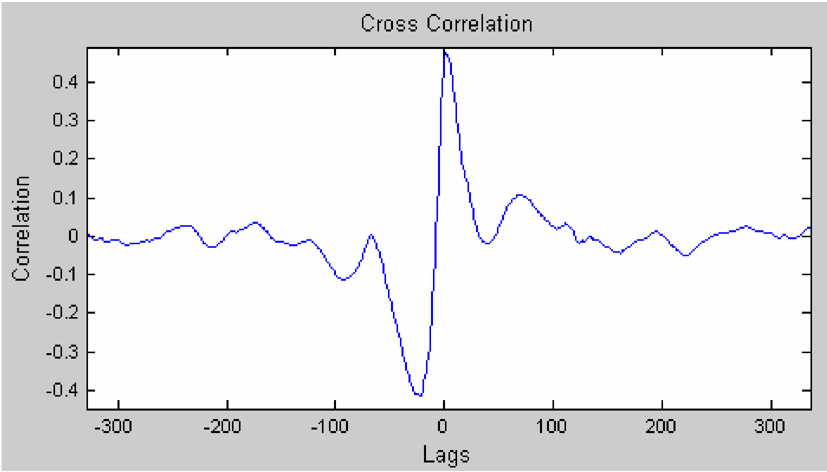

Figure 4- 16: Cross correlation curve for F1021

From this figure the even nature of the oscillation interpreted by the *ΔQ* and *ΔT* tests can be confirmed. In contrast to L1030, the oscillations for F1021 decays rapidly either side of zero lag. The lack of oscillatory information in this loop is confirmed by evaluating the statistical and time plot information (not shown).

Figure 4-17 shows the cross correlation information for F1038.

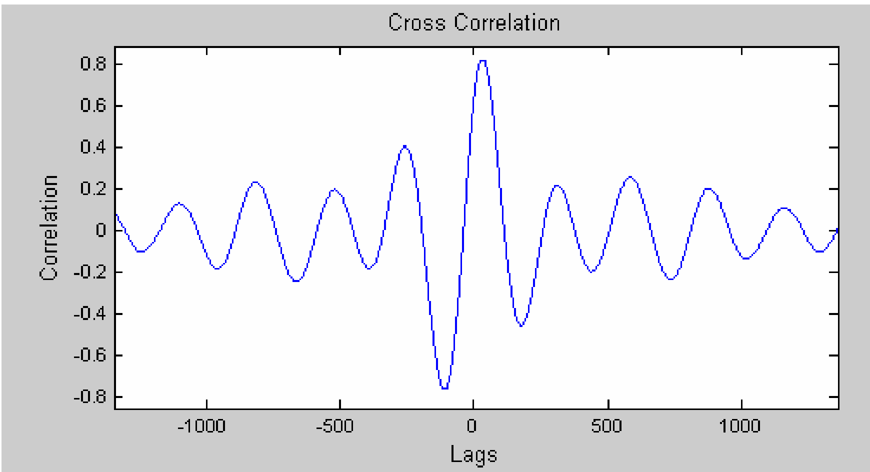

Figure 4- 17: Cross correlation curve for F1038

From table 4-5 it is clear that while the ΔQ and ΔT tests do not point directly to a process non-linearity, there is some doubt in the interpretation of the ΔQ test information. In figure 4-12 the histogram of F1038's error was shown. This indicated a high probability of valve induced non-linearity, based on the pattern/distribution of the error around the mean.

This highlights the danger in using limited diagnostic information when attempting to isolate the cause of process variability. Loop F1038 requires further investigation in order to determine the cause of the oscillation.

Figure 4-18 shows the cross correlation information for loop F2024.

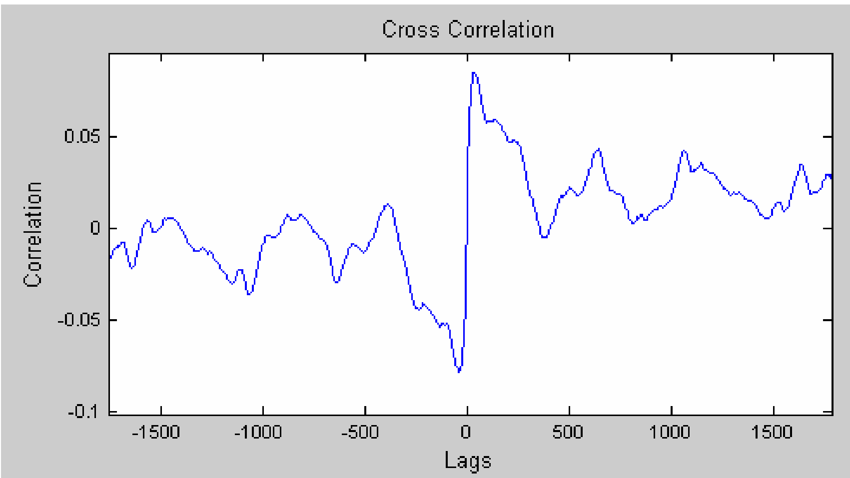

Figure 4- 18: Cross correlation curve for F2024

This loop is an example of conflicting information between the two tests. Based on this test method, it is not possible to draw a confident conclusion that F2024 has stiction induced oscillation. Using the histogram method, there is no indication of non-linearity either (not shown).

Figure 4-19 shows the cross correlation curve for the final loop, L1043.

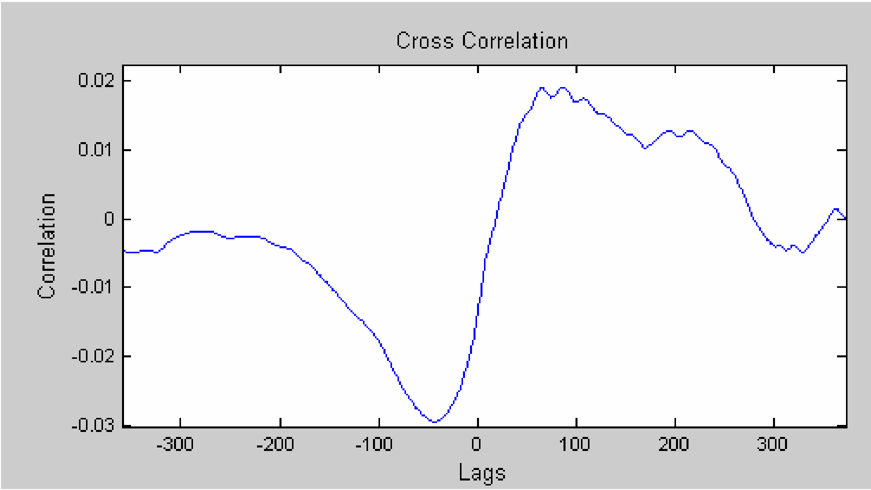

Figure 4- 19: Cross correlation curve for L1043

This loop was specifically chosen to highlight another one of the pitfalls of evaluating mathematical interpretations without human intelligence to prevent incorrect conclusions from being drawn. L1043 is indicated as having a non-linearity by the ΔQ test, and is a suspect in the ΔT test. Both indicate there is a possible valve induced non-linearity present.

Figure 4-20 shows the Gaussian distribution of the error around the mean.

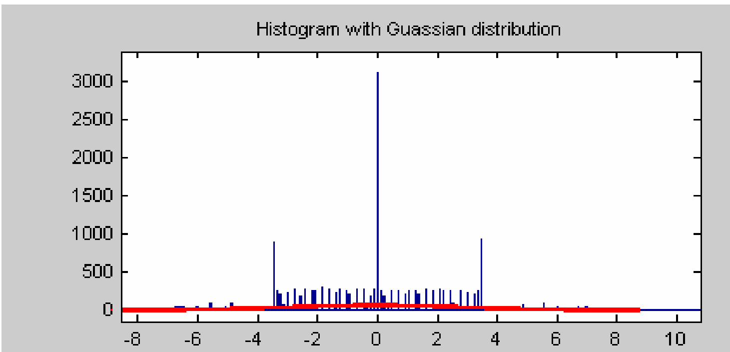

Figure 4- 20: Gaussian distribution of L1043

This distribution graph also shows the presence of an unusual distribution of error around the mean.

In the case of L1043, a unique implementation of a proportional only controller has been utilised. At every execution, the SP is written equal to the PV, and an externally normalised OP – PV mapping is used to control the level. This has produced the unusual distribution seen in figure 4-20, and interpreted by the cross-correlation analysis as a valve induced non-linearity.

5.6 Frequency analysis

5.6.1 Autocorrelation analysis

While autocorrelation analysis does not really reflect a frequency based technique (there is no transformation of the time data into the frequency domain), the natural ability to generate information about the response patterns and dominant oscillations in the data, means that this grouping makes sense.

