Development of a Regulatory Performance Monitoring Structure

by

Ruan Minnaar du Toit

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

Master of Engineering (Control Engineering)

in the

Department of Chemical Engineering
Faculty of Engineering, the Built Environment and Information Technology

University of Pretoria
Pretoria

December 2005

Development of a Regulatory Performance Monitoring Structure

Author:	Ruan Minnaar du Toit
Supervisor:	Prof PL de Vaal
Department:	Department of Chemical Engineering University of Pretoria
Degree:	Master of Engineering (Control Engineering)

Synopsis

A number of factors have contributed to increased pressure on plant operating efficiency in the chemical processing industry. These factors include more stringent environmental and safety regulations, global economic pressures and downsizing of many support services in order to save money. Control performance monitoring is a tool that is used to keep automated control systems performing as optimally as possible. Various performance metrics and methods exist to evaluate plant operation. In essence, however, they all refer to the same principle which is to indicate how far a plant is operating from its inherent optimum and what can be done to ensure that the gap between the optimum and the current operation is as small as possible for the longest possible period.

Performance monitoring is, although well researched, not yet a generic, complete and specific application. Current shortcomings of monitoring applications are that it is process or unit operation specific and that it provides a local indication of performance and not a plant wide evaluation of how close the plant is operating to its inherent optimum. Performance reports are usually in terms of statistical measures and graphics which are usually abstract and vague. For high level decision making (on operation end economic investment) simple and quantifiable measures are needed that are repeatable and transparent.

The focus of this project was to develop and implement a regulatory performance monitoring structure for real-time application on an industrial pilot scale chemical process.

The structure was implemented by means of two graphical interfaces. The first provides a holistic plantwide indication of performance and indicates sources of poor performance in the regulatory control structure. The plantwide interface includes a proposed plant wide performance index (PWI) that reduces operational efficiency to one specific number. The second interface supplements the plantwide interface by providing statistical information on individual loop performance. The individual loop interface is a tool to locate causes of poor performance in the regulatory control structure to aid controller and plant maintenance.

Keywords: regulatory performance, statistical process control, plant wide control

Acknowledgements

I would like to thank Professor Philip de Vaal for his guidance and insight, my family and friends for their amazing support and also Fluor SA (PTY) LTD for providing financial assistance.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

$\bar{P}$	Average of the steam inlet pressure	kPa
$\bar{T}$	Average of the plate temperature	°C
Δt	Sample period for a discrete signal	sec
$\hat{y}$	Output prediction	-
$\overline{x}$	The mean of a discrete variable x	-
a	Oscillation amplitude	-
$A,\ B,\ C,\ G,\ F$	Coefficient polynomials in the backward shift operator	-
bot	Bottom plate	-
c	Simplified kurtosis or fourth moment around the data mean	-
C_i	The cumulative sum at time instant i	-
C_{xy_k}	The cross correlation coefficient between the variables x and y for a lag of k	-
C_{xy}	Coherence for two variables x and y	-
$Cov(x, y)$	The covariance between the discrete variables x and y	-
D	Deadtime	-
d	Load disturbance	-
E	Residual matrix	-
e	The control error	-

F	Feed flow rate	kg/hr
$Feed_i$	A feed value factor for stream i	-
G_R	Desired closed loop dynamic transfer function	-
h	The rise distance for construction of a V-mask	-
H_2	The optimal norm that indicates minimum variance	-
J	Goal or cost function	-
k	The slope for construction of a V-mask line	-
k_1	Constant used to construct control chart limits	-
$kurtosis$	The fourth moment around the data mean	-
$load$	Load detection term (either 1 or 0)	-
m	Number of feeds into the plant	-
n	Number of product streams leaving the plant	-
n	The number of discrete samples taken	-
n_{lim}	Number limiting the allowed oscillations	-
NN	Nominal operating state	-
OI	Operability index in terms of operating spaces	-
p	loading vector	-
$P_c,\ F_c$	Weighting factors	-
P_{xy}	Cross power density for variables x and y	-
P_{yy}	Power density	-
PI	Performance Index	-
$Prod_i$	A product value factor for stream i	-
Q	Sum of the product flows	kg/hr
q	Number of utility streams entering the plant	-
$q^{-1}\ or\ z^{-1}$	Backward shift operators	-

R_k	The autocorrelation coefficient for a lag of k	-
S	Standard deviation	-
s	Sensitivity factor	-
S_{hi}	V-mask higher limit	-
S_{lo}	V-mask lower limit	-
t	Score	-
t_a	Starting instant for an evaluation period	-
t_b	Ending instant for an evaluation period	-
T_u	Ultimate period of oscillation	-
T_{sup}	Supervision time for oscillation detection	-
top	Top plate	-
U	Cooling water flow rate	kg/hr
u	Process input	-
UPI	Unit performance index	-
$Util_i$	A utility value factor for stream i	-
V	Volume of a surge tank	-
w	Any discrete variable used as a performance characteristic	-
w_i	Weighting factor for the plant wide index	-
x	Random external disturbance	-
Y	The fast Fourier transform	-
y	Process output	-
Z	The Z-transformation for non-normal distributions	-
Z_i	The exponentially weighted average at time instant i	-

Subscripts

act	Actual operating point	-

cap	Capability of the process	-
d	Calculated in the disturbance space	-
fbc	Deviation from minimum variance	-
k	The lag for correlation calculations	-
lim	Limiting bound on a variable	-
opt	Optimal operating point	-
tot	Calculated over the entire sample population	-
u	Calculated in the input space	-
$user$	User specified	-
y	Calculated in the output space	-

Greek

γ_{skew}	The result of the third moment about the data mean	-
λ	Constant used to calculate the EWMA	-
μ	Measure function to calculate the size of a space	-
ω	Frequency	-
ω_u	Ultimate frequency	-
σ_i	Standard deviation of variable i	-
τ	Time constant in the LaPlace domain	unit of time
τ_i	Controller integral time constant	-
υ	Controllability index in terms of surge volumes	-

Abbreviations

AIS	Available input space	-
AOS	Available output space	-
APC	Advanced process control	-
ARL	The average run length	-

CBA	Cost benefit analysis	-
$CUSUM$	Cumulative sum	-
CV	Controlled variable	-
CW	Cooling Water	
DCS	Distributed control system	-
DIS	Desired input space	-
DOS	Desired output space	-
DP	Differential pressure	-
EDS	Expected disturbance space	-
$EWMA$	Exponentially weighted moving average	-
FFT	Fast Fourier transform	-
ISA	Integral of the absolute error	-
ISE	Integral of the error squared	-
LAN	Local area network	
LCL	Lower control limit	-
LQG	Linear quadratic control	-
MPC	Model predictive control	-
MV	Manipulated variable	-
MVC	Minimum variance control	-
OPC	Object linking and editing for process control	
PFD	Process flow diagram	-
PID	Proportional integral derivative	-
PSD	Power spectral density	-
PWI	The plant wide evaluation index	-
TDS	Tolerable disturbance space	-
UCL	Upper control limit	-

CHAPTER 1

Introduction

1.1 Background

Performance monitoring, assessment and operational diagnosis of chemical plants has become one of the most active research areas in the field of process modelling and control in recent years. The reason for this is due to the more stringent requirements on plants to become more profitable. These requirements are due to many factors which include more stringent environmental and safety regulations, global economic pressure to operate as efficiently as possible and downsizing of many support services in order to save money.

Control performance monitoring/assessment (CPM/CPA) is a tool that is used to keep automated control systems performing as optimally as possible. Performance monitoring exists under a number synonyms in industry and in literature which include loop monitoring, loop auditing, loop management, performance assessment, etc. They all basically refer to the same principle which is to indicate how far a plant is operating from its inherent optimum and what can be done to ensure that the gap between the inherent optimum and the current operation is as small as possible over the longest possible period of operation.

Recent advances in process control technology have made application of advanced process control (APC) techniques as well as process modelling and characterisation more sustainable and implementable. These advances have identified a need for more effective performance evaluation systems to identify possibilities for APC and to sustain/maintain successful APC implementations.

Performance evaluation techniques can be separated into two main areas of application:

- A real-time, on-line type evaluation of plant operation. This includes loop monitoring, early fault detection and justification of advanced control investments.

- Performance evaluation in the design stage of a process to aid in decision making on various process and control configurations. Performance measures in the design stage of a process is a very handy tool to ensure more effective process and controller design integration. Performance monitoring at the design stage of a process is usually referred to as controllability analysis.

1.2 Problem statement

Performance monitoring is, although well researched, not yet a generic, complete and specific application. Current shortcomings of monitoring applications are that it is process or unit operation specific and that it provides a local indication of performance and not a plant wide evaluation of how close the plant is operating to its inherent optimum. Performance reports are usually in terms of statistical measures and graphics which are usually quite abstract and vague. For high level decision making (on operation and economic investment) simple and quantifiable measures are needed that are repeatable and transparent.

1.3 Research objectives

The focus of this project is to develop and implement a performance monitoring structure that functions as a real-time application for performance assessment of an industrial pilot scale chemical process.

The performance structure will be transparent and generic so that it can be easily interpreted to aid in the validation of changes to the general control structure or in the validation of implementation of advanced control applications. It will therefore cater for management level decision making of plant operation and investment as far as possible on a plant wide scale.

1.4 Method

1.4.1 The process

The process consists of a 10-plate glass distillation column. It will separate binary mixtures of ethanol and water. The column is fully equipped with the latest in control instruments that communicates with a distributed control system (DCS). The digital communications are enabled by DeltaV operating software which allows for efficient data capturing and application of various regulatory and advanced controller function blocks.

1.4.2 Performance monitoring structure

The structure will comprise of data retrieval from the process through the OPC communication protocol. The retrieved data will then be reported and statistically manipulated through the MATLAB environment to deliver measures of performance. The performance information will be interpreted through central graphical interfaces which will contain the whole control structure. From the central interfaces report generation is possible that will give a holistic summary of present operating performance.

CHAPTER 2

Background

This chapter provides background on performance monitoring. It defines what an effective plant performance evaluation system is and why there is a need for effective plant performance monitoring. Problems with implementing and maintaining the monitoring structure are also identified. Finally the similarities between plant performance and the general optimisation problem are assessed.

2.1 Defining plant performance

Chemical plants have an inherent maximum throughput and operating efficiency that are determined by the process design and equipment used. This inherent capability of a plant is independent of the control system and it is important to note that poorly designed plants cannot be compensated for by control systems. A control system's purpose is to ensure that the plant is operating as close to its inherent optimum as possible while keeping operation stable, safe and within environmental constraints (Blevins, McMillan, Wojsznis, and Brown, 2003).

A performance monitoring system should continuously evaluate product quality and the production levels of a plant. The base of any performance monitoring system consists of measurable plant data that are reworked (usually statistical) to provide a clear and simple picture of the operating states of a plant. With the operating states known, one needs a means of comparison to evaluate current performance. This is why most performance systems have a certain performance benchmark that current operation is compared with. The benchmark is an indication of the inherent optimum that is set by the process design and equipment. With the states and operating regions known together with the plant's inherent optimum capability defined, operational problems and poorly performing plant sections can be identified. A method of diagnosis is then necessary to find the cause of bad performance (Perry and Green, 1998).

Current performance monitoring systems mostly provide an indication of current regulatory performance. This fits into the general loop monitoring approach that has been well studied ((Mosca and Agnoloni, 2003), (Xia and Howell, 2003) and (Salsbury, 2004)). Loop monitoring is a very effective methodology to locate loops that are not operating to potential but is limited to a discrete part of the control system and fails to give an overall plant performance indication. This is where advanced control techniques fit in. Advanced control fits on top of the regulatory control level to make sure that the plant as a whole is optimised and not just localised feedback loops. Advanced control can in some cases replace the traditional regulatory feedback system, where inputs to final control elements are written by considering a number of process variables and conditions. Performance monitoring definitely also covers advanced techniques, for instance the accuracy of plant models that are used in control algorithms need to be determined. Obviously the base layer and regulatory performance monitoring structures always need to be in place. Advanced control is useless without a proper base layer system which is functioning properly.

With the loop or plant efficiency known and causes of the poor performance located, recommendations can be made to improve current plant operation. An efficient and fixed plant performance structure will provide the platform for decision making with a view to better plant performance. The monitoring and evaluation structure should be a continuous and iterative procedure seeing that performance needs to be evaluated constantly and if changes to the process are made, the performance structure should be adapted to reflect the new system and the new improved operating states. A well structured process monitoring structure is discussed in chapter 6.

The decisions that will be aided by a proper performance structure will be related to changes to the process design, changes in the operating philosophy, capital investment justification, etc. The structure should be of such a type as to be understandable by all the parties involved in decision making of general plant operation. These parties include management, operations and the engineering/design office. The structure should also be completely generic and be applicable to any type of unit operation. The individual unit operations combined can then give a general idea of where the plant is operating at and where it is heading.

If plant performance can be defined in one sentence it will be something like the following:
An effective plant performance monitoring structure provides a generic, simple and complete illustration of how far a particular process is operating from its inherent capability, with an ultimate aim of locating the cause of bad performance and adjusting the current control system to bring the plant closer to its inherent capability which is determined by the design of the process.

2.2 A general plant performance evaluation methodology

Although performance evaluation techniques are vast and based on numerous different statistical and mathematical methods a clear methodology is apparent. A general universal methodology that illustrates the general method for determining plant performance can be set-out as follows (Schäfer and Cinar, 2004) (Harris, Seppala, and Desborough, 1999) (Blevins, McMillan, Wojsznis, and Brown, 2003):

- Setting a benchmark - The first step is to obtain a benchmark of optimal operation and control. The optimal base case is where the plant should be operating under the design, safety and environmental constraints that apply.