One of the beauties of using the autocorrelation analysis technique is the intuitive interpretation of the visual data generated. The following 3 control loops were evaluated:

- F2002
- L1025
- P1001

Intuitively, a control engineer wishes to see a control action forcing zero error after a set period of time. In the case of flow loops, this is often as quickly as possible. Autocorrelation analysis is an easy way of visualising how quickly a control loop is forced to zero error.

The autocorrelation of F2002 is shown in figure 4-21.

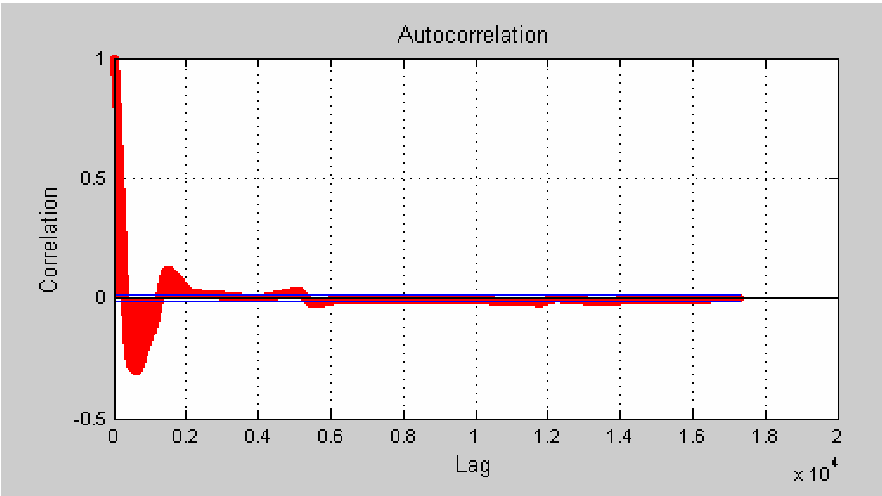

Figure 4- 21: Autocorrelation of F2002

From this analysis it is clear that F2002 has been tuned to aggressively remove any disturbances that enter the system. However, due to the aggressive nature of the tuning, there is over shoot and gradual decay of error over time. Figure 4-21 is a good example of quarter amplitude damping (Ziegler Nicols tuning), which is often too aggressive on the initial proportional action.

L1025 has a totally different problem that can be diagnosed using the autocorrelation approach. The autocorrelation function is plotted in figure 4-22.

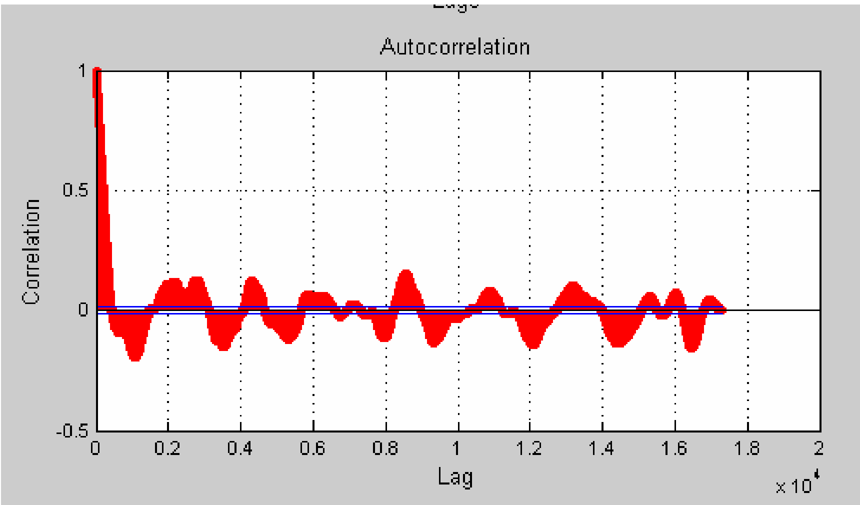

Figure 4- 22: Autocorrelation of L1025

L1025 has a regular periodic error. Industrial processes, specifically disturbances to those processes, are very rarely regular. Normally regular patterns are introduced through incorrect control strategies, either in the loops being analysed, or upstream from that control loop. The correct approach to tuning level loops is the topic of several papers. Often the tried and tested method of simple proportion only control tuning is rejected in order to ensure some SP tracking. While this is often necessary, it can lead to the over-aggressive integral tuning seen in L1025. A rule of thumb solution to solving this error would be to increase the integral time up to the period of the oscillation. The initial rise time (time taken to initially eliminate error) is more than acceptable. If less aggressive action can be tolerated, the proportional action could also be decreased.

Figure 4-23 shows the autocorrelation of P1001.

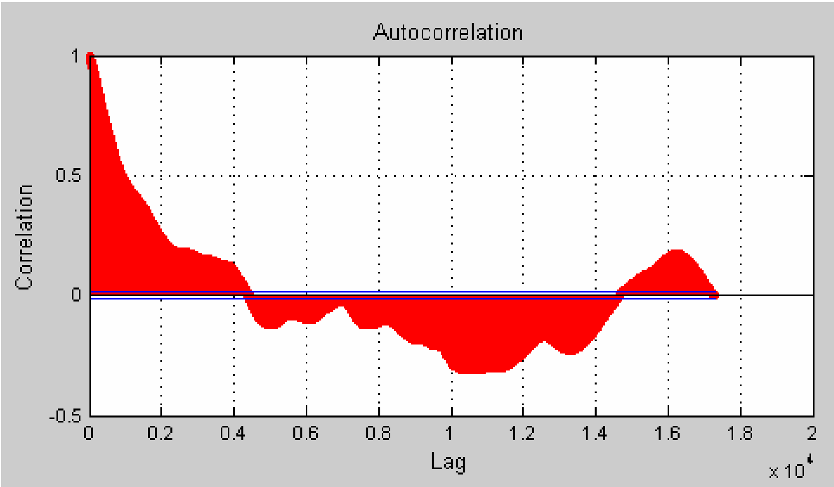

Figure 4- 23: Autocorrelation of P1001

The autocorrelation of P1001 shows an extremely slow settling time (the x-axis is in multiples of sample size, and in this case the rise time is roughly 6 hours). This could be due to extremely ineffectual control action (CLP = 0.09). This also highlights the need to evaluate pieces of control diagnostics as a package. Figure 4-24 shows the time plot for P1001.

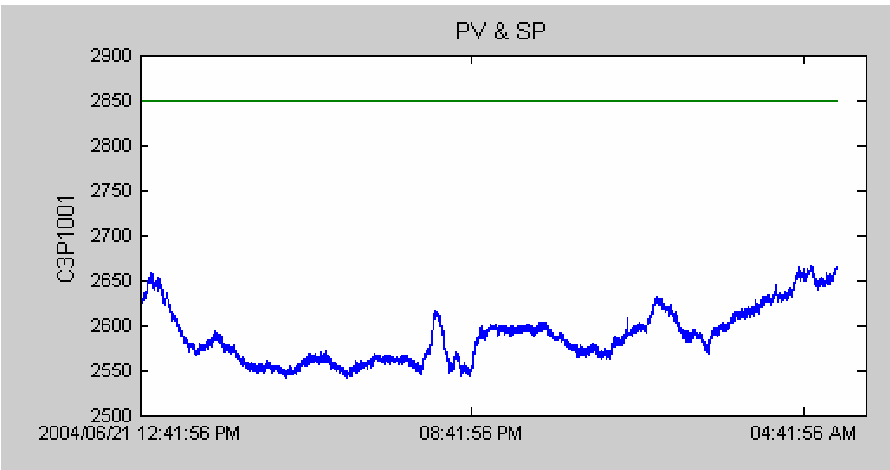

Figure 4- 24: Timed data for P1001

It is clear that P1001 PV never reaches the SP. The reason for this is that P1001 is an over-pressure controller. The OP is also zero over the entire period (not shown). Thus the diagnosis of P1001 could more easily be achieved by simply evaluating the figure 4-24.

5.6.2 Power spectrum analysis

The power spectrum analysis is a useful technique to determine if there is significantly more power at a certain frequency region than at others. Common peaks in the power spectrum of different control loops may indicate a common cause. In interpreting the following results, note the choice of seconds as the x-axis, rather than Hz. The following 3 control loops have been chosen to indicate the interpretation of the power spectrum:

- F1020
- F1035
- L2036

Figure 4-25 shows the power spectrum for control loop F1020.

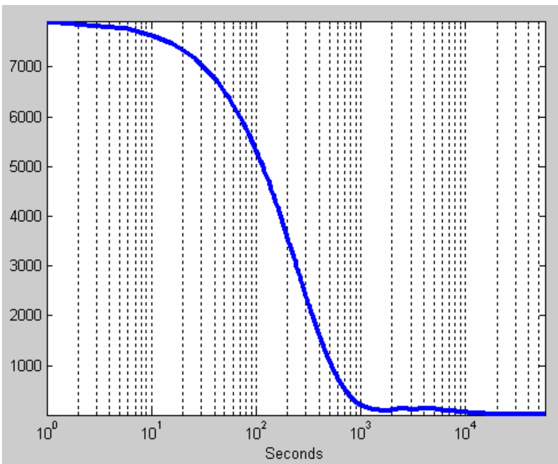

Figure 4- 25: Power spectrum analysis for control loop F1020

From this analysis it is clear there is a steady decay of power in this signal from high to low frequency. There is negligible low frequency information, showing that this control loop is effective in removing long term oscillations from this loop. The power spectrum analysis does not add any additional value in troubleshooting this loop, other than to confirm that there is no excursions in the frequency domain.

Figure 4-26 shows the power spectrum of F1035.

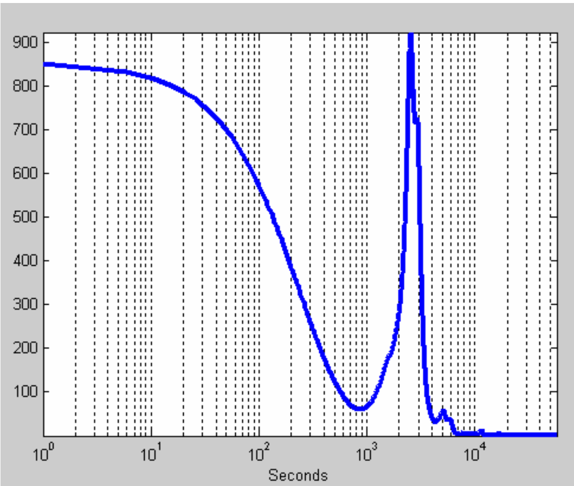

Figure 4- 26: Power spectrum analysis for control loop F1035

The left hand portion of figure 4-26 is very similar to that shown for F1020 in figure 4-25. However, it is clear there is a significant disturbance at lower frequencies. In this case, the frequency information is at 43 minutes.

The PID tuning constants typically used for flow control are designed to remove high frequency information, and are not well suited to removing longer sustained oscillations. Further analysis of F1035 showed extremely aggressive integral tuning. As mentioned earlier, integral action can be used to damp lower frequency oscillations. Slowing down the integral action in F1035 would be a sensible next step, based on the power spectrum analysis.

Figure 4-27 shows the power spectrum for L2026.

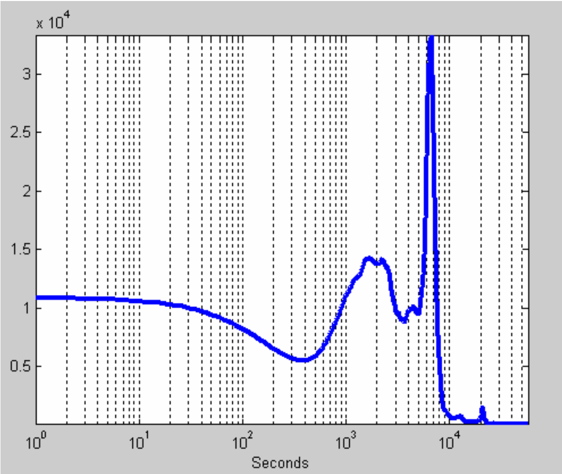

Figure 4- 27: Power spectrum analysis for control loop L2026

As with power spectra shown in figure 4-25 and figure 4-26, the high frequency portion decays fairly rapidly. However, L2026 has an interesting double peak in the frequency information. The first peak occurs at 26 minutes. Level loops can be designed to assist in removing higher frequency oscillations, and there-by stabilising down-stream processes. 26 minute oscillations would be long enough to be difficult to remove with normal, quick PID tuning constants (typical in most flow and pressure loops). However, by allowing the level to move, it is possible to allow the liquid capacity to damp these oscillations.

The second peak is at roughly 2 hours (133 minutes). Sustained oscillations of such a low frequency are extremely difficult to remove. The residence time in most drums (L2026 is an example), do not have the capacity to remove oscillations at such low frequency (obviously depending on how violent the oscillation is). Addressing such low frequency oscillation would need to be done in the upstream culprit.

The most prevalent low frequency disturbance in chemical industries is the day night swing. While none of the control loops analysed in this section showed signs of a 12 hour swing, temperature and pressure loops are susceptible to this swing.

5.6.3 Common oscillation detection

When monitoring hundreds, or even thousands of different control loops, it is impossible to pickup loop interactions without some assistance. Using this technique of common oscillation detection, it is possible to graphically view where groupings of control loops appear to have common behaviour.

Figure 4-28 shows all the loops available in the data set used for the purpose of this report. 127 loops in total are represented. Each axis represents one principle component.

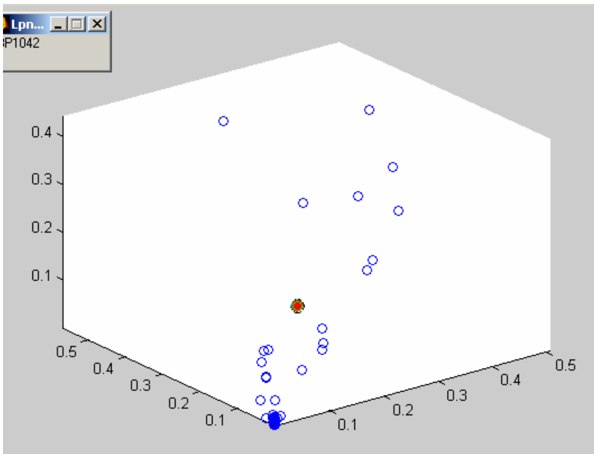

Figure 4- 28: Three dimensional common oscillation plot

In figure 4-28, pressure control loop P1042 is highlighted. Unfortunately the grouping and identification process is manual. Another problem with the above graphic is that since the dominant frequencies are so different between the different loops, the resolution is poor. In figure 4-28 the majority of visible control loops are pressures and levels.

By grouping loops by control type, ie flow, pressure, temperature and level it is possible to get more resolution. This is done in figure 4-29.

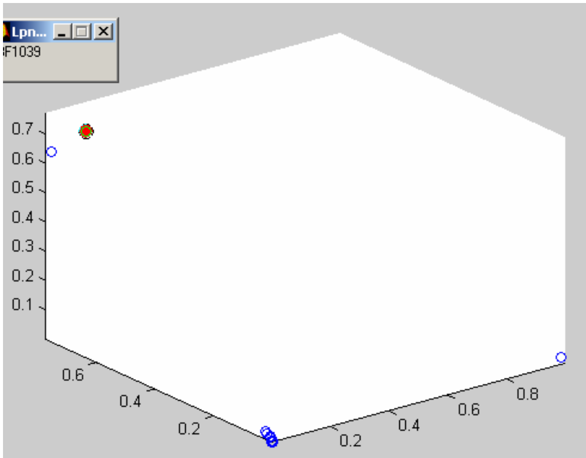

Figure 4- 29: Flow loops three dimensional common oscillation detection

The selected flow loop has an obvious common pattern with the loop just left in the diagram, which is the identical flow loop on the sister plant. Since both have identical tuning, and the plant has a common feed, this is not unexpected. Once again there is a cluster of loops – zooming into these yields figure 4-30.

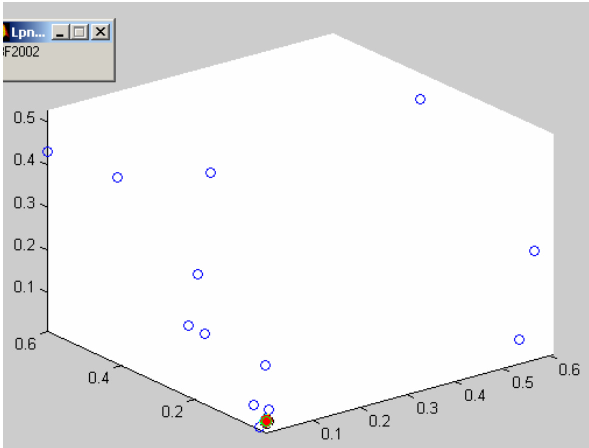

Figure 4- 30: Zoomed three dimentional analysis

Figure 4-30 yields several useful groupings. F2020, F2021 and F2002 are all related flow loops on the plant, and are tightly clustered. The relationship between F1035 and F1081 is also clear.