- Current operation - After the optimal benchmark has been set-up, the current operating point of the plant is determined. This is done through field measurements that are gathered and manipulated through statistical means to give a clear indication of how the plant is operating.

- Comparison - The real plant operation is then compared with the benchmark to determine the performance of the plant. It is in this step where the actual performance is determined and decisions are made as to whether the plant operation is satisfactory.

- Diagnosis - If performance is not up to standard the causes for suboptimal performance are identified in this step. Possible improvements and solution methods are proposed and evaluated. There should be a strong economic factor in this step to aid the decision making.

- Implementation - The solutions as well as a post installation framework are implemented to determine what the real benefit of the changes are and how it compares with the initial estimates that aided the decision making.

- Maintenance - This is where the initial success that may have been achieved is made sustainable. For instance, the initial monitoring system should be adapted to include the new methodologies and technologies.

This general methodology is what performance monitoring structures usually comprise of to ensure that the plant is operating as close as possible to the inherent optimum. The structure should automatically detect degradation in performance and locate the cause. The cause should then be rectified by plant maintenance. A complete and detailed performance structure is developed and implemented on a lab scale unit operation in chapter 6 and 7.

2.3 Performance and the plant control hierarchy

A successful performance monitoring structure is an automated, continuous, active process that is functioning right through the control hierarchy of the process that it is implemented on. This section explains what the function of performance monitoring is at various levels of the control hierarchy. The general plant control hierarchy is shown in figure 2.1.

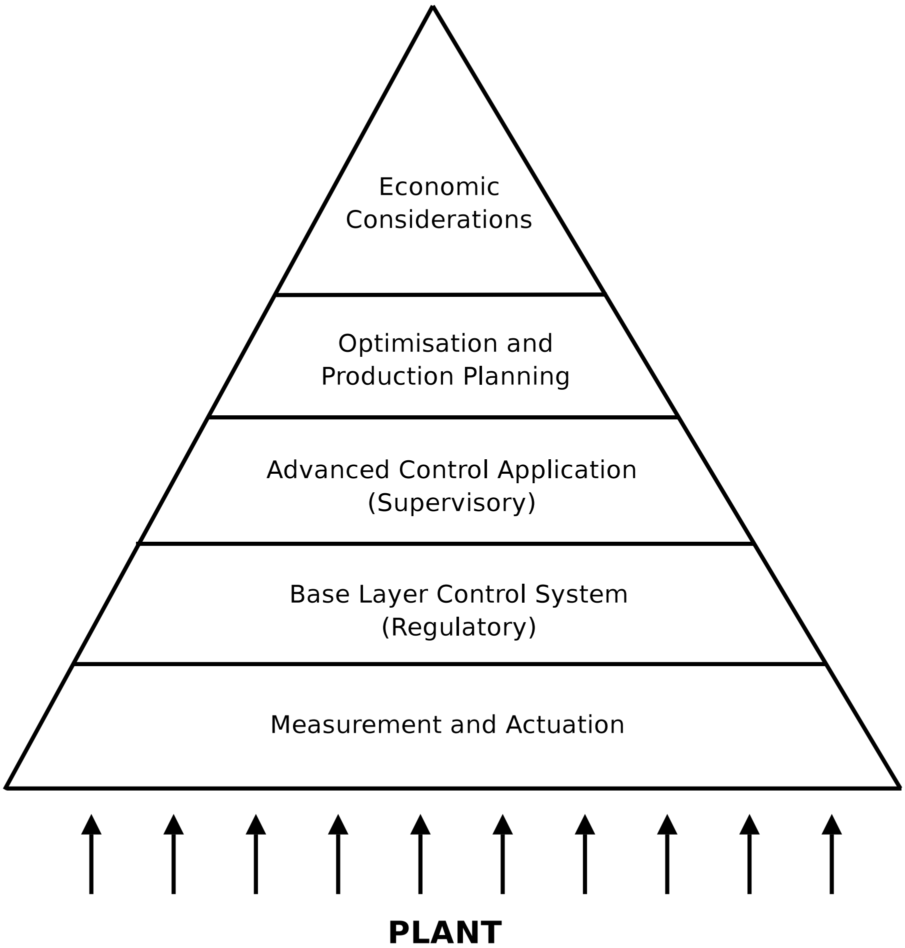


Figure 2.1: Plant control hierarchy

From figure 2.1 it is apparent that plant instruments form the base or foundation of the control system and is of prime importance for successful plant operation and control. Without operational and accurate instruments there is no means to know what the operating states of the plant are and no adjustments to process variables will be possible. The monitoring structure's role in this foundation level of control will firstly be to automatically detect faulty measurement, final control element failure, etc. Then it should localise the faulty instrument with a diagnosis of what the problem is, for instance valve stiction.

The next level of operation is the base layer control system where regulatory control is applied to compensate for load disturbances on the process. Normal feedback and conventional control algorithms are used to adjust final control elements based on plant measurements all within a distributed control system (DCS). The function of the performance structure will be to determine how effective the regulatory control system is performing. This type of monitoring is traditionally referred to as loop monitoring. The performance structure will typically be of use in this level to indicate which feedback loops need retuning or indicate which loops are running on manual, etc.

The next level of operation is the advanced process control (APC) level. At this level advanced control algorithms are implemented like for instance model predictive control (MPC), feedforward control, internal model control, cascade control, etc. This level was traditionally seen as the control level that writes set-points to the base layer (DCS). Therefore it is imperative that levels beneath the APC level are fully functional and working efficiently to ensure success of APC implementation. The performance monitoring structure at this level is much the same as for the base layer in terms of identifying poor performance but the causes of poor performance may be different. For instance a badly tuned PID loop will give rise to bad performance in both the base layer level and in the APC level, but an inaccurate model which is used in a MPC algorithm will only be detected in the APC level.

The upper levels in the control hierarchy is where financial aspects of plant operation are of main consideration. This is where, for instance, it is decided how operation should change to satisfy changes in market demand. Approval of changes to plant design and major capital investments happen at this level. Traditional cost benefit analysis (CBA) is a tool that is used in this level. The one shortcoming in the control hierarchy has been the lack of communication between the advanced control layer and its subsequent lower levels and the top managerial layers. The reason being that it is difficult to relate the performance measures used in the regulatory and supervisory levels to economic type measures which the top decisions makers can understand. That is where the performance monitoring structure should come in and aid in rectifying this communication breakdown. The structure should provide a clear, simple and generic indication of how the plant is performing. The structure should therefore provide a holistic indication of plantwide performance and provide possible solutions to improve operation.

2.4 Performance and plant operation

Figure 2.2 shows how the general flow of information in daily operation happens and how a performance monitoring structure fits into it (Perry & Green, 1998).

As can be seen from figure 2.2 the evaluation procedure is a closed loop that needs to be executed on a continual basis to improve performance. Firstly data is gathered from past periods of operation (historian) to obtain a benchmark. Next, current data

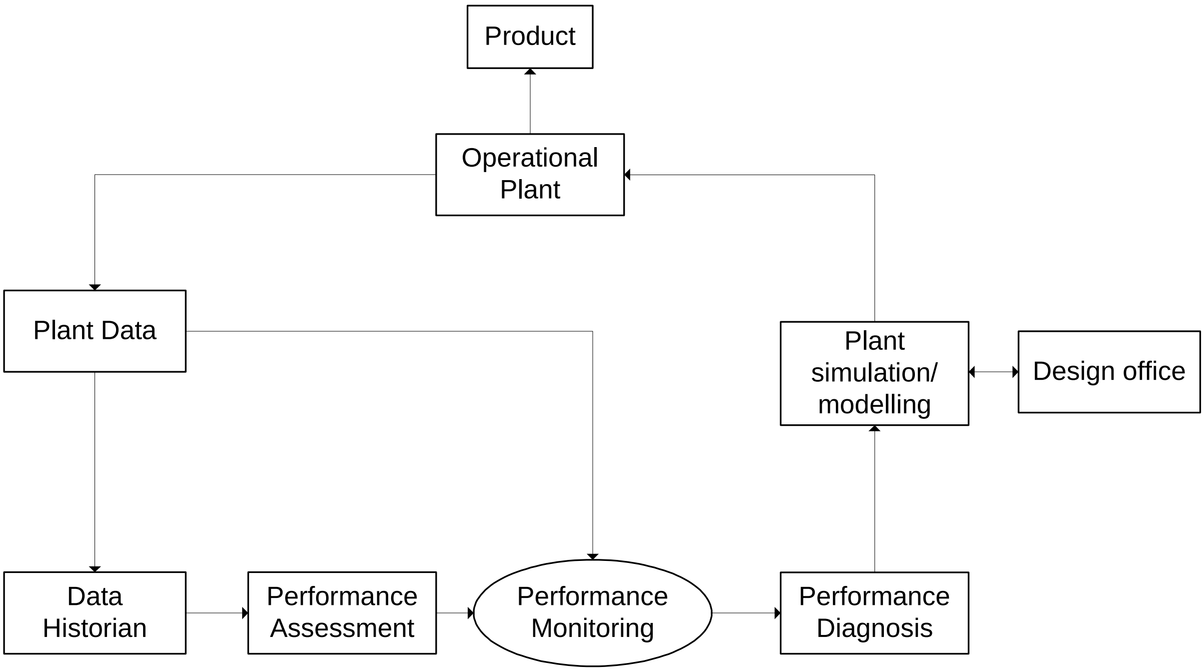


Figure 2.2: Plant performance and operation diagram

is gathered from the plant and compared with the benchmark. After the comparison is made, diagnosis can be performed to locate root causes of poor performance. When the root causes have been identified the information is used to develop possible solutions to the problem. This is done by means of plant modelling and simulation. If the root causes cannot be eliminated by the normal control applications one should turn to APC or plant design alterations.

2.5 Benefits of performance evaluation

There are a number of benefits to having a proper performance structure in place. The benefits to operation originate from two main sources. The first is that the control structure is optimised to enable closer operation to the process constraints (inherent optimum). This will provide the following benefits for production:

- Increased throughput
- Better and more consistent product quality
- Reduced energy usage
- Reduced waste products

The second major source of benefit is enabling more efficient operational maintenance. This will:

- prevent operator information overload
- lessen the workload on control and instrumentation technicians
- reduce dimensionality of optimisation problems
- provide specific fault detection
- better the planning of scheduled downtimes for maintenance

The result of an effective plant monitoring system will be a more efficient, safe and profitable plant (Perry & Green, 1998).

2.6 Problems with implementing performance monitoring strategies

2.6.1 Data dimensionality

The number of control loops on processing facilities are large and the information available for a single loop is substantial in itself. This makes extracting the right information from datasets very difficult. It is impossible to evaluate each and every loop individually seeing that the amount of information available is so much. The number of control and instrumentation staff on a plant is usually also limited which makes loop maintenance even more difficult.

Data capturing and storage is also a problem due to the sheer amount of data available and the communication between the historian and the DCS system can sometimes be lacking. This is due to data compression and filtering to preserve storage space which can sometimes lead to loss of important data related to dynamic behaviour.

With advances in process control technology the amount of information that is available from the process has grown. This is not necessarily a good thing seeing that one can get lost in the sheer amount of data. Commercial developers of performance monitoring products should realise that more information is not necessarily a good thing. The more information available the more time will be needed to retrieve and evaluate it.

The challenge to implement a successful performance monitoring structure is to reduce the dimensionality of the plant information and to transform it into useful and applicable information.

2.6.2 Multi-variable performance evaluation

Practically all data on a plant are interconnected. This means that when considering a single loop's performance, often a lot of interaction factors need to be considered to find possible sources of performance degradation. That is why an integral part of a successful performance monitoring structure should be some kind of multivariate analysis to give a more holistic idea of where the plant is operating and how the variables interact.

There has been success in the application of univariate performance assessment research in recent times, but the application of multivariate assessment research technology remains a hard task. There has been a reasonable amount of work published for the multivariable case but very little of the work has been successfully implemented in commercial packages or in customer installations (Huang, Ding, and Thornhill, 2004). Multi-variate systems show a definite difficulty if one needs to obtain a benchmark by means of the minimum variance control (MVC) technique. For the univariate case the parameter that is needed is an estimate of the dead-time which can be obtained from the closed loop response without disturbing normal operation. For the multi-variate case this is not the case. There needs to be more information known about the process than just the delays between each input-output pair. This information is locked within the termed interactor matrix which characterises the deadtime structure of the process. Some of the work done in this area was done by Harris, Boudreau, and MacGregor (1996) and Huang and Shah (1999).

2.6.3 Non-linear systems

Normal plants are non-linear but most of the techniques used to evaluate performance are usually based on linear methods. These techniques work well if processes do not deviate too much from their expected operating point.

Non-linear controllers are rarely found in industry mainly because of their complexity as well as the difficulty in obtaining non-linear models. Performance monitoring methods are more often than not based on obtaining some sort of control benchmark. Setting a non-linear control benchmark will be just as complex as the implementation of non-linear controllers and therefore not common practice in industry.

An example of a monitoring technique that was adjusted to compensate for non-linearity is principal component analysis (PCA). Normal PCA is a linear technique that is used to reduce process dimensionality and to visualise the dynamic movements of process states for fault detection purposes. Researchers like Zhang, Martin, and Morris (1997) have identified the need to adapt the linear PCA technique to a non-linear technique with better results.

2.6.4 Performance evaluation of closed loop data

Numerous performance evaluation and monitoring techniques go along with dynamic modelling and characterisation. Techniques to determine models are often intrusive to normal operation. A successful performance monitoring structure should be able to function on normal closed loop data alone with no unnecessary plant tests. This is why statistical process control (SPC) techniques are often employed seeing that most of the methods are non-intrusive and provides performance assessment based on routine operating data.