5.6.4 Bicoherence

Bochoherence analysis is a high order frequency based technique which is useful for highlighting instances of process non-linearity. By evaluating the correlation between offset frequency decompositions of the error signal, it is possible to determine where areas of high interaction lie. If the control loop is performing perfectly there should be no interaction between different frequencies, and any off diagonal correlation would be a sign of some form of process non-linearity. The following control loops are analysed:

- F1021
- F1035

Figure 4-31 shows the bicoherence contour for F1021.

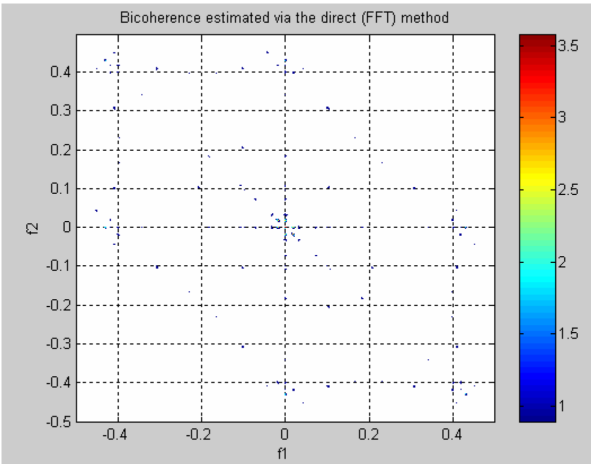

Figure 4- 31: Bicoherence for control loop F1021

It is clear from this figure that there is almost no information contained by evaluating the interactions in the frequency domain. Using other analyses it is possible to show that F1021 is an excellently performing flow loop, with no signs of stiction, and close to minimum variance performance.

Figure 4-32 shows the bicoherence contour map for F1035.

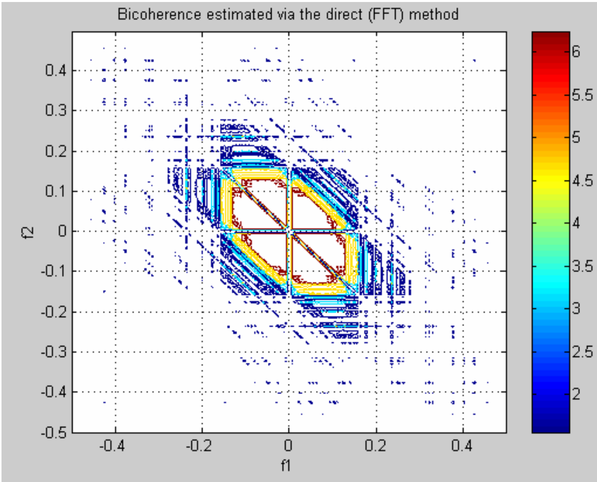

Figure 4- 32: Bicoherence for control loop F1035

From previous analysis (as well as that contained in the section dealing with Wavelet analysis) it is clear F1035 has unusual patterns. The bicoherence shows clear indications of non-linearity in the process. Unfortunately, without further or supporting analysis, there is no way of determining the source of this non-linearity.

5.6.5 Wavelets

While it is possible to dedicate much more to the use of wavelets, only a few examples will be presented. Since the use of the wavelet is not dissimilar to other frequency analysis techniques, it makes sense to group this discussion along with the other frequency techniques.

Only 1 loop is analysed, namely F1035. This loop was chosen due to clear frequency and time based information. Figure 4-34 shows the normal trend of process information for F1035.

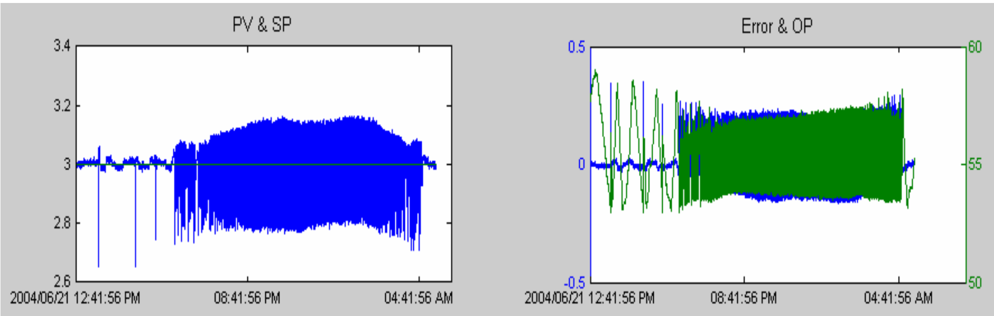

Figure 4- 33: Time information for control loop F1035: PV, SP and OP information

From figure 4-34 it is clear that there was a major change around sunset. The normal frequency based techniques cannot isolate an event such as the one presented above. When analysing the power spectrum, shown in figure 4-26, it was clear that there was high frequency content, but it was not discrete. The lower frequency oscillation, shown on the left hand side of both graphs in figure 4-33, was identified.

Dealing with localised frequency domain phenomenon is exactly what wavelet analysis is designed to handle.

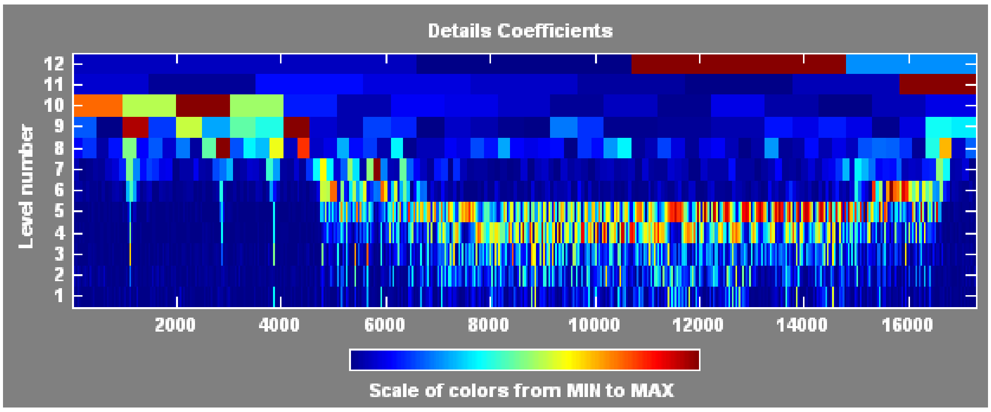

Figure 4- 34: Full wavelet decomposition for the F1035 error signal

The wavelet decomposition shown in figure 4-35 was compiled using a Daubechies 4th order mother wavelet. The method used is exactly the one introduced earlier. Each level is a subsequent detail obtained from the decomposition. As such, level 1 shows the 'highest frequency' information, and the colours show the level of 'information' contained at that time, for that frequency.

From figure 4-35 it is clear that the first portion of the time window analysed contains mainly information in the level 8, 9 and 10 region. In order to show more clearly, only level 10 is used. Figure 4-36 shows the approximation and detail for level 10 versus the entire time span.