Another indication of good performance monitoring structures are that they work well over any period of evaluation. This means it works when pure regulatory control conditions exist or when set-point changes occur. The latter is very difficult seeing that when a set-point change is applied to a process it moves from one state to another, with different control objectives to which the performance structure must adapt. To make a performance structure independent of periods of operation is a definite implementation difficulty.

2.7 Origins of performance degradation

The causes of bad performance are numerous and can be inherent to a particular control loop itself, to the plant design, to interaction, etc. Some of the common causes of poor performance are mentioned below:

- Field instruments - Field instruments form the basis of the monitoring structure and should at all times be in operation and accurate. Without proper measurement the control hierarchy will be ineffective. The performance structure should be able to pick up measurement inaccuracies like calibration faults etc.

- Process design - The process design places an inherent limit on performance. It determines an optimum operating state that cannot be improved by any control system. It is of utmost importance to have a well designed process to ensure good performance.

- Control system design - The control system design should always be considered while the process design is performed and vice versa. This will ensure that the states that were meant to be attained are actually obtained and kept there by the control system in general plant operation.

- Loops not in normal mode - Often control loops and APC algorithms get implemented with wonderful results but are not maintained over time. The reason for this is that the plant changes over time because of process parameter changes that cause different operating conditions. The operator then sees that the control is not doing its job and switches the control loops to MANUAL instead of going through the right channels to fix the problem. The performance structure should cater for this by detecting poorly tuned loops automatically and letting the correct maintenance parties know about possible defects in the control structure.

- Manipulated variables - Manipulated variables are often a large source of variance and poor performance. They can get saturated due to poorly tuned loops and inaccurate scaling. Also, if the controller is too aggressive, actuator wear is accelerated, which causes valve stiction that hampers good control.

- Abnormal situations - When plants cross certain operating limits they may go into abnormal modes such as trip conditions or shut down procedures, etc. If a particular unit is in a start-up or a shut-down mode it not only means that the particular unit's performance is bad; it means that the other units downstream also gets affected together with their performance. The performance structure should be implemented as a tool to reduce the time the process spends in an abnormal situation.

The performance structure should be able to detect bad performance and locate the sources of poor performance mentioned above.

In the case of model predictive controllers the performance is obviously directly related to the base layer performance. So the origins of performance degradation are the same as were mentioned above. Model inaccuracy is however also a big issue with MPC and model validation should form part of a performance monitoring structure.

2.8 Performance monitoring and the general optimisation problem

Performance monitoring and general optimisation of chemical processes are closely related. This is true for both design stage and on-line operation optimisation problems. Design stage application of performance monitoring and evaluation is discussed in more detail in chapter 3. The mathematical representation of optimisation problems can directly be related to performance evaluation type problems.

2.8.1 Representing the optimisation problem mathematically

Biegler and Grossmann (2004) have combined most types of optimisation problems in one diagram that is shown in figure 2.3.

The mathematical representations that are shown in figure 2.3 can be divided into problems that only contain discrete variables or problems that contain continuous variables or problems that have both. The representation pertaining to discrete systems is

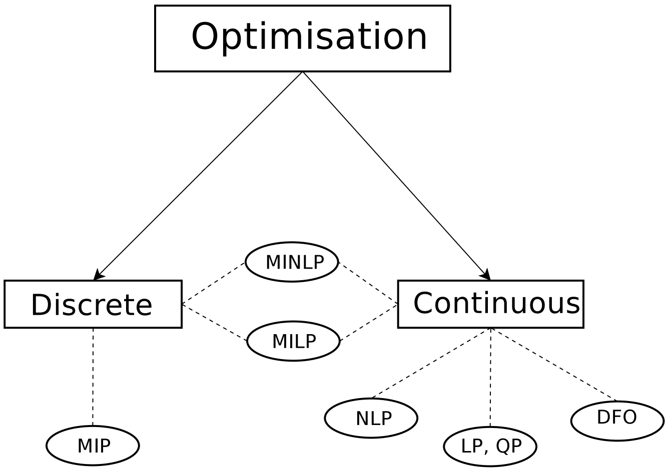


Figure 2.3: Mathematical representations of optimisation problems (Biegler & Grossmann, 2004).

usually the mixed integer programming (MIP) approach. Continuous systems are represented by means of either linear programming (LP), non-linear programming (NLP) or derivative free optimisation (DFO). The problems represented in figure 2.3 are independent of the solution method used to solve the actual problem (Biegler & Grossmann, 2004).

If there are both discrete and continuous variables in the problem formulation a combination of both MIP and LP is used. All the above-mentioned methods can be represented in the algebraic form of the MIP problem. The general algebraic form of a MIP problem is shown in equation 2.1 (Biegler & Grossmann, 2004).

$$min\ Z = f(x,y)\ s.t. \begin{cases} h(x,y) = 0 \\ g(x,y) \leq 0 \\ x \in X,\ \ y \in \{0,1\}^m \end{cases} \tag{2.1}$$

In equation 2.1, $f(x,y)$ is the objective or cost function that needs to be minimised, $h(x,y) = 0$ are the performance equations of the process that indicate production rate, utility usage, etc. In other words $h(x,y)$ are the equations that model the process. The constraints on the optimisation problem are given by $g(x,y) \leq 0$, which will determine feasible solutions to the optimisation problem. The variables x and y are the process variables that need to be optimised. x refers to continuous variables (usually state variables), while y refers to the discrete variables such as scheduling of equipment and seasonal operating changes. y is generally restricted to take values ranging between 0 and 1.

Equation 2.1 corresponds to the MINLP problem if the functions are non-linear and there are both continuous and discrete variables, if the functions are linear, however, it will become a MILP problem. If only continuous variables are present it is a LP or NLP problem depending on whether the functions are linear or not. The general MIP form given in equation 2.1 contains steady-state models. Important extensions on MIP are to include dynamic models and to optimise the system dynamically as well as to include uncertainty in the process models and optimisation constraints (Biegler & Grossmann, 2004).

2.8.2 Relating performance monitoring to optimisation

There are definite similarities between the general optimisation philosophy and a performance monitoring structure. They both have an ultimate goal of making changes to the plant (control structure or process design) to minimise some objective function. The objective function can usually be related to more efficient production and therefore to more profitable plants. This is not always the case seeing that optimisation and performance are subject to constraints like regulations, design limitations, product demand, etc.

This is why a mathematical representation of performance evaluation will fit in nicely into the MIP problem that is shown in equation 2.1. This mathematical representation fits into the general process monitoring methodology where comparisons with the base case are made as well as in the diagnosis and solution steps.

CHAPTER 3

Design stage applications

Performance evaluation in design stage application is a handy tool to compare competing process and/or control design configurations. The techniques that are discussed in this chapter illustrate some of the current work that is being done to consider and quantify the inherent limitations of particular process designs. The foundation of performance in process design is to ensure proper control and process design integration. The performance will indicate the inherent optimum of a design and how easily this optimum will be maintained by the control system if at all. The off-line type of measures discussed in this chapter also fits into the on-line performance structure seeing that these techniques can be applied to rectify problems with performance highlighted by continuous on-line type performance monitoring.

3.1 Operability indexes according to operating spaces

A novel technique to quantify controllability in terms of operating spaces has been proposed by numerous researchers and the basic technique is discussed in this section. The publications include Georgakis, Uztürk, Subramanian, and Vinson (2003), Vinson and Georgakis (2000), Uztürk and Georgakis (2002) and Subramanian and Georgakis (2001). The operability measure quantifies the inherent ability of a process to move from one steady state to another and to reject any of the expected disturbances in a timely fashion with limited control action available. This measure should consequently be independent of the controller type used.

The technique uses four operating spaces as a method to define an operability index. The available inputs to any process can only vary over certain ranges. These ranges were denoted the available input space (AIS). A corresponding available output space (AOS)

can then be determined using the AIS and the model of the process. However a desired output space (DOS) also exists which represents the control needs and is defined at the start of the design of a process. The union of the DOS and the AOS then indicate the area that should be attainable with control while the areas that don't coincide indicate uncontrollable operating points. The operating space concept is illustrated in figure 3.1. In figure 3.1 the grey area indicates the union of the DOS and AOS. This area indicates

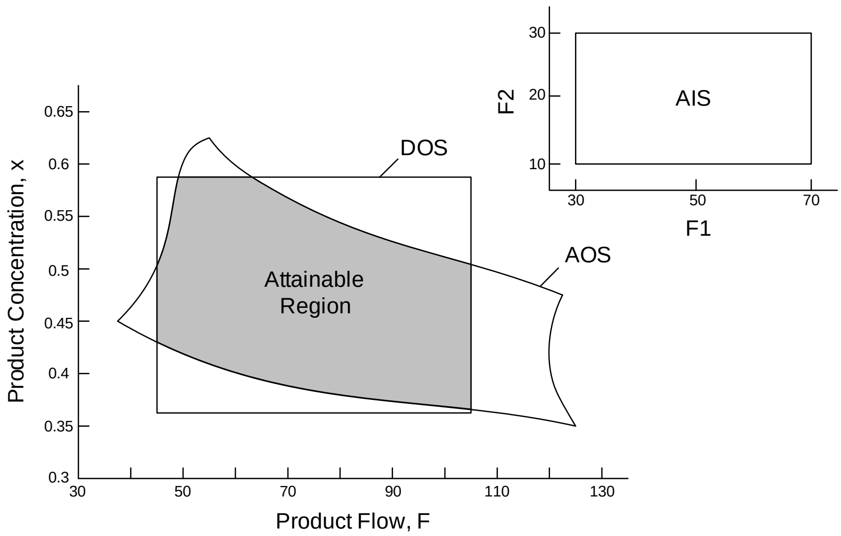


Figure 3.1: Controllability according to operating spaces (Georgakis et al., 2003).

the operating points that can be obtained by the control system by taking into account the plant design (plant model).

3.1.1 Steady-state operability

If we consider all the predefined spaces as steady state operating regions we can calculate both servo and regulatory operability indexes as follows.

The servo operability index

Firstly we define a servo operability index in the output space in equation 3.1 by using the AOS obtained by using the entire AIS (all possible values of u) and the disturbances at their nominal values (d^{NN}).

$$servo : OI_y = \frac{\mu[AOS_u(d^{NN}) \cap DOS]}{\mu[DOS]} \tag{3.1}$$

μ in equation 3.1 is a measure function to calculate the size of union of spaces. If the index is less than one the desired performance will not be met seeing that some desired outputs will lie outside the attainable output region.

The servo operability index can also be calculated in the input spaces as is shown in equation 3.2

$$servo : OI_u = \frac{\mu[AIS \cap DIS_y(d^N)]}{\mu[DIS_y(d^{NN})]} \tag{3.2}$$

In equation 3.2 the opposite route is followed to compute the OI. This time the DIS is determined as a function of all the desired outputs in the DIS, with the disturbances at their nominal values. Computing the OI in terms of the input spaces can be used for new plants or existing ones where changes are being considered. The index is an indication of the extent to which the desired inputs are covered by the available ones. If the index is smaller than one, changes to the design of the plant need to be considered.

The regulatory operability index

Operability indexes to predict the regulatory operability of a process can also be defined. Before regulatory operability can be defined we need to define another operating space which is the expected disturbance space (EDS). The EDS is not only the space that defines all external disturbances acting on the process but also contains all uncertainties in the process model. The OI can now be defined as in equation 3.3 with DIS the input space needed to compensate for all expected disturbances defined in the EDS and the plant at its nominal set-point (y^{NN}).

$$regulatory : OI_u = \frac{\mu[AIS \cap DIS_d(y^N)]}{\mu[DIS_d(y^{NN})]} \tag{3.3}$$

Alternatively the regulatory OI can be determined by making use of a tolerable disturbance space (TDS) together with the range of expected disturbances (EDS). These disturbance operating spaces can then be used as in equation 3.4 to calculate a regulatory OI. The TDS is determined by finding the region of disturbances that can be compensated for by the available inputs (AIS) while the outputs remain at their nominal values (y^{NN}).

$$regulatory : OI_d = \frac{\mu[TDS_u(y^{NN}) \cap EDS]}{\mu[EDS]} \tag{3.4}$$

As was the case before, if the OI in equation 3.4 is smaller than one then some redesign work is necessary.

The overall objective is to reject all expected disturbances and to reach all the set-points in the DOS. First we need to define a union of the desired input spaces, each calculated in the output space for all expected disturbances. Or equivalently we can calculate a desired input space in the expected disturbance space for outputs. This total

DIS is mathematically represented by equation 3.5.

$$DIS = \bigcup_{y \epsilon DOS} DIS_d(y) = \bigcup_{d \epsilon EDS} DIS_y(d) \tag{3.5}$$

The DIS is therefore determined by compensating for regulatory and servo changes. Now that we have defined a total DIS the total operability index can be calculated using equation 3.6.

$$OI = \frac{\mu[AIS \cap DIS]}{\mu[DIS]} \tag{3.6}$$

Again, an OI near zero is bad while close to one is good.

3.1.2 Dynamic operability

A dynamic OI based on the steady-state concepts discussed in the above sections was also formulated. This was done by first defining dynamic operating spaces and from these spaces operability indexes are calculated (Georgakis et al., 2003).

3.2 A performance index in terms of surge volumes

A quantitative controllability index in terms of surge volumes was developed by Zheng and Mahajanam (1999). This index describes a plant design in terms of a total surge volume that would be needed to make a plant controllable. Conversely the volume can be an excess capacity that can be taken away to still leave the plant controllable.

The controllability index, υ, was defined as follows by Zheng & Mahajanam (1999):

"Consider a continuous plant. Given a set of disturbances, a set of constraints, a set of control objectives and a control system, the dynamic controllability index, υ, is defined to be the smallest additional surge volume required to meet all the control objectives and constraints dynamically for all of the disturbances."

The basic idea behind the index is that it indicates the minimum surge capacity that is needed to meet all the control objectives. The theory behind surge volumes is motivated by Buckley (1964) and his dynamic process control concept that implies that poor product quality control can be overcome by installing sufficiently large surge capacities. Although this concept is very old it is still relevant seeing that surge volumes provide a universal quantification of controllability. The surge volumes does not get implemented on the real plant they are just a method of controllability assessment, especially seeing that modern day plants try to reduce excess buffer capacities as much as possible.

Properties of υ

From the definition, the following properties were defined by Zheng & Mahajanam (1999):

- A process is controllable if and only if $\upsilon \leq 0$. If $0 < \upsilon < \infty$, then the process can be made controllable by installing additional a surge volume of υ. If $\upsilon < 0$, then a surge volume of υ can be removed from the design without affecting controllability.

- υ is bounded if and only if the steady-state control problem is feasible and the closed system is stable.

- If $\upsilon > 0$, then the cost associated with achieving controllability equals the cost associated with installing surge tanks with a total volume of υ (although some surge capacity may also be removed). This method is a very rough method to predict the cost of making a process controllable. Process design modifications and controller cost should also be considered. This method can however be seen as an estimate of the upper bound on the cost.

- The transfer function for a surge tank is, $\frac{1}{\tau s+1}$, where the time constant, τ, is proportional to its volume, V. The proportionality constant depends on the process variable variation that needs to be damped for instance product composition, or flow rate.

To illustrate the above definitions and concepts of υ an illustrative example is given by Zheng & Mahajanam (1999). This index is most useful in the later design stages to rank different process and controller designs according to economics.

3.3 Defining an operability framework to evaluate competing designs

Swartz (1996) has also used the philosophy of the integration of plant design and closed loop performance. This philosophy was the main motivation for development a systematic approach to evaluate competing designs.

CHAPTER 4

Performance monitoring and evaluation tools

This chapter highlights the current state of the art of performance monitoring technologies. The technologies are the tools that are used to generate the measures of how effective a plant is operating by indicating how close the plant is from its inherent optimum. The inherent optimum is a defined benchmark for comparative operation evaluation. Most of the current methods are based on statistical manipulation of data.

The tools discussed indicate whether the process is operating sub-optimally and provide suggestions as to the possible causes for this behaviour and how they are to be corrected.

4.1 Controller performance concepts

The general concepts of controller performance are well known and the basic principle is clear; the dynamic responses (CV's) of the plant have to be maintained at their set-points for as long as possible. This is achieved by the control system. This is done by making controllers more aggressive, by making them change manipulated variables (MV's) quickly and with large steps based on what the deviation of the controlled variables (CV's) from the setpoints are. This is good for quick responses but usually makes responses much more oscillatory and may lead to instability. The process and control system also become much more sensitive to external disturbances and noise. So there is always a trade-off between quick responses and oscillatory behaviour (instability). Controllers are therefore always tuned on the conservative side as to leave a margin of robustness to cater for uncertainty in inputs (disturbances, noise, etc.).

4.1.1 Traditional methods

Traditional methods are well known and can be used successfully to indicate the speed of responses. Most introductory literature to the field of process control mention these methods (Stephanopolous (1984), Luyben (1990), Marlin (1995) and Seborg, Edgar, and Mellichamp (2004)). Some of the methods that are commonly used include:

- Rise time
- Overshoot
- Settling time
- Integral of the error squared (ISE)
- Decay ratio, etc.

These methods work well but it is usually still the experience of the control engineer that needed to interpret the measures in order to determine how well a controller is performing. In most cases all of the above measures have to be considered to provide enough information to give an indication of a single controller's performance. A reasonably sized plant usually has thousands of controllers. To apply the above measures to each of the controllers on a regular basis is a near-impossible task seeing that the measures do not provide a straight quantitative value. Experienced control engineers need to do the evaluation and they are usually limited in numbers on most plants.

The techniques that probably provide the best overall indication of controller performance are the error integral techniques (ISE) seeing that they cover both poles of the general controller optimisation problem (robustness and speed of response). The reason for this is because they consider the deviation of the CV from set-point over time. If the standard deviation of a CV or control error is considered one sees that it is similar to the error integral methods seeing that it is a quantity that provides an indication of the deviation of a CV from its set-point over time. For instance if we have a very robust process the CV is slow to reach set-point which will cause large standard deviation, while a process that is quick and oscillatory will also have a large standard deviation. Therefore the optimisation problem of minimising the standard deviation or variance is ideal seeing that it finds the optimum between a too robust, slow response and a quick oscillatory (unstable) process. That is why the current trend is to look at variance of CV's as a direct indication of controller performance. Some of the leading researchers in the field of variance isolation and statistical process control include Harris (1989), Huang and Shah (1999) and Thornhill, Oettinger, and Fedenczuk (1999).

4.2 Statistical process control (SPC)

The main application of SPC is to improve the quality and productivity of a process. This is achieved by a set of tools that is used to increase the capacity and stability of a process by considering and reducing the process variability (do Carmo C. de Vargas, Lopes, and Souza, 2004).

Variance in process variables has two sources, namely (Shunta, 1995):

- Common variance (inherent variance) - Frequent, short-term, random disturbances that are inherent to the process. These types of disturbances can normally not be compensated for by the control system and limit the control system to a certain lower bound of variance that is the optimum. The variability can be reduced by changing the processing equipment or making use of other control instruments. The control system is not useless against these types of disturbances seeing that without the control system there would have been much larger deviations, but the control system cannot eliminate the variance completely.

- Special causes (controllable variance) - Regular load disturbances that are larger, less frequent and more specific. These can normally be catered for by the control system seeing that they are predictable. Examples of special causes are equipment failures, raw material feed inconsistency, steam pressure fluctuations, catalyst decay, etc.

Most of the techniques that are discussed in this section assume that the data are normally distributed and are only useful if the data are random. This is not always the case for signals that are captured. To cater for this, data are usually transformed by fitting a curve (mostly linear) and using the residual data for further analysis purposes.

4.2.1 Control charts

Statistical process control charts are used to indicate changes or variations in the operation of a process. The chart consists of a graphical display of some quality characteristic which is measured or calculated from a sample and then plotted against the sample number (Montgomery, 1985). Some process parameters can be estimated from the charts and the process capacity can then be inferred from that. The charts could also indicate possible changes to operation to reduce variability. The type of control chart that is used is dependent on the process characteristic that is to be monitored and the way the sample are gathered (do Carmo C. de Vargas, Lopes, and Souza, 2004).

Control charts can also be used as modelling tools to estimate certain process output parameters such as the mean, standard deviation, percentage nonconforming, etc. (do Carmo C. de Vargas et al., 2004) (Montgomery, 1985).

A control chart usually contains three horizontal lines that provide guidelines on the quality of the characteristic to be monitored. One of the lines is known as the centre line and is calculated as the average of the quality characteristic in the presence of common, inherent variance. The centre line indicates the value from which the quality characteristic should deviate as little as possible from. The two other lines indicate control bounds that should not be violated by the quality characteristic. The lines are known as the upper control limit (UCL) and the lower control limit (LCL). A typical example of a control chart is shown in figure 4.1. The control limits can be used as a guide for controller

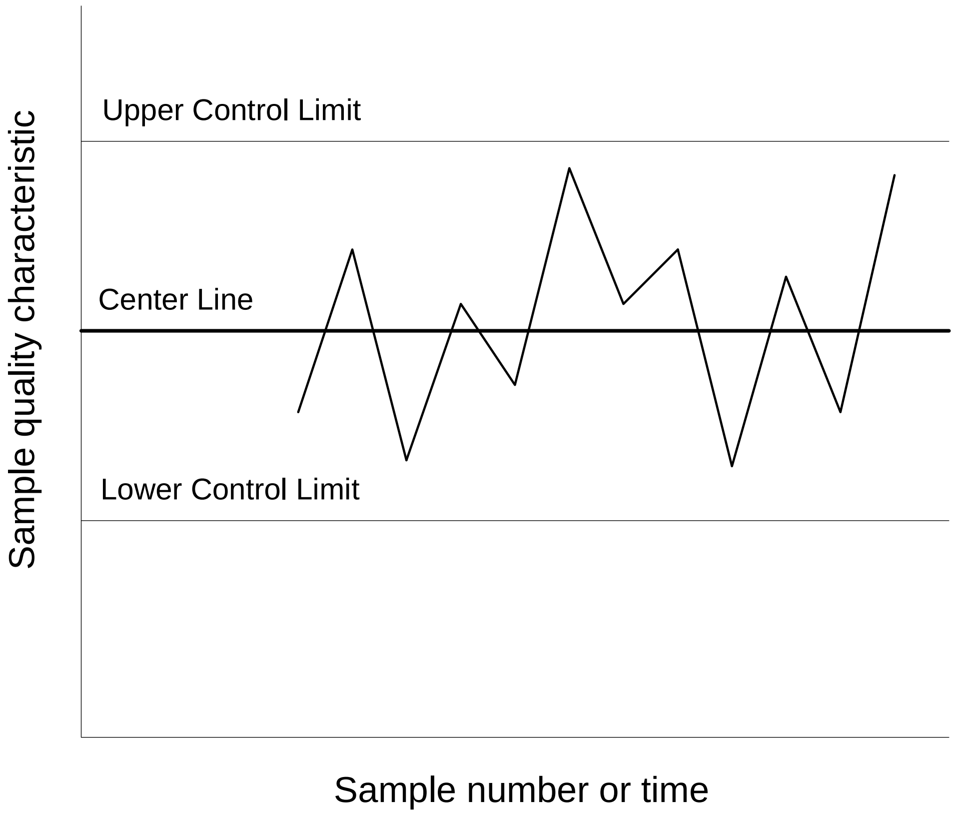


Figure 4.1: Control Chart

performance. If points lie outside the limits the process is said to be out of control and investigation is necessary to find the cause of this violation. It does not mean however that if no violations of the control limits occur that the system is in perfect control. If, for instance, several consecutive samples lie above or below the centre line and do not violate the limits, it usually means that the control is suboptimal and that there can be some improvement. It can be concluded therefore that the system is performing optimally when the quality characteristic will have a random distribution around the centre line with no violations of the control limits (Montgomery, 1985).

Control chart application in the chemical processing industry and in the aforementioned monitoring structure has definite potential; especially seeing that it is a direct representation of process variability. In the application of the chart the control limits will represent the area of allowable variance. The allowable variance is that which originates from the plants inherent design. This is the variance that is uncontrollable. A control chart is therefore a good indication of how far a plant is operating from its inherent optimum by looking at the number of control limit violations.

From the above discussion it is clear that a number of factors need to be considered in the design of a control chart, which include:

- The actual quality statistic to be sampled or calculated. This statistic is the value to be plotted against sample number. This can for instance be the sum of the product flows from the plant multiplied by their purity or the error signal of a PID loop.
- The control limits that should not be violated.
- The sampling frequency. In other words, how often does one need to compare the quality against the limit.
- The sample size, determined by how many samples should be taken to give an appropriate representation of operation.

Modified Shewhart chart

Control charts can be represented by a general model that is shown in equations 4.1 to 4.3. This general model was first proposed by Shewhart (1931) and control charts that are based on it are referred to as Shewhart control charts.

$$UCL = \overline{w} + k_1\sigma_w \tag{4.1}$$

$$Center\ line = \overline{w} \tag{4.2}$$

$$LCL = \overline{w} - k_1\sigma_w \tag{4.3}$$

The equations that describe the general model are for a sample statistic w that measures some quality characteristic, with a mean of $\overline{w}$ (target value) and a standard deviation of σ_w. It has become a general standard in industry to use a value of 3 for the constant k and is generally known as the 3σ limit (Montgomery, 1985) (NIST-SEMATECH, 2005).

The classical Shewhart approach is to compare samples to the target value and if the absolute value of the difference is larger than a constant times the standard deviation the process is said to be uncontrollable. The modified Shewhart chart takes exactly the same approach with the only difference being that the samples are compared to the standard error of the time series of the process (Kramer and Schmid, 1997). The modified Shewhart criterion is shown in equation 4.4.

$$|w_t - \overline{w}| > k_1 \sigma_{w_t} \tag{4.4}$$

The problem that can be identified with Shewhart type charts is that they consider single samples. The control criteria are sample specific and do not consider past sample values or sequences of sample values (do Carmo C. de Vargas et al., 2004). That is why Shewhart type charts are almost always used in conjunction with other control charts that are discussed in the subsequent sections.

Cumulative sum chart (CUSUM)

The formula that is used to calculate a CUSUM chart is shown in equation 4.5 (do Carmo C. de Vargas et al., 2004).

$$C_i = \sum_{j=1}^{i} (\overline{x}_j - \overline{x}_{tot}) \tag{4.5}$$

The variable $\overline{x}_j$ is the average of the variable that is being sampled at sampling instant j, while $\overline{x}_{tot}$ is the desired target value of the variable, x (almost always the mean of the population). The reason for the mean of x being used at each sample point is that the formula is usually applied in the manufacturing industry where more than one sample is taken at each sampling instant. In the chemical process control environment continuous single samples are taken, for example for measuring temperature. The average of x will therefore only be the measured value at a particular time.

From equation 4.5 it is clear that if the cumulative sum, C_i, is near zero for large samples, it means that the actual value is deviating around the desired value. If C_i is increasing with samples it means that the process is moving away from $\overline{x}_{tot}$ by increasing and when C_i is decreasing with increasing samples, x is moving away from $\overline{x}_{tot}$ negatively.