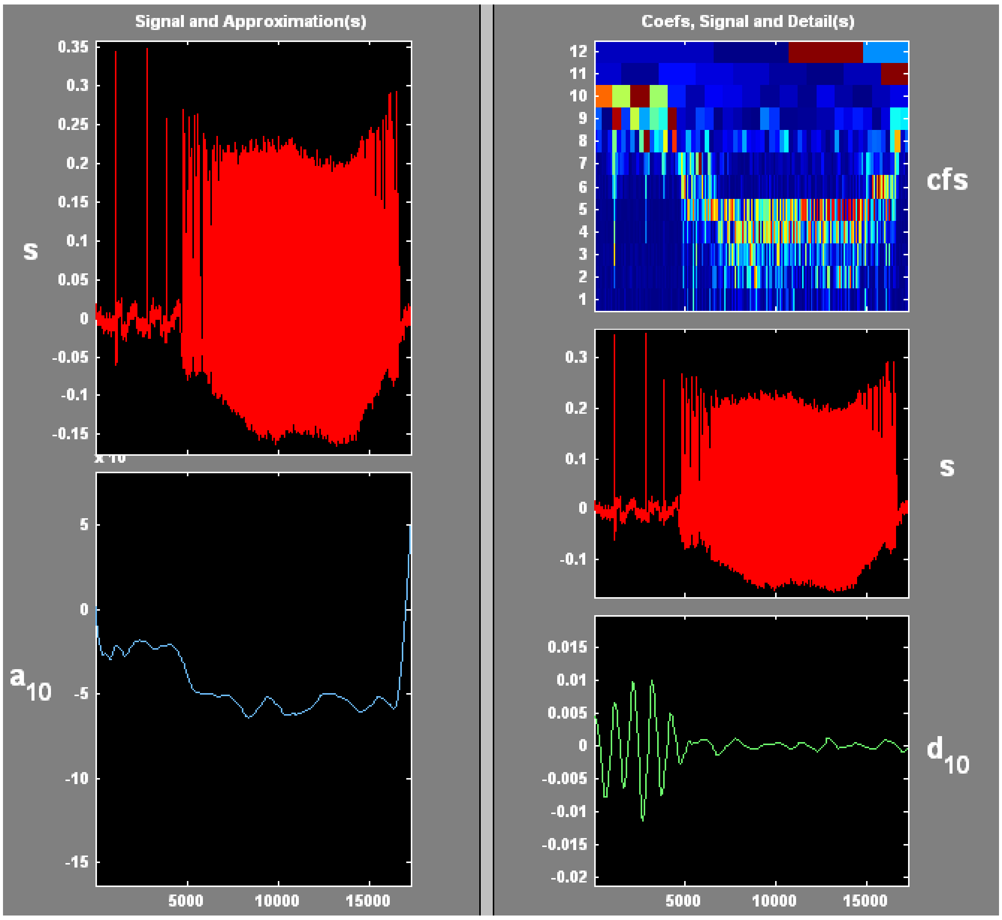

Figure 4- 35: Level 10 of control loop F1035 wavelet decomposition

From figure 4-36 it is clear that the information content, especially contained in the detail (d_{10}) is richest over the first part of the time series. Notice also that the approximation at level 10 (a_{10}) which has had the detail present in level 9's approximation removed, contains almost no information.

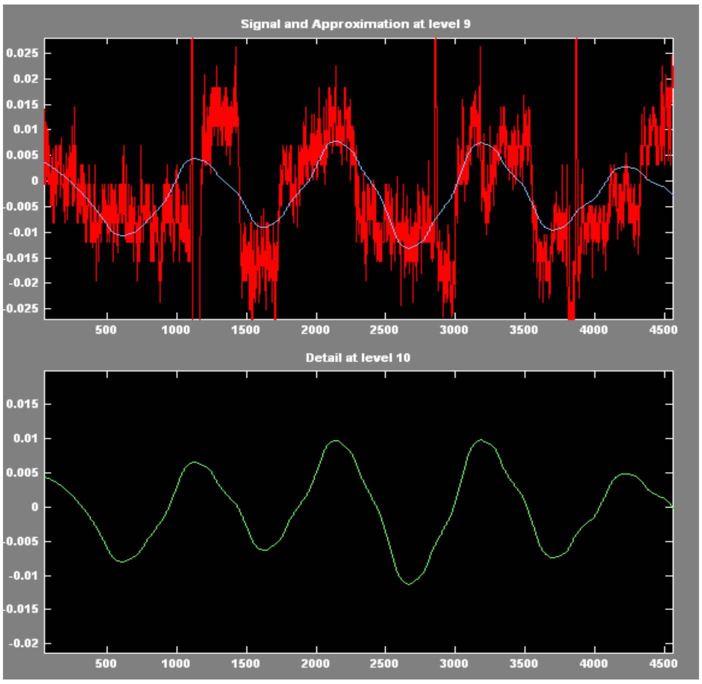

Figure 4- 36: Signal estimation of F1035 using detail at level 10 of the wavelet decomposition

Figure 4-37 compares this lower frequency information in the level 9 approximation and level 10 detail, with the original signal.

From figure 4-37 it is possible to identify that the information captured over this period represents the majority of the information contained in the signal.

It is possible to find the relationship between the level and the frequency. It must be stressed that this is an approximation. The results were generated, and are shown in table 4-6.

Table 4- 6: Pseudo-frequencies translation for wavelet decompositions for Daubechies 4th order mother wavelet

Level	Centre frequency (Hz)	Centre frequency (min)
1	0.0714	0.23
2	0.0357	0.47
3	0.0179	0.93
4	0.0089	1.87
5	0.0045	3.73
6	0.0022	7.47
7	0.0011	14.93
8	0.0006	29.87
9	0.0003	59.73
10	0.0001	119.47
11	0.0001	238.93
12	0.0000	477.87

In figure 4-26, F1035 was identified to have a dominant oscillation of 43 minutes. While the resolution using the wavelet analysis (in terms of exactly identifying a precise frequency) is not very high, this analysis does show that during the initial stages of the time frame analysed there was indeed a low frequency disturbance. Using the Fourier transform approach of power spectral analysis, the absence of this low frequency information during the later parts of the signal, blur the results. From the wavelet analysis the true oscillation period appears slower – between 60 and 120 minutes (ie between approximations level 9 and level 10) rather than 43. This is confirmed by visual inspection of the signal (4 swings in 4500 samples, 5 seconds per sample) giving a period of oscillation of 94 minutes.

The same analysis can then be performed to identify in more detail the cause of oscillation in the centre portion of the time signal in figure 4-34. From figure 4-35 it is clear that the dominant information is in levels 4 and 5. The view shown in figure 4-36 is repeated in figure 4-38, with focus on the approximation and details relevant for levels 4 and 5.

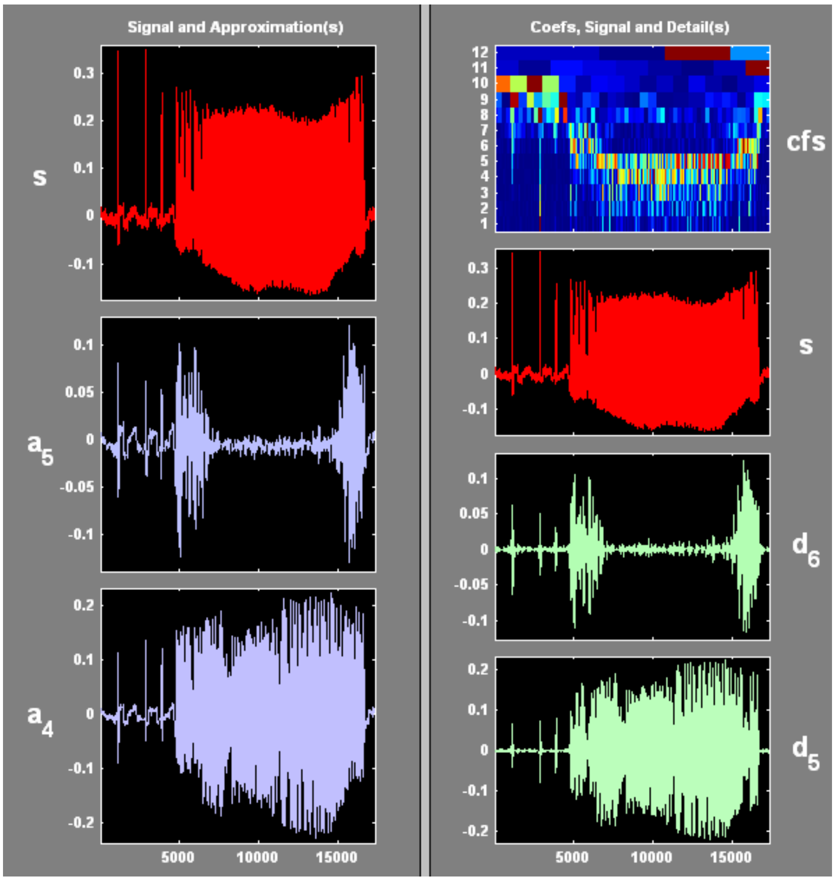

Figure 4- 37: Analysis of higher frequency oscillations in levels 4 and 5 of wavelet decomposition for control loop F1035

Since this information dominates the signal space, it is not necessary to zoom further into the data. A wealth of information is available from this analysis. During the transitions between high and acceptable variability – day night transition, there is a shift in the content of the frequency information. This is also clear in the cyclic pattern present in coefficient information in figure 4-35.

From the analysis in figure 4-38 it is possible to identify that the levels between 4 and 10 are contain most information in the transition periods (late evening and early morning). Level 4 oscillations are prevalent in the evenings, and level 10 oscillations during the day.

From table 4-6 we can determine there is a steady shortening of prevalent oscillation from 100 minutes at mid day, to about 5 or 10 minutes at sun set, down to less than 2 minutes during the course of the night. At sun rise the period of oscillation starts increasing again.

6 Discussion

The previous section presented the results of a detailed study of a number of problematic and well performing loops, using varied techniques to highlight the way various techniques can be used to troubleshoot controller induced process oscillation.