The normal horizontal control limits are not usually applied to CUSUM charts. The method normally used is the V-mask. The V-mask can be applied visually or in tabular form. Figure 4.2 shows the application of the V-mask to determine the control limits (NIST-SEMATECH, 2005). As can be seen from figure 4.2 the two lines are plotted with equal gradient magnitudes but opposite in sign. This causes a V-shaped plot that is superimposed on the control chart. If any of the sample values lie outside the V the system is said to be out of control. To plot the boundaries, two parameters are needed together with the origin as the last sample value plotted. The two parameters that are usually defined is the rise distance, h and slope of the bottom arm, k (see figure 4.2).

The graphical use of the V-mask has become outdated and the V-mask is almost exclusively implemented in a tabular or spreadsheet format. To generate the table the

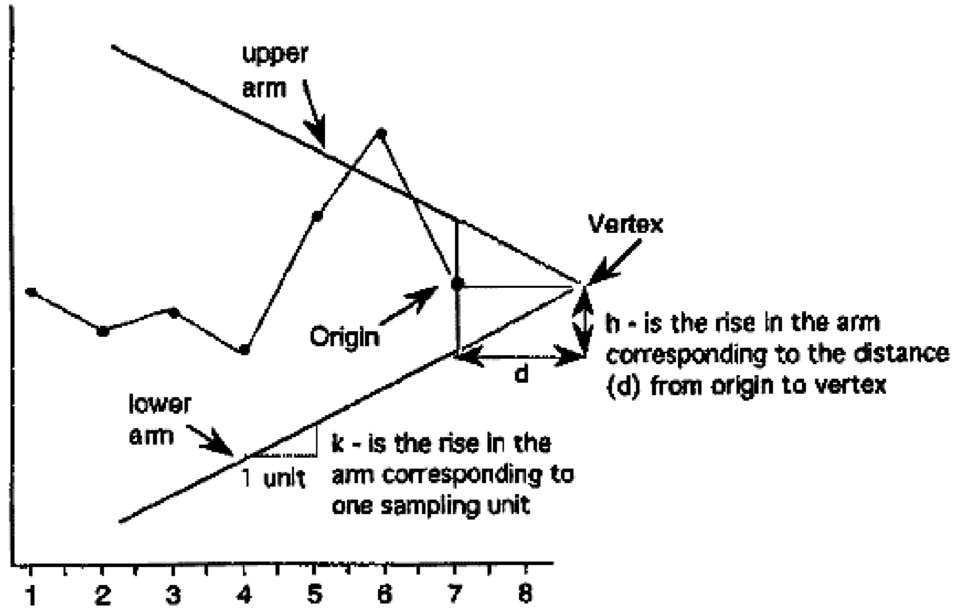


Figure 4.2: The V-mask control limits for a CUSUM chart (NIST-SEMATECH, 2005).

same two parameters, h and k, are specified which can then be used to calculate the quantities in equations 4.6 and 4.7.

$$S_{hi}(i) = max(0, S_{hi}(i-1) + x_i - \hat{\mu}_0 - k) \tag{4.6}$$

$$S_{lo}(i) = max(0, S_{lo}(i-1) + \hat{\mu}_0 - k - x_i) \tag{4.7}$$

For the first sample, $S_{hi}(0)$ and $S_{lo}(0)$ is taken as zero. The criterion then becomes, if $S_{hi}(i)$ or $S_{l0}(i)$ exceeds h, the process is deemed to be out of control.

CUSUM charts have shown to be more sensitive to small changes in the sample value mean compared to Shewhart type charts and of course it considers past sample instants through the cumulative sum (NIST-SEMATECH, 2005).

Exponentially weighted moving average (EWMA)

As is the case with the CUSUM chart, the EWMA also considers past samples to judge the control state. It applies an exponential weight to past samples and then combines all the past weighted values including the most recent into one EWMA statistic. The unique difference with the EWMA chart is that it applies less and less weight to sample values that are more and more removed from the current sampling instant (NIST-SEMATECH, 2005).

The EWMA can be defined as in equation 4.8 and was first proposed by Roberts

(1959) (do Carmo C. de Vargas et al., 2004).

$$Z_i = \lambda x_i + (1 - \lambda) Z_{i-1} \tag{4.8}$$

To start computing the average, Z_0 needs to be specified and is usually chosen as the process target SP, or the initial process variable average, $\overline{x}$. λ is a constant between 0 and 1 that needs to be specified and is usually chosen to be in the region of 0.2 to 0.3. Note that the closer λ is to 1 the less weight is applied to the past sample values. If λ is equal to 1 the EWMA chart reduces to the classic Shewhart chart.

The centre line and the control limits are given by equations 4.9 to 4.11 (NIST-SEMATECH, 2005).

$$UCL = \overline{x} + L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \tag{4.9}$$

$$Center\ line = \overline{x} \tag{4.10}$$

$$LCL = \overline{x} - L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \tag{4.11}$$

Evaluation and interpretation of the control chart

The common technique to evaluate the performance of control charts is to compute a measure called the average run length (ARL). The ARL is a measure of how long on average we will plot samples before we detect a sample that violates the control limit (NIST-SEMATECH, 2005) (Yang and Makis, 1997).

The ARL should be large if the quality statistic is on target and short if the mean is deviating, in other words if a disturbance has occurred. The ARL can therefore be used to make a decision as to what type of control chart to use for a process. The calculation of the ARL is usually quite involved and different for each control chart type. Tables and graphs (nomographs) are available to read of ARL values for specific chart types (NIST-SEMATECH, 2005).

If we look at the ARL from a control loop evaluation point of view it seems like a good measure of performance. Say that we assume the quality statistic is the deviation from set-point (control error). If the ARL is large it will mean that the loop is performing well and the control error does not violate the control limit frequently. It has to be determined however what control chart will be the best for a PID control loop and when this is established how complex is the ARL calculation.

4.2.2 Statistical distributions

Normal distributions

When a data set is independent and not-correlated the standard deviation and mean of the sample is always the same as the standard deviation and mean of the population for big enough sample sizes. The frequency plot of these data sets are called Gaussian or normal distributions. A normal distribution has the general shape that is shown in figure 4.3 (Shunta, 1995). Figure 4.3 shows the average value of the sample, $\overline{X}$, that is

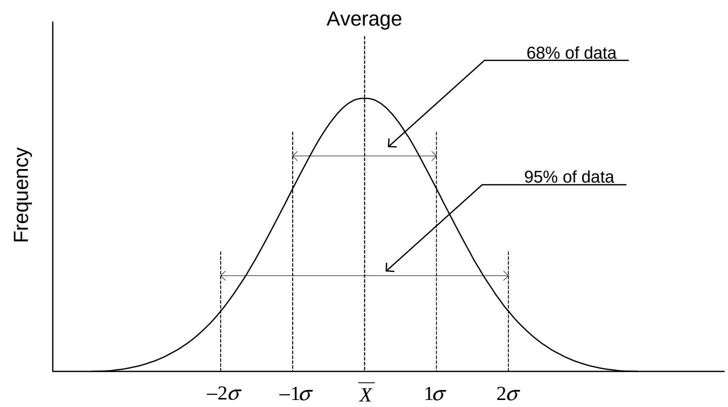


Figure 4.3: A typical normal distribution (Shunta, 1995).

also the symmetry axis of the distribution. The standard deviation, σ, influence on the shape is also shown.

We can make several qualitative assumptions on the quality of control by examining the distribution of the control error or the CV of a normal PID regulatory control loop. If the distribution is normally distributed as discussed above with the mean the set-point for a CV distribution or zero for the control error, the system is under regulatory control. The shape of the normal distribution can however give us further information as to how well the regulatory loop is performing. Obviously the less variation in the loop the better. So the 'thinner' the distribution, the better the control. An example of comparative analysis for different distributions for the same loop is shown in figure 4.4. The value of reducing the variance is also pointed out in figure 4.4. Set-points can be moved much closer to constraints due to the reduced variance. So the process throughput can be increased, yield increased, waste reduced, etc. if the variability is reduced by still keeping the process stable and in safe operation(Shunta, 1995).

A useful tool when working with normally distributed data is the confidence interval approach which enables us to predict with a certain probability what the value of the variable will be at a specific sampling instant. The method is to compute the Z value for a certain value of the variable which shown in equation 4.12.

$$Z_h = \frac{x_h - \overline{x}}{\sigma_x} \tag{4.12}$$

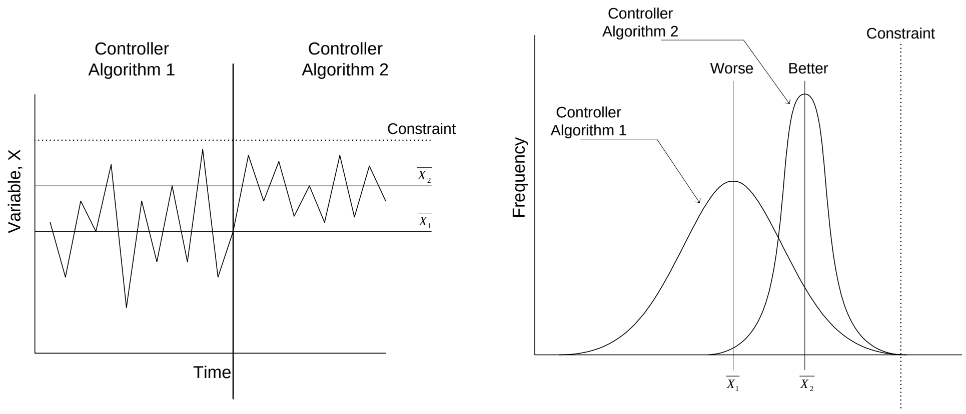


Figure 4.4: Comparing loop performance through distributions.

Equation 4.12 can for example be used to calculate Z_h and with probability tables found in most statistical texts, the probability that the value at a specific sampling instant will lie above or below the value x_h can be determined.

When considering a real process there are various sources of variability that the control system is trying to compensate for. So we should be careful in assuming normality for response data, especially seeing that most statistical measures are based on the assumption that the data is normally distributed. Distributions that are not completely independent appear skewed or may have more than one peak. This usually indicates that there is a problem with the regulatory control system and further investigation is necessary. Non-normal distributions are briefly discussed in the next section.

Non-normal distributions

If a frequency plot of some variable that is under regulatory control, is a non-normal distribution then this usually means that there is some sort of problem with the control system. Possible causes for non-normal distributions are (Shunta, 1995):

- Presence of outliers - Outliers cause normal distributions to have heavier tails. Outliers are not always real and engineering judgement should be used to determine whether they represent real variability. If they are incidental incidents then they need not be considered and can be filtered out. Real outliers on the other hand need to be investigated.

- Measuring elements - Measuring elements cause problems when they are not calibrated for the current operating range or they do not have sufficient sensitivity to measure the correct magnitudes of variable movement. Measuring sensitivity problems will usually cause a distribution with a sharp drop off to the one side.

- Variable characteristics - The process variable may have a physical constraint like a measurement of length or weight that cannot be negative, or a purity above 100%.

Distributions that are affected by these limitations will have a fixed boundary on the one side of the frequency plot.

- Nonlinearity - If a linear PID controller is used to control a nonlinear process the resulting distribution plot will also be skew. The reason for this is that the controller makes proportional changes to the MV depending on the error magnitude. So the signal to the final control element will be the same in magnitude for positive and negative error signals with equal magnitude. The process however, will not react in the same manner. Say a nonlinear process is disturbed by moving a MV by a positive change of magnitude, A. The output difference between the new steady state and the current steady state is C. If the MV is moved in the opposite direction with a value of $-A$ the nonlinear process output steady state difference will not be $-C$. This is the reason why linear controllers (PID) only work efficiently for specific disturbances. Nonlinear processes are compensated for by inverting the non-linear effect of the process through the measuring element to produce a linear output from the measuring element. High purity distributions almost always have skew distributions.

- Final control element - Valve stiction is also a source of non-normal distributions. Valve stiction usually causes the CV to be on either side of the set-point and therefore produces a distribution that is shaped like an upside down bell. A histogram of this type is shown in figure 4.5 (Expertune, 2005) .

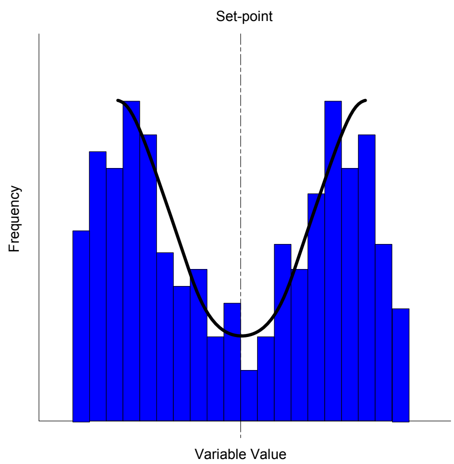


Figure 4.5: An example of an histogram of a control loop containing a valve that exhibits stiction (Expertune, 2005).

There are several ways to quantify the amount of skewness of a distribution. The following three properties of frequency plots can be used as an initial indication of skewness (Albright, Winston, and Zappe, 2002):

- Mean - The mean is the normal algebraic average of the data that have been sampled.
- Median - The median is the "middle" sample number of the population data that has been arranged in ascending or descending order. The median is the middle sample for odd numbered population sizes and the average of the two middle samples for even numbered populations.
- Mode - The mode is the value interval that occurred the most frequent.

These three values are very close to each other for normal distributions, so if they differ by substantial amounts one should suspect that the distribution is skew. A more commonly used technique to test the skewness of data set is to take the third moment around the mean. Equation 4.13 shows the third moment around the mean (Duval, 2005) (NIST-SEMATECH, 2005).

$$skewness = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{(n-1)\sigma_x^3} \tag{4.13}$$

If the skewness calculated in equation 4.13 is positive it means that the distribution has a heavier tail to the right and if the skewness is negative the tail is heavier to the left. A skewed distribution because of variable limitations like weight measurement that cannot go negative will have a heavier tail to the right seeing that the measurement has a lower bound. The skewness for any symmetrical (e.g. normal distribution) distribution is near zero.

A simplified formula for calculation of the third moment is shown in equation 4.14.

$$\gamma_{skew} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{\frac{3}{2}}} \tag{4.14}$$

Another handy statistic to measure the amount of deviation from a normal distribution is the kurtosis formula shown in equation 4.15.

$$kurtosis = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{(n-1)\sigma_x^4} - 3 \tag{4.15}$$

The reason for the second constant term of 3 is to standardise the formula to a result of zero for standard normal distributions seeing that the first term equals 3 for normal distributions. The kurtosis formula provides an indication of how the peaks in the distribution compares to the standard normal distribution. It can also be said that the kurtosis gives an indication of how the gradients of a distribution's side compare to the normal distribution. If the kurtosis is positive it indicates a "peaked" distribution while a negative kurtosis indicates a "flat" distribution.

A simplified formula for measuring the kurtosis is shown in equation 4.16.

$$c = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{(\sum_{i=1}^{n}(x_i - \overline{x})^2)^2} \tag{4.16}$$

To illustrate the use of the skewness measures some well known skew distributions and their corresponding skewness and kurtosis values are shown in figure 4.6 (NIST-SEMATECH, 2005). From figure 4.6 it is clear that for the normal distribution the value

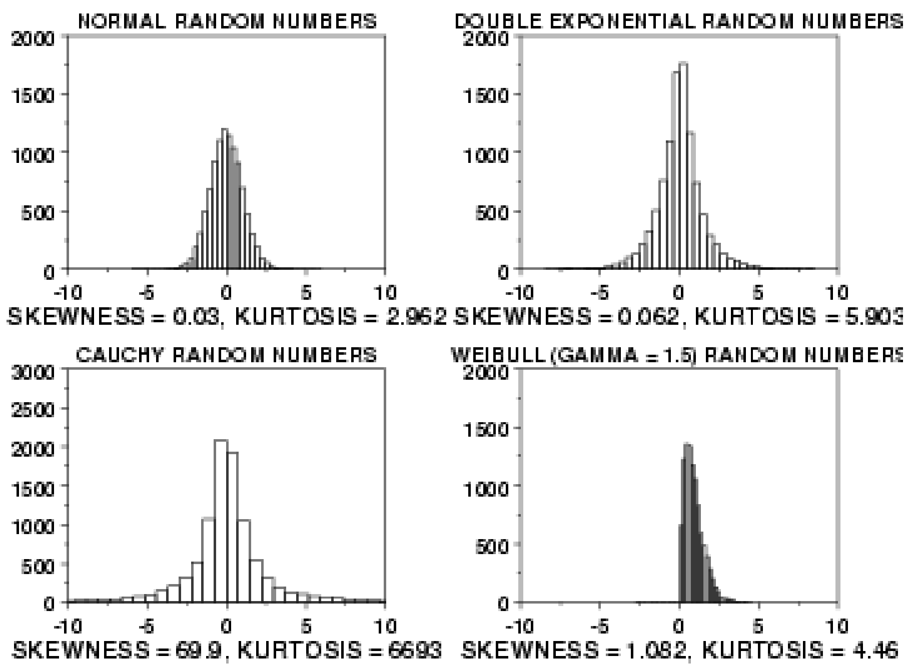


Figure 4.6: Skewed distributions together with their skewness and kurtosis values.

for the skewness is near zero and the kurtosis (first term in equation 4.15) is near 3. For the double exponential distribution the skewness is also near zero which makes sense seeing that the plot is symmetric, but the slopes of the sides are not normal and the plot seems peaked which is related in the larger positive kurtosis value of 5.9. The Cauchy distribution shows elongated heavy tails and has a skewness and kurtosis of 69.99 and 6 693 respectively. The skewness was suspected to be near zero due to the symmetry but in fact has a large positive value. The kurtosis is also very large. These extremely large values are due to the heavy tails of the distribution. The last distribution shown is the Weibull distribution which has a heavier right side. So we suspect the skewness value to be positive. This is indeed the case with a value of 1.08. The kurtosis is 4.46 which indicates a slightly peaked distribution.

If the dataset is skew as well as the residuals of the data there are some transformation techniques that exist to approximate a normal distribution. Various transformations exist for various types of skewness. The power and logarithmic transforms are two examples that are shown in equations 4.17 and 4.18 (Shunta, 1995).

$$T_1(x) = ax^p + b \tag{4.17}$$

$$T_2(x) = c\ log(x) + d \tag{4.18}$$

Now if we want to use the normal distribution probability tables mentioned in section 4.2.2 we can if the transformed distribution is normal by using equation 4.19.

$$Z_h = \frac{T_i(x_h) - T_i(\overline{x})}{\sigma_{T_i(x)}} \tag{4.19}$$

4.2.3 Correlation

To quantify and measure the amount of interaction between two variables various statistical measures exist. Two popular measures that are commonly used are covariance and correlation (Albright et al., 2002). Two variables are said to be correlated if a change in one variable is reflected by a change in the other variable. In terms of control loop interaction, correlations are useful tools to determine the relative amount of interdependency either between two loops, or the relative amount of variability that the loop causes with its own control action.

Another functional aspect of determining correlated variables is that it identifies possibilities to infer variables from each other. It therefore means that only one of the correlated variables has to be monitored to give a representation of all the other correlated ones. This reduces the dimensionality of the general control problem and simplifies it especially seeing that normal control problems are of large dimensions. Inferred variables also enable us to choose which of the variables are to be monitored (Stapenhurst, 2005). The obvious choice are those variables that are inexpensive, easy and repeatable (e.g. temperature measurement as an indication of concentration).

The covariance between two variables can be defined as in equation 4.20.

$$Cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1} \tag{4.20}$$

It is clear from equation 4.20 that the covariance of two variables is the average of the product of the deviation of the variables from their mean. It can therefore illustrate to us whether the variables are at respective sampling instants above or below their mean and by what magnitude. So variables that are completely random will have a covariance close to zero. Variables that are both decreasing or both increasing will have a positive covariance, while variables that are changing in opposite directions will have a negative covariance. The problem with the covariance function is that it is not a dimensionless entity. Therefore, if the numerical value of the two variables differ by a large amount the magnitudes won't contribute the same weight to the function result. Some sort of scaling is necessary (Albright et al., 2002). This is why the covariance function is not a commonly used function but rather the correlation function shown in equation 4.21.

$$C_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \tag{4.21}$$

As can be seen from equation 4.21 the correlation equation is scaled with the standard deviations of the respective variables. The correlation function will always be between -1 and 1. With 1 a perfect positive correlation with both variables changing in the same direction at the same rate and -1 a perfect negative correlation with variables changing in opposite directions while zero indicates no relationship at all.

Cross correlation

The correlation functions found in the process control field correspond directly to the above discussion and is found in discussions on time series analysis. The result of the correlation function in equation 4.21 is known as the cross correlation coefficient and is a single value that describes the linear relationship between the variables for a specific sample delay. (Box, Jenkins, and Reinsel, 1994) (NIST-SEMATECH, 2005) (Blevins, McMillan, Wojsznis, and Brown, 2003)

The cross correlation function will be applied in this investigation as an indication of the relationship between two particular process variables. The correlation coefficient for a CV/MV pair will for example be calculated with equation 4.22.

$$C_{MV/CV_k} = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(MV_i - \overline{MV})(CV_{i+k} - \overline{CV})}{\sigma_{MV}\sigma_{CV}} \tag{4.22}$$

The cross correlation coefficient in equation 4.22 contains an extra index, k, which indicates what the sample time delay should be for comparison. In equation 4.22 the MV sample at instant i is compared to the CV value at k time instants later, $i+k$. The index k is known as the coefficient lag and causes comparison of the variables at different sampling instants. The coefficient can be calculated for all time lags smaller than the sample size, n, but this is unnecessary seeing that the effect of the MV on the CV will only be seen for the time that the process takes to reach steady state. It can therefore be estimated that the correlation coefficients should be calculated for all lags smaller than the longest time to steady state for the process. Variables usually only show correlation at specific values for k if at all. To illustrate the complete correlation between variables the cross correlation coefficient is plotted against the values for the lag. The figure will have peaks for lags that show correlation. An example of a cross correlation coefficient plot is shown in figure 4.7. The cross correlation coefficient can be calculated by considering any

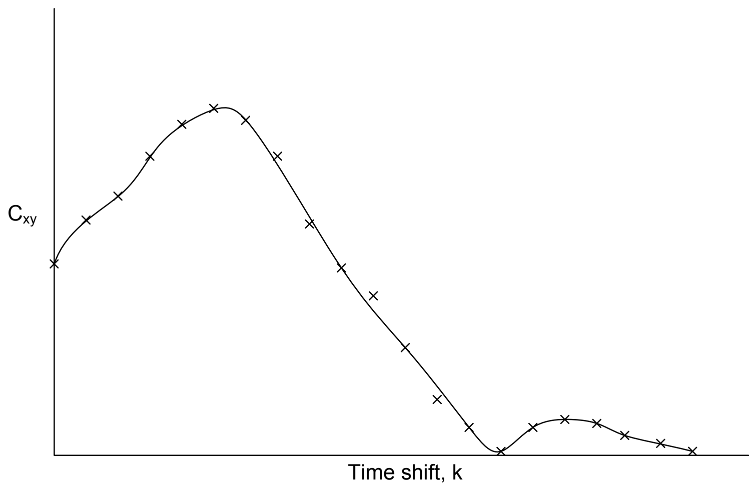


Figure 4.7: The cross correlation coefficient plot vs. the lag.

two variables. Two main interaction characteristics can be observed from the correlation plot. The first is the amount of delay involved before the one variable's movement is noticed in the other variable by considering the lag where peaks are observed and the second is the "strength" of the interaction by considering the heights of the peaks.

If a CV/MV pair of the same feedback loop is considered the effect of that feedback control algorithm is illustrated in the coefficient plot. Firstly the lag plot will give an indication of how much deadtime is contained in the loop (peak occurs at the lag value that estimates the deadtime) and secondly it indicates how much of the CV shifts/moves are due to the MV in that loop. We can conclude that peaks close to 1 or -1 will indicate strong control action while smaller peaks will indicate weaker control due to external disturbances on the loop.

To give an indication of possible causes of interaction and variance in a particular CV, the cross correlation can be done for CV/MV pairs that are not linked by feedback or any other disturbance/CV pair that is suspected to have an effect on the CV. The cross correlation function will indicate the strength of interaction and is a good way to locate the cause of performance degradation. This is useful to identify opportunities for advanced control applications like decoupling or model predictive control (MPC). The obvious prerequisite for disturbance/CV cross correlation is that the disturbance has to be a measured variable.

The benefit of the cross correlation interaction analysis technique is that the calculations can be done under closed loop conditions with no process modelling necessary. An estimate of the process deadtime would however be a handy parameter to know before the correlation coefficient plot is performed seeing that it will provide an estimate of where to expect peaks on the plot and therefore confirm whether the initial deadtime estimate is correct.

Autocorrelation

Autocorrelation is a form of correlation that is another important function that should be considered when doing statistical analyses. The difference between the cross correlation function and the autocorrelation function is that the autocorrelation considers a single variable and the cross correlation two variables. A variable is autocorrelated when a variable's value at a specific time can be related from another value of the same variable at some other sample period. The auto-correlation is calculated with a coefficient that is similar to the cross correlation coefficient and also varies between -1 and 1. The coefficient is an indication of the randomness of a variable dataset. Random datasets have a autocorrelation coefficient that is close to zero for all time lag coefficients. Most statistical analysis techniques assume that the data are not autocorrelated and the coefficient plot is a useful tool to test for randomness in data. One method to calculate the autocorrelation coefficient is shown in equation 4.23 (NIST-SEMATECH, 2005).

$$R_k = \frac{\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.23}$$

The coefficient in equation 4.23 is for a time lag of k time sample periods. To test for randomness of data, consecutive samples (a time lag of $k = 1$) are sufficient. A coefficient value of $R_k = 0$ will indicate a random dataset.

Another tool that is frequently used to check for randomness of data that is similar to a autocorrelation is the a lag plot. The lag plot is a graphic that plots samples from a univariate dataset that is separated by a constant sample difference. For example the lag plot for dataset Y for a lag of 1 will be a plot of $y(i)$ vs. $y(i+1)$. This will give an indication of the randomness of the data as well as indicate outliers in data. An example of a lag plot for a dataset that exhibits non-random cyclic nature and contains a few outliers is shown in figure 4.8 (NIST-SEMATECH, 2005). The cyclic nature that was picked up in the lag plot is also visible when the autocorrelation coefficient plot is done in figure 4.9 (NIST-SEMATECH, 2005). As mentioned before, the autocorrelation coefficient tells us about the randomness of the data and if we look at this from a process control point of view, it will tell us about the dependency of the data on time. If we consider a controlled variable signal that is at set-point with only random noise

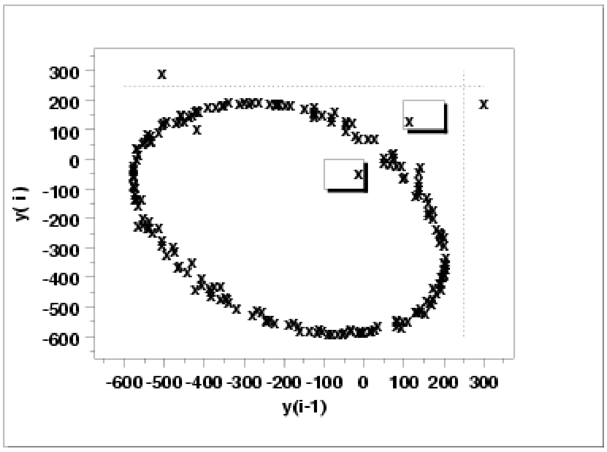


Figure 4.8: The lag plot for y for a sample lag of 1 (NIST-SEMATECH, 2005).