6.1 Results

In each of the subsections highlighting the use of different techniques, a detailed discussion of interpretation was given. In total, 127 control loops were analysed. Detailed analysis was performed on loops showing interesting characteristics, and 18 different control loops were analysed in detail.

The loops chosen for detailed analysis were selected to show specific behaviour or results within a given analysis method. Of these 18 loops, F1035 was used repeatedly to show examples of how analysis points to poor performance.

The lower order statistical information showed signs of non-linearity, but the absolute standard deviation of the flow was low. There were no set point moves, so it is not possible to see how well the loop responds to set point activity, but the disturbance rejection is good.

From a minimum variance perspective, F1035 also performs well. This is indicative of how close to best achievable performance F1035 gets.

The PV-OP plot investigation seemed to indicate a squarish pattern. This is typical of a loop with some degree of stiction induced non-linearity.

The power spectrum investigation showed a prevalent oscillation at a period of 43 minutes. The conclusion was that even though the performance of the loop is excellent, PID tuning is not well suited to removing slower oscillations.

The bicoherence investigation (higher order statistics), indicated clear non-linear behaviour in the control loop. However, it is not possible to diagnose the source simply using this one measure.

Finally, the wavelet investigation showed a very interesting pattern. During the day, the control loop performs optimally in terms of classic PID (some indication of very low frequency oscillation). During the night, higher frequency cyclic information is prevalent.

The above discussion is indicative of how important a total analysis of a control loop is, using all the information and methods at the control engineer's disposal. Using only a single method may lead one to conclude that the loop is in dire need of attention. All investigation into the non-linearity of the loop indicated the need for action, yet a simple statistical break-down shoed that the actual variability in the control loop was low. Minimum variance confirmed limited room for improvement.

6.2 Industrial requirements

Following on from the previous section, and the call for a holistic evaluation of loop performance, the requirements of an industrial base-layer performance monitoring system is presented. This discussion aims to expand on the practical issues involved in obtaining, interpreting and acting on the results and conclusions that can be drawn from the technical results obtained from a monitoring solution.

Most maintenance groups are structured to allow the maximisation of asset utilisation within their area of focus. This is even more prevalent amongst groups affecting the control, and thus the stability, of a production facility. The Strategic Asset Management Group (SAMI), suggest that five stages of operation mastery must be obtained prior to maximum asset utilisation. Figure 5.1 shows a diagrammatic summary of the SAMI thinking.

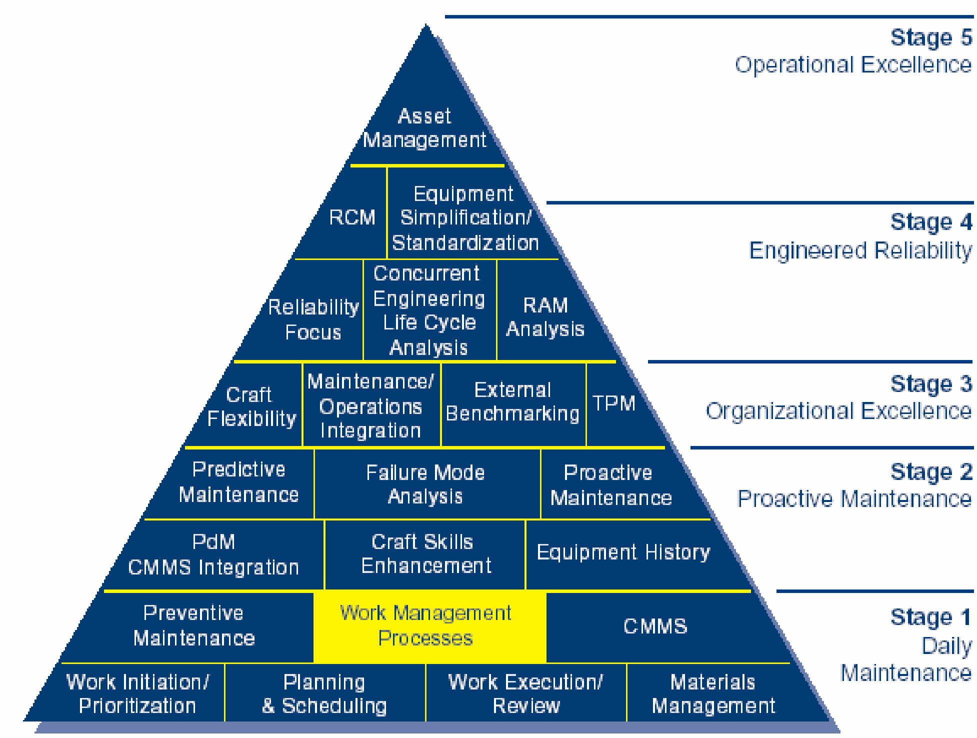

Figure 5- 1: SAMI process

Another view on obtaining operational excellence is a process know as DMAIC (Define, Measure, Analyse, Improve and Control). The basis of this improvement process is that if there is no measurement, there can be no improvement, and no control of the work environment.

By evaluating this statement, it becomes clear that all industrial sites which do not monitor the performance of their control layer, are pre-stage 1 in the SAMI triangle.

By simply introducing a measurement system, ie a purely technical solution, it may be possible to start performing some stage 1 functions, but the progression through organisation excellence, engineered reliability to operational excellence will not be supported by a measurement tool alone.

As such, any industrial solution should enable:

1. Making the highest value decisions (based on the intelligence built into the solution). This should be enabled by the thinking and business processes that are built into the monitor.
2. Executing the highest value decisions (based on the work flow that accompanies the solution). In other words, how does the monitoring tool assist in structuring the business to optimally realise benefit.

Based on these requirements, any monitoring solution (Harris, T. J. & Seppala, C. T., 2001) should perform strongly in the following areas:

1. Connectivity
2. Configuration
3. Analysis
4. Reporting (Shook, D. *et al*, 2002)
5. Auditing and work flow structure

Each of these areas of solution performance are handled separately below.

6.2.1 Connectivity

Any monitoring system requires data upon which to perform analysis. Most vendors should be capable of connecting to almost any control system, or to any historian. However, the ability to collect data is only one of the issues to be considered. The following questions can help guide the decision when evaluating different monitoring systems:

- What is the method of data collection?
- Does the collection of data have an impact on the performance of the actual control system? It is important to note what level of performance impact is acceptable, and ensure this is never exceeded (ie, does the collection of data impact available bandwidth on the control network, there-by decreasing controller schematic update periods from 1 second to

2 seconds). Does the monitoring system provide a method of protection against such network overloading?

- What is the impact of the monitoring system on the historian? Some systems collected data via an existing data collection system such as a historian, and can negatively affect the performance of such systems.
- Is it possible to collect high frequency data, and is this required in order to achieve acceptable performance?
- What is the integrity or the fidelity of the data collected using the method required by the system? What is the requirement of the monitoring system for high fidelity data, and how are errors or irregularities in the data handled? Is the collection system prone to failures in the network, data lagging, or filtering?
- Does the system provide continuous monitoring of the system performance? It is important to note the advantages and disadvantages, and the impact this could have on the business processes. If system audit are performed adhoc, this dramatically reduces the data and computational intensity of the solution. Personnel may not be in a position to deal with data collected at a higher frequency than that set for the audits. Balanced against this is the ability to monitor the performance of the control system during upset conditions.

6.2.2 Configuration

This section deals with the configuration and setup of the monitoring system. Many of the issues affect how the system will be maintained, and may affect the way in which maintenance is performed on specific loops.

- Is it possible to configure the monitoring system, and request specific information, from a central point? Is this important for the business? Many plants operate with several control rooms, and there may be a need to configure the monitoring system from a central point.
- How does the system direct maintenance attention to the most critical loops? Maintenance effort should be directed on a balance of performance and criticality. An example would be, How does the monitoring system ensure that maintenance effort is not expended on a poorly performing level controller that determines the flow of waste material to a drain?
- How does the system capture and highlight the interaction between different loops within the system? A common control setup is to have temperature control in a distillation column cascaded to a steam flow controllers. When this cascade (slave master) configuration is broken, there is effectively no temperature control, even though the temperature controller calculates required moves. How does the system handle these kinds of interactions (slave/master, feedforward)?
- How easy and intuitive is the use of the engineering front end? What additional calculation/diagnostic facilities are provided and do these provide a differentiating factor when evaluating system performance? This set of questions is largely to do with the aesthetics and usability of the system.
- Does the configuration of the system lend itself to easy use by the intended users? The configuration complexity that is sustainable may be different if the end user is a panel operator than a post graduate process/chemical/control engineer (Shook, D. *et al*, 2002).

6.2.3 Analysis

While it is clear that data analysis and measurement methods are not a differentiating factor, basic checks for completeness should be conducted. In addition, the presentation and the work flow associated with analysis can present a clear business advantage. The possibility for development within the analysis environment can also no be discounted, though progress in this area is more likely to come from an academic source than from industry. The following points about analysis should be evaluated:

- Which of the known measures are used in the performance measurement system? The list of measures presented in the literature survey can form a basis for compiling such a list.
- What is the intended audience of the measurement results? Within an industrial environment several different audiences have different information requirements. Basic control maintenance often forms the exclusive maintenance domain of personnel without a formal qualification in process control theory. As such, detailed and un-interpreted data may be overwhelming, or even unintelligible. Conversely, highly interpreted data may frustrate expert users. It is critical that the output of data analysis be matched to the intended audience.
- Is the plant interaction and loop design incorporated in the analysis? This is a similar question to that was posed in the previous section about configuration. The focus here is about algorithm and calculation specific techniques tailored to different control loop designs (gap controllers, gain scheduling).
- What new/solution unique measures are made use of? This is more prevalent where data is being interpreted to some degree. Pattern recognition, system identification and advanced time series analysis are currently intensely focused on in academia, and will increasingly filter down into industrial applications.

- What is the scope for implementing custom calculations? Control problems often don't have single solutions, and since the requirements for a certain control response may be unique, there is often a requirement for unique monitoring techniques.
- How much maintenance, in terms of process information, is required to keep the analysis accurate? Can the solution gather loop specific data automatically (controller mode, algorithm etc)? This touches on the give and take nature of the solution. Advantages can be gained by incorporating the controller structure into a monitoring package, but this must be balanced against the need to maintain changes when maintenance is done to the plant.
- Is the analysis robust enough to deal with standard control configurations? This deals with issues such as APC handles within the base layer. The set point of an APC application is regularly changed. This may interfere with standard analysis techniques, and should not necessarily be flagged as poor performance. In addition, poorly performing APC variables should be highlighted.
- Is it possible to customise analysis and benchmarking on a loop per loop basis? The occasion does arise where it is useful to benchmark the performance of a loop. If certain performance measures do not apply, then the specific loop potential should be considered. Non-linear process may be an example where this benchmarking is critical. Controller tuning is determined as a compromise between performance and robustness. While the potential may exist to improve control in the current operating region, once the operating region changes, the effect on plant performance may be negative.

Formatted: Bullets and Numbering

6.2.4 Reporting

Irrespective the design considerations regarding the prime users of a monitoring system, several parties have a vested interest in the performance of the basic control layer (Shook, D. *et al*, 2002). This section deals with the reporting capabilities of the monitoring system. The following reporting issues should be investigated:

- What technology (software) is used to present the results, and what platform (hardware) is required to accommodate that presentation? Once again there is no single perfect approach when choosing a reporting environment. Advanced visualisation and intelligent reporting can best be implemented in a custom environment. This requires local databases and has a maintenance and platform implication. In addition to the hardware requirements to run custom reporting, upgrades and additional calculations my not be explicitly handled. The flip side to custom reporting is using a flexible web-type front end. Advanced visualisation and custom reports are sacrificed for a more flexible reporting engine.
- Manufacturing companies normally have a significant investment in data management, both in the production (Manufacturing Execution Systems) and the business (Enterprise Resource System) environments. These data management systems form an important starting point when deciding on the best technology. Does the reporting environment allow easy integration into the existing data management systems.
- How does the grouping and drill down functionality of the reporting work? Plants and maintenance environments are structured according to function. Does the monitoring solution mirror the company structure?
- Does the information at each level of detail make sense? The information critical to the instrumentation maintenance foreman is totally different from that of the CEO of the company. Both may be interested in certain information, and the level of interpretation needs to be geared to their function. The maintenance personnel working the coal front and using the detailed reporting should also not be neglected.

- Are interactions and priorities carried through each level of detail? While prioritisation within different value chains and functional areas is critical, there may be instances when it is useful to view relative performance across functional areas. The drill down functionality within the solution, from a high level alert to the detailed problem analysis should be intuitive.
- Is the information displayed in a manner that allows for visual trend spotting and analysis? Is the visualisation technique effective and innovative? When presented with tabular data the human eye is at it's least effective for making conclusions and suggesting action. The naturally ability of the human brain for trend spotting should be exploited wherever possible.

6.2.5 Auditing and work flow structure

The above sections have dealt with how problems are measured, identified and then reported. In order to ensure there is continuous improvement of controller performance, work should be regularly audited. The following auditing and work flow pointers should be investigated:

- Does the system allow adhoc auditing of work recently performed? It is often difficult to judge the success of a maintenance action in the couple of hours spent executing the work on the plant. The ability to monitor the success of maintenance effort immediately after execution is important.
- How is performance information archived and compared to the current performance? Long term improvement needs to be measured, but data storage and detailed information/reports should be managed sensibly to keep infrastructure requirements reasonable. What level of performance information is available for what length of time, and how easily can summaries be extracted from the system?
- Does the vendor have a suggested work flow structure to accompany the solution? When first starting to monitor and improve basic controller performance, most companies are not structure and geared to deal with

the new information. Suggestions about how to structure, staff, train and deploy new personnel can be critical.

- Is it possible to benchmark plant controller performance against other players in the industry. Keeping track of the differences between current performance, best in class and best achievable performance, gives an accurate indication about how effectively the solution has been deployed.

7 Conclusions

Several different analysis techniques were explained and practically demonstrated. From the results and discussions, it is clear that there are advantages to several of the techniques presented. Each piece of analytical data presents a complementary piece of information with which the trouble-shooter can analyse the problem.

Several examples of unusual or difficult to troubleshoot examples were chosen and examined using different techniques.

In the case of flow F1002, first glance statistics indicated that there was high variability, and potentially high non-linearity. All this indicated a loop in strong need of maintenance.

The CLP of the process was analysed, and indicated that there was significant room for improvement, specifically in a well designed control loop. However, by examining the process over the entire time period, and investigating the range of the SP and OP, it was clear that there had been an abnormal situation in the middle of the data set. The possibility of removing this type of information by means of well designed filters exists, but was not investigated.

Plotting the CLP over time (in this case achieved by windowing the time data) highlighted the areas of poorer performance. This is recommended as a standard procedure for maintenance personnel. By tracking performance information over a significant period of time, clear abnormal situations can often be identified. This will prevent incorrect system changes that are based on erroneous information provided by aggregation over unstable conditions.

Flow loop F1038 shows significant probability of sticition (or other non-linearity) when examining the histogram of the process error. However, when performing the both cross-correlation tests, the information is much less damning. This contradictory information is indicative of the uncertainty inherent in the measurements.

Contrasted to F1038 is flow loop F2024, where the histogram is almost perfectly normal and the cross-correlation test seems to indicate that there may be cause of stiction investigation. Further analysis shows high process variability, especially for a flow loop. However, when the above is read in conjunction with the CLP information, it is clear that the loop is performing close to minimum variance, and that the opportunity for improvement is likely to be small. If the technical staff had acted on any of the information without considering all the facts, an incorrect conclusion was likely.

Pressure controller P1001 is just one example about the difficulty of applying standard techniques to the process industry, where unique setups and special control strategies abound. P1001 is an over pressure controller, and when analysed using standard techniques shows signs of sluggish control, high non-linearity and incorrectly sized valves. However, this controller is designed never to reach SP under normal process conditions, and clearly needs to be monitored differently.

L2051 and L1043 are 2 level controllers which are also configured unusually, and where the standard techniques fell flat. L2051 is a proportional only action controller, which has no feedback action. As such, the explicit OP-PV mapping leaves analysis of errors meaningless since there is no input from error into the control algorithm. L1043 has an external control algorithm which sets the SP equal to the PV on every execution cycle, and controls the OP remotely. This makes traditional analysis very difficult.

Flow controller F1035 was used as an example through most of the measurements, due to its unique type of problem. Firstly the statistics showed that the loop was performing fairly well. However, analysis of the non-linearity showed uncertainty as to the perfect performance of the control loop. Based only on this information, the technical staff would have left the loop unattended.

The bicoherence technique makes use of higher level statistics on frequency based information, and this technique indicated significant levels of non-linearity within the process information.

Applying a power spectrum decomposition on the frequency domain information, yielded results that showed a dominant oscillation at 43 minutes. However, this was contrary to the visual inspection, which indicated a much lower frequency disturbance.