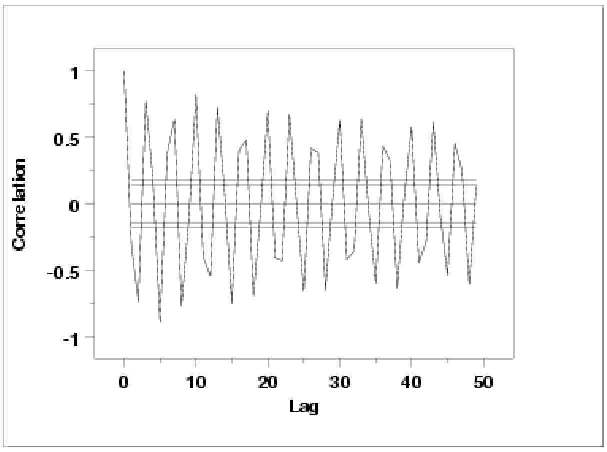


Figure 4.9: The autocorrelation plot for a cyclic non-random dataset (NIST-SEMATECH, 2005).

adding variance to the signal the autocorrelation coefficient will be zero. If the signal is however transient or even oscillatory or unstable the coefficient will be non-zero. The auto-correlation coefficient can therefore be used to determine whether the variance that is contained in a signal is due to random noise or due to time dependent behaviour of the process (load disturbances). Huang and Shah (1999) provided a typical example of implementation of the autocorrelation function as a performance measure by considering a control system for the Wood-Berry column. In literature they usually refer to a random variable (autocorrelation zero) in statistical control.

The autocorrelation coefficient can be used as an early fault or disturbance detection measure to determine whether a loop is deviating from set-point due to factors that is not random noise. The autocorrelation is once again a useful measure seeing that it can be applied to regular closed loop data.

4.3 Frequency analysis

4.3.1 Power spectrum analysis

The Fourier transform is a mathematical tool that is used to investigate the frequency behaviour of a continuous function or a discrete set of sampled values. Seeing that the monitoring structure developed in this study makes use of sampled signals in the discrete domain the discrete Fourier transform will be implemented. Equations 4.24 and 4.25 show the discrete Fast Fourier transform (fft) for the sampled signal x.

$$Y(k) = \sum_{j=1}^{n} x(j)\omega_n^{(j-1)(k-1)} \tag{4.24}$$

$$\omega_n = e^{\frac{-2\pi i}{n}} \tag{4.25}$$

The same type of "normalisation" or de-trending of data has to be performed as was mentioned in the section on statistical control (section 4.2). This is typically done by taking the difference between each sample and the first value or mean of the dataset. Straight lines or polynomial fits are also used and the transform is then applied to the residual of the curve fitted.

The magnitude plot of the Fourier transform gives an indication of the power of the frequency components of the measured signal and can be calculated with equation 4.26. The plot is referred to as the power spectral density (PSD).

$$P_{yy} = \frac{Y \times conjugate(Y)}{Resolution} \tag{4.26}$$

As can be seen from equation 4.26 the power density, P_{yy} is calculated by multiplying the

FFT of the variable, Y, with its complex conjugate and dividing it with the resolution used to perform the fft. The magnitude, P_{yy}, is then plotted against frequency to determine frequency components. The region of interest in the plot is the low to medium frequency range seeing that peaks in the high frequency range are usually related to noise. The plot is done for all frequencies smaller than the sampling frequency because no useful information with respect to process dynamics can be obtained for frequencies higher than the sampling frequency. It is important to have a sampling frequency that is high enough to capture all useful process dynamics for monitoring purposes.

The PSD will show if there is non-random oscillatory behaviour hidden in the noise by showing large peaks at frequencies where oscillations are found while the height of the peak is an indication of the amplitude ("power") of the oscillation.

The example of the non-random cyclic behaviour that was detected by doing the lag plot and the autocorrelation plots in section 4.2.3 is also shown in the spectrum plot in figure 4.10 (NIST-SEMATECH, 2005).

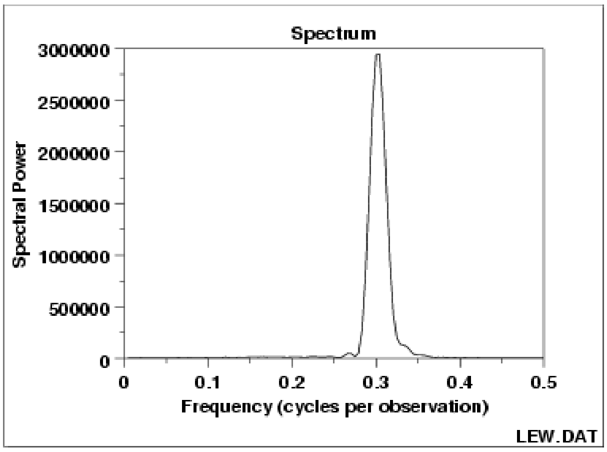


Figure 4.10: Power spectrum for cyclic dataset.

MATLAB has numerous built-in functions to compute a PSD. They all perform PSD calculations but use different algorithms that differ according to the method of filtering, execution, windowing, etc. The choice of a PSD calculation function depends on the input vector that the PSD is performed for. In this investigation none of the standard PSD functions will be used. The built-in fft function of MATLAB will be used to perform the discrete fast Fourier transform of signals. If this function is used, one of the input parameters to the function is the resolution. The fft algorithm is executed in such a way that it converges the quickest for resolutions that are factors of two. It is preferable to choose a resolution that is a factor of two and greater than the length of the considered signal vector. If the resolution is smaller than the signal vector it gets truncated and if the resolution is larger the vector gets packed with zeros for the extra elements of the vector. After the fft has been calculated the PSD is calculated with equation 4.26.

The power spectrum is also done from regular closed loop data and therefore very functional to point out oscillatory behaviour in a loop on a continuous on-line basis without disrupting normal operation. The power spectrum is a useful tool to help identify possible performance shortcomings like over tuned loops or cyclic disturbances, etc.

The cross spectral density (usually denoted P_{xy} for signals x and y) is a function that provides an indication of the relative power between two signals. It can be used to give an indication of variable interaction. The function can be applied as another test together with the cross correlation function to determine variable interaction.

The coherence function is a "normalised" function of the cross spectral density and is given by equation 4.27.

$$C_{xy} = \frac{|P_{xy}|^2}{P_{xx}P_{yy}} \tag{4.27}$$

The coherence function cater for differing peak heights in the two individual power densities. If the one peak is small relative to the other the interaction is not easily detected in the cross spectrum density, however, dividing by the individual densities provides a scaled function for better interaction detection.

4.3.2 Oscillation detection

Detecting oscillation is an important tool in a performance monitoring structure and was briefly touched on in section 4.3.1. Oscillations in a loop is not only a source of increased variance but is also related to increased MV movement that causes wear which eventually relates to stiction and hysteresis. Oscillations usually have a negative effect on operation seeing that it causes increased energy consumption, waste of raw material and increased variance in production rates and quality. The frequency response techniques discussed in section 4.3.1 are good methods to determine oscillatory behaviour but they are computationally intensive if functional mathematical software is not available.

Faulty control valves and actuators are the most common causes of oscillatory behaviour in control systems while other causes include bad controller tuning and external oscillations affecting the controller. Final control elements cause oscillations because of stiction (stick-slip) motion of the element due to high friction because of wear (Hägglund, 1995). It is interesting to note that wear in final control elements is increased when a loop is oscillatory and faulty final control elements is also a cause of oscillatory behaviour to cause a snowball effect that can quickly get out of control. It is therefore important to detect, diagnose and rectify oscillatory behaviour as soon as possible to minimise wear on final control elements. Oscillatory behaviour is rectified by doing regular valve maintenance, controller retuning or using feedforward control. It is important to note that controller retuning will not take away the oscillatory behaviour if it is a final control element problem or external oscillatory disturbances (Hägglund, 1995).

Detecting oscillatory behaviour by the frequency of control error sign changes

Hägglund (1995) proposed a robust and easy to use oscillation detection method that can be applied to normal regulatory closed loop data. The procedure is based on a simple principle of determining what the frequency of control error sign changes are, which is directly related to oscillatory behaviour. The region of the frequencies that we want to consider is those close to the cross over or ultimate frequency. These are the oscillations that cannot be compensated for by the controller seeing that the frequency of the disturbances is too fast and filtering can't be used seeing that the frequency is too low.

The first step in the procedure is to determine whether a disturbance has occurred or not, in other words, to determine whether the CV has deviated from set-point significantly. This can be done by computing the integrated absolute error (IAE) shown in equation 4.28.

$$IAE = \int_{t_{i-1}}^{t_i} |e(t)|dt \tag{4.28}$$

The integration interval is the time between control sign change instants, t_{i-1} and t_i. Equation 4.28 is for the error defined as the difference between the measured value and the set-point for controllers with zero offset (PI controller). For controllers with an offset (P controller) the error should be defined as the difference between the measurement and the mean of the measurements.

The IAE is a functional method for determining significant disturbances seeing that it incorporates the magnitude of the disturbance as well as its duration. So high frequency disturbances will have smaller IAE values than lower frequencies due to the small integration interval for high frequencies and the larger interval for lower frequencies. So instrument noise will normally have small IAE values which we don't want to consider. Disturbances of larger magnitude will have large IAE values while smaller disturbances will have small IAE values. Now the question arises what value for the IAE is sufficient to indicate that a disturbance has occurred.

Hägglund (1995) proposed that we consider the error to be a pure sine wave with amplitude a and frequency ω. The half cycle of the sine wave will then represent a load disturbance. The sine wave is therefore a series of load disturbances. From the definition of the IAE, the limit in equation 4.29 was proposed.

$$IAE_{lim} \leq \int_{0}^{\frac{\pi}{\omega}} |a\ sin(\omega t)|dt \tag{4.29}$$

Seeing that we are only interested in disturbances that occur in the low to medium frequency ranges, frequencies up to the ultimate frequency, ω_u should be detected. Hägglund (1995) found that an appropriate value for the amplitude will be 1% of the value around which the oscillation is occurring. This will be 1% of the set-point for zero off-set controllers otherwise the average value of the CV. With the above values for the specified parameters the IAE_{lim} become as in equation 4.30.

$$IAE_{lim} = \frac{2a}{\omega} = \frac{2(0.01)(SP)}{\omega_u} \tag{4.30}$$

If the ultimate frequency is not known the integral time constant of the controller, τ_i, should be used as an estimate of the ultimate frequency, $\omega_i = \frac{2\pi}{\tau_i}$. The estimated frequency will roughly be the same as the ultimate frequency if the controller is well tuned. The load detection can now be performed by computing the IAE for a period between control error sign changes and then to compare it with the IAE_{lim}. If the IAE is greater than the limit a significant disturbance is said to be detected.

After defining the disturbance detection procedure the question arises, how do we implement this load detection procedure to detect oscillatory behaviour? This is done by defining an evaluation time, T_{sup}, in which the number of detected disturbances are counted and if they exceed a certain amount, n_{lim}, an oscillation is said to be present in the loop. The evaluation time needs to be carefully chosen. Hägglund (1995) proposed a lower limit on the evaluation time and is shown in equation 4.31

$$T_{sup} \geq \frac{n_{lim} T_u}{2} \tag{4.31}$$

Oscillations in real signals are often not pure sinusoids at their ultimate frequency but rather ragged functions of lower frequencies than the ultimate. Therefore the evaluation time needs to be much larger than the lower limit. Hägglund (1995) proposed evaluation times of 50 times the ultimate period of oscillation, or, if the ultimate frequency is unknown the integral time constant of the controller.

If the evaluation time principle is followed, as mentioned above, the disturbances have to have a time stamp to keep track of how many disturbances have occurred in the evaluation time. Hägglund (1995) proposed an exponentially weighted function that can be implemented as a "counter" to determine the number of disturbances in the evaluation time period. This function is shown in equation 4.32.

$$x = \lambda x + load \tag{4.32}$$

Equation 4.32 is executed each time an error sign change occurs after the IAE has been calculated for the error sign change interval. If the integral is big enough to signify a disturbance the value for the parameter, *load*, is set equal to 1 else it is zero. λ is a constant that is related to the evaluation time by equation 4.33.

$$\lambda = 1 - \frac{\Delta t}{T_{sup}} \tag{4.33}$$

Where Δt is the sample period of the signal. The weight adds less and less significance to disturbances that occurred further and further back in the past. The oscillation detection algorithm is now complete and a final decision on oscillatory behaviour can be made by considering the criterion in equation 4.34.

$$x > n_{lim} \tag{4.34}$$

If the criterion in equation 4.34 is true then there is oscillatory behaviour for the proposed evaluation time. As a summary of the oscillation detection algorithm of Hägglund (1995) the diagram in figure 4.11 can be considered.

4.4 Performance benchmarks

4.4.1 Minimum variance performance (MVC) benchmark

Performance benchmarks are necessary to determine what the optimum operating state of the process is. This optimum is the operating state containing the inherent variance that cannot be compensated for by controllers. The operating optimum is then used as a benchmark to do comparative evaluation of the current operation. Various methods exist and most of the methods originate from the minimum variance theory that was developed by the popular work done by Harris (1989).

Standard deviation is a common and understandable way to evaluate process variability as has been seen in the preceding performance evaluation methods. The MVC benchmark tries to quantify the minimum variance of a process due to common causes and is an inherent property of the process. The real total variance of the process (common and special causes) is then calculated and compared to give an indication of performance.

Minimum variance control

The MVC benchmark concept was developed by Harris (1989) in his now classic article on control loop performance assessment. The original MVC benchmarking method uses closed loop data from a linear process under normal linear time invariant feedback control. This means that no extra perturbations are necessary to determine performance; only routine closed loop operating data is used. The method uses a univariate time series model to estimate the number of whole periods of dead time which is used to calculate the optimum capability of the controller. The MVC method for benchmarking is a completely non-intrusive technique for performance assessment which is precisely the reason for its popularity.

MVC can be expressed as the control law that minimises the cost function shown in

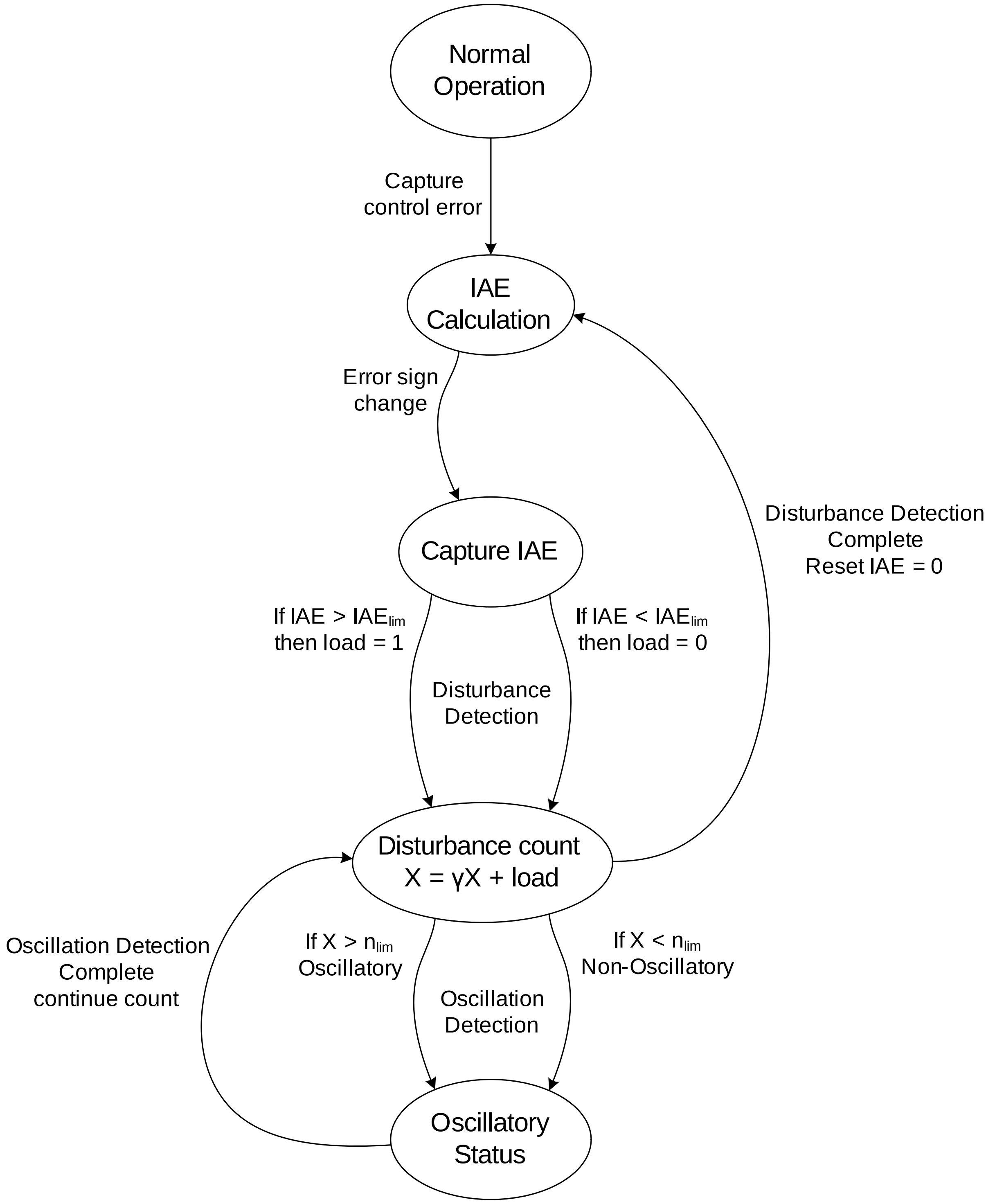


Figure 4.11: The oscillation detection algorithm of Hägglund (1995)

equation 4.35 (Åström and Wittenmark, 1984).

$$J_{mvc} = Ey^2(k) \tag{4.35}$$

y in equation 4.35 refers to the controlled variable or the process output. The scaling should be chosen such that $y = 0$ when the process is at the desired set-point or steady-state. E is a weighting that penalises the deviations. As one can see from equation 4.35 the cost function is the equation that describes the deviation of the output from its desired value and therefore contains the variance that is to be minimised by the controller.

Equation 4.35 only acts as an indication of the variance at one specific time instant, k. To take a larger evaluation horizon into account the equation can be rewritten as in equation 4.36.

$$J_{mvc_\infty} = \lim_{n\to\infty} E\frac{1}{n}\sum_{1}^{n} y^2(k) \tag{4.36}$$

n in equation 4.36 is the number of sampling instants of the horison over which the variance is calculated.

The next step is to determine what the controller algorithm should be to achieve the MVC objective. The formulation of the controller was done in accordance with a document compiled by (Hong, 2005).

In order to determine what the controller algorithm should be, we consider a general controlled autoregressive moving average model of the closed loop response (equation 4.37).

$$A(q^{-1})y(t) = q^{-D}B(q^{-1})u(t) + C(q^{-1})x(t) \tag{4.37}$$

In equation 4.37 y is the output, u is the manipulated input. x is a random external disturbance on the control loop with standard deviation, σ_x and A,B and C are polynomials in the backward shift operator q^{-1} shown in equations 4.38 to 4.40.

$$A(q^{-1}) = 1 + a_1q^{-1} + ...a_{na}q^{-na} \tag{4.38}$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + ...b_{nb}q^{-nb} \tag{4.39}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + ...c_{nc}q^{-nc} \tag{4.40}$$

Already we foresee that control won't be completely free from variance seeing that the manipulated variable takes D time instants to affect the process output, y.

The next step is to develop the controller algorithm that minimises the cost or goal function shown in equation 4.35. Let's consider the controller in equation 4.41.

$$u(t) = -\frac{G(q^{-1})}{B(q^{-1})F(q^{-1})}y(t) \tag{4.41}$$

G and F are also polynomials in the backward operator q^{-1} and are shown in equation 4.42 and 4.43.

$$G(q^{-1}) = g_0 + g_1 q^{-1} + ...c_{ng}q^{-ng}\ ; \quad ng = max(na-1, nc-1) \tag{4.42}$$

$$F(q^{-1}) = 1 + f_1 q^{-1} + ...f_{D-1}q^{-(D-1)} \tag{4.43}$$

B is the same as for the closed loop model in equation 4.37. Now we specify the polynomials G and F to satisfy equation 4.44.

$$C = AF + q^{-D}G \tag{4.44}$$

If we rearrange the closed loop model (equation 4.37) and use equation 4.44 for C the following equation for y at time instant $t + D$ can be formulated as in equation 4.45.

$$y(t+d) = \underbrace{\frac{B(q^{-1})F(q^{-1})}{C(q^{-1})}u(t) + \frac{G(q^{-1})}{C(q^{-1})}y(t)}_{\hat{y}(t+D|t)} + Fx(t+D) \tag{4.45}$$

As can be seen from equation 4.45 the first two terms is the prediction of the output at D time instants in the future based on the present and past knowledge of the variables u and y. This prediction is represented as $\hat{y}(t + D|t)$. The last term is however not known as it lies in the future and is known as the output prediction error. This is the inherent variance of the loop and can't be compensated for by the controller algorithm. The first two terms ($\hat{y}(t + D|t)$) are available and can be compensated for by the MV.

Now if we consider the objective function in equation 4.35 to be minimised for the value of the output at D time instants in the future it can be represented as in equation 4.46.

$$J = Ey^2(t+D) = \underbrace{E\hat{y}^2(t+D|t)}_{controllable} + \underbrace{\sigma_x^2(1 + f_1^2 + ...F_{D-1}^2)}_{uncontrollable} \tag{4.46}$$

The reason why the objective function simplifies to a function of the predicted output and the standard deviation of the disturbance alone is because the disturbance is assumed to be an independent random variable. The task is now to minimise the cost function by reducing the predicted output to zero (see equation 4.47) by changing the MV.

$$\frac{BF}{C}u(t) + \frac{G}{C}y(t) = 0 \tag{4.47}$$

The minimum variance controller that results is the same controller proposed at the start of the of the analysis (equation (4.41)). We have therefore proved that the suggested controller is the minimum variance controller for the proposed process model and the polynomials can be solved recursively by solving for corresponding coefficients by using equation 4.44.

If we consider the objective function in equation 4.46 it is clear that it consists of two parts, one being the controllable variance part of the output signal and the other the uncontrollable part because of the time delay of the process. The aim is now to determine what the variance is of the uncontrollable part of the output which will be the theoretical lower bound performance benchmark. Then we determine the normal variance of the output signal and compare it to the benchmark. The application of the benchmarking method is discussed in section 4.4.1.

Application of the minimum variance benchmark

For the application of the minimum variance benchmark it is useful to consider the power spectral density of the normal output of a control loop shown in figure 4.12 (Clegg, Xia, and Uduehi, 2005). The uncontrollable variance pointed out at frequencies higher than

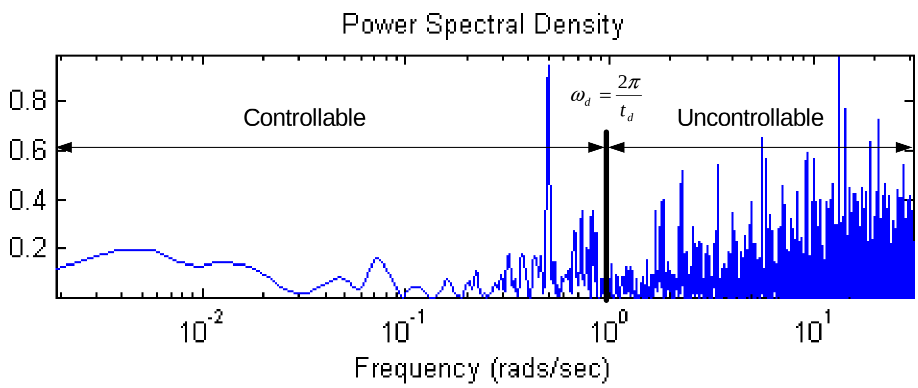


Figure 4.12: Power spectrum of the controlled variable or the control error.

the deadtime frequency, ω_D, needs to be quantified. An estimation of this minimum variance in one dead-time increment is shown in equation 4.48. It considers the mean square successive difference of variable, x.

$$S_{cap} = \sqrt{\frac{\sum_{i=2}^{n}(x_i - x_{i-1})^2}{2(n-1)}} \tag{4.48}$$

Shunta (1995) refers to this quantity as the “capability standard deviation”. We will make use of S_{cap} as an indication of the inherent standard deviation that cannot be compensated for by the control algorithm.

Now that we have defined the minimum variance we need the current operational variance which can be calculated with the standard deviation, S_{tot}, of the output, x, shown in equation 4.49.

$$S_{tot} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \tag{4.49}$$

As was mentioned before if we are considering single loop performance, x, will typically be the CV value or the control error. S_{tot} is a quantified indication of what the current operating condition for the control loop is. If the variable that is to be monitored is not a regularly measured variable and is usually inferred (for example, concentration, via temperature) the additive property of variance can be used. This property enables us to use the independent variables to approximate the inferred variable variance if we know what their corresponding mathematical relationship is (Shunta, 1995).

Since we now have calculated the real variance, S_{tot}, as well as the capable variance, S_{cap}, we can perform a comparison between the two. Fellner's formula can be used to achieve this and is shown in equation 4.50.

$$S_{fbc} = S_{cap}\sqrt{2 - \left[\frac{S_{cap}}{S_{tot}}\right]^2} \tag{4.50}$$

The variable S_{fbc}, is an indication of the standard deviation from minimum variance control (Shunta, 1995).

To combine all the standard deviations mentioned above into one index that gives the loop performance as a percentage of its full potential, equation 4.51 can be used (Blevins et al., 2003).

$$PI = 100\frac{S_{fbc} + s}{S_{tot} + s} \tag{4.51}$$

In equation 4.51, s is a sensitivity factor that can be used to control how sensitive the PI is to changes in the standard deviations. The closer the value is to 100% the better the loop is performing. The PI should not be used as a stand-alone index but rather used to compare periods of operation for possible fault detection, like poorly tuned parameters or uncertain models used in MPC algorithms, etc.

To summarise the minimum variance benchmarking method we conclude that good performance relative to minimum variance control indicates that process variability is inherent and cannot be reduced by controller action except for APC techniques like preventative feedforward (ff) control. This highlights