In order to provide further insight into the performance of this control loop, a novel wavelet based technique was applied. The error signal of F1035 was decomposed into the various wavelet basis functions, and these were firstly visually displayed. This allowed the trouble-shooter to easily determine the interaction between the time of day and the frequency of the oscillation. By further investigating the nature of the oscillation, by interpreting the detail and approximation information at each decomposition level, it was possible to determine that the dominant low frequency disturbance was between 60 and 120 minutes. This was confirmed by visual inspection (ie 93 minutes).

What the standard visual inspection, statistics and frequency domain techniques failed to highlight was the presence of a high frequency disturbance during the evenings. While PID control is ill suited to removing slow continuous oscillations, it should remove higher frequency disturbances. The wavelet decompositions clearly showed that the performance during the evenings could be improved.

The comparison between the wavelet decomposition and the normal frequency based techniques highlights a major shortcoming in attempting to group common oscillations within processing plants. Most plant information is likely to be dominated by day-night cycles, as well as other repetitive, and time-localised, disturbances. Since the translation to the frequency domain disregards all time based information, the frequency decomposition is in effect an aggregation of performance information. This makes the grouping, and thereby the isolation of the source of oscillation, very difficult.

The examples discussed above, as well as those that only appeared in the discussion section, highlight several important characteristic of closed loop performance management:

- Each closed loop performance technique provides a portion of unique information – a partial view of the whole truth.
- Unusual performance problems may be characterised by conflicting results.
- The aggregation of results can lead to erroneous conclusions. Firstly, in the example of interpreting the CLP, and secondly from the use of non-localised frequency based techniques.

It is clear that by analysing several different closed loop performance measures, one is able to arrive at a better picture of the true performance, provided that the analysis is interpreted by a knowledgeable specialist.

Several best practices for the choice of a performance framework were suggested. Along with the conclusion of multiple techniques to arrive at a true conclusion, comes the burden of interpreting the results from these techniques. Many of the measures presented as part of this study require some level of control theory in order to interpret.

Therefore, when purchasing or developing a closed loop performance measurement environment, the key question needs to be the skill level of the audience. Combined with this are the considerations of data collection, business processes, data visualisation and work flow. All are important points to consider when purchasing or developing a monitoring package. The most successful implementations of these techniques will always be found in combination with sound technical and process knowledge. Both are key to interpreting the results presented by the monitoring system. Without this human intellectual property combining with the intelligence captured in the system, maintenance cannot occur. Either in isolation are not totally effective.

8 References

Bakshi, B. R. and Stephanopoulos, G. (1996) "Compression of Chemical Process Data by Functional Approximation and Feature Extraction" *AIChe Journal*, 42 (2), 477 – 492.

Carrier, J. F. and Stephanopoulos, G. (1998) *'Wavelet-based modulation in control-relevant process identification'*, AIChE Journal, *44* (2), .341 – 360.

Choudhury, M. A. A. S., Shah, S. L. and Thornhill, N. F. (2002) "Detection and Diagnosis of System Nonlinearities using Higher Order Statistics", collaborative paper from *University of Alberta,* Edmonton, Canada and *University College,* London, England.

Daubechies, I. (1988), *'Orthonormal bases of compactly supported wavelets'*, Communications on pure and applied mathematics, Vol. XLI, 909 – 996.

Dauberchies, I. (1992) Ten Lectures on Wavelets, SAMI, USA.

Desborough, L. and Miller, R. (2001), "Increasing Customer Value of Industrial Control Performance Monitoring – Honeywell's Experience", paper presented at *CPC VI Meeting,* 2001.

Desborough, L., Nordh, P. and Miller R. (2001) "Control Systems: Process out of Control" *Industrial Computing,* http:// www.isa.org/journals/ic [2003, June 4].

De Vries, W. R. and Wu, S. M., (1978) *Evaluation of process effectiveness and diagnosis of variation in paper basis weight via multivariate time series analysis,* IEEE Transactions on Automation and Control, 23, 702-708.

Gustafsson, F. and Graebe, S. F. (1998) "Closep loop performance monitoring in the presence of system changes and disturbances, *Automatica,* 34 (11), 1311-1326.

Hägglund, T. (1995) *'A control-loop performance monitor'* Control Engineering Practice, *3*(11), 1543-1551.

Harris, T. J. (1989) "Assessment of Control Loop Performance" *The Canadian Journal of Chemical Engineering, 67.*

Harris, T. J. and Seppala, C. T. (2002) *Recent Developments in Controller Performance Monitoring and Assessment Techniques*, Project Report, Dept. Chem. Eng., Queen's University, Kingston, USA.

Harris, T. J., Seppala, C. T. and Desborough, L. D. (1999) "A Review of Performance Monitoring and Assessment Techniques for Univariate and Multivariate Control Systems", *Journal of Process Control*, 9, 1-17

Horch, A. (1998) *Extensions of a performance index and detection of static friction in process control loops*, PHD Thesis, Royal Institute of Technology, Stockholm, Sweden.

Huang, B. and Shah, S. L. (1999) *Performance Assessment of Control Loops Theory and Applications*, Springer-Verlag, London, 9-15.

Huang, B., Shah, S. L., Badmus, L. and Vishnubhotla, A. (2000) *Control Loop Performance Assessment: An Enterprise Asset Management Solution*, Research Report, Dept. Chem. & Mat. Eng., University of Alberta, Canada.

Hugo, A.J. (2002) "Process Controller Performance Monitoring and Assessment", *Control Arts Inc.*

Ingimundarson, A. (2001) *Performance Monitoring of PI Controllers using a Synthetic Gradient of a Quadratic Cost Function,* Project Report, Department of Automat Control, Lund Institute of Technology, Sweden.

Kaiser, G. (1994) *A Friendly Guide to Wavelets*, Birkhauser, Boston.

Kozub, D. J. (1996) "Controller Performance Monitoring and Diagnosis: Experiences and Challenges", paper presented at *Fifth International Conference on Chemical Process Control,* 7-12 January, 1996, Tahoe City, California, USA.

Ljung, L. (2002), *ARMAX, Output error and Box-Jenkins*, Matlab Systems Identification Toolbox, The Mathworks Inc, Natick.

Luyben, W. L. (1989) *Process Modeling, Simulation and Control for Chemical Engineers 2nd Edition*, McGraw-Hill, Singapore, 615-629.

Mallat, S. G. (1989) *'A theory for multiresolution signal decompostion: The wavelet representation',* IEEE transactions on pattern analysis and machine intelligence, 11 (7).

Misiti, M. and Misiti, J. and Oppenheim, G. and Poggi, J., (1996) *Wavelet Toolbox User Manual*, The Mathworks Inc, Natick.

Shah, S. L., Patwardhan, R. and Huang, B. (2001), "Multivariate Controller Performance Analysis: Methods Applications and Challenges", paper presented at *CPC-6 Conference,* January 2001, Tuscon Arizona, USA.

Shook, D., Shetsen, A., Balabuck, I. and Khalili, R. (2002) "A Web-based Approach to Control Loop Assessment in a Mineral Processing Facility" paper presented at *CIM Conference,* May 1, 2002, Vancouver, B.C, Canada

Stenman, A., Forsman, K. and Gustafsson, F. (2004) "A Segmentation-based Approach for Detection of Stiction in Control Valves: *Journal of Signal Processing and Adaptive Control,* 17 (7-9), 625-634.

Stephanopoulos, G. (1984), *Chemical Process Control An Introduction to Theory and Practice*, Prentice-Hall, New Jersey, 570-588.

Stephanopoulos, G and Carrier, J. F. (1998) "Wavelet-Based Modulation in Control-Relevant Process Identification" *AIChE Journal,* 44 (2), 341 – 360.

Strang, G. (1989) '*Wavelets and dilation equations: A brief introduction'*, Society for industrial and applied mathematics review, 31 (4).

Tham, M. T. (1999) *'Minimum Variance and Generalised Minimum Variance Control Algorithms',* Project Report, Department of Chemical and Process Engineering, University of Newcastle upon Tyre, 1999.

Thornhill, N. F., Shah, S. L. and Huang, B. (2001) "Detection of Distributed Oscillations and Root-Cause Diagnosis" paper presented at *CHEMFAS 4,* June 7-8, 2001, Jejudo (Chejudo) Island, Korea.

Tyler, M. L. and Morari, M. (1995) 'Estimation of cross directional properties' *AICHE Journal*, 41, 846 – 854.

Qin, S. J. (1998) "Control Performance Monitoring – A Review and Assessment" paper presented at *NSF/NIST Measurement and Control Workshop*, March 6-8, 1998, New Orleans, USA.

Wan, S. and Huang, B. (2002) "Robust Performance Assessment of Feedback Control Systems", *Automatica,* 38 (1), 33-46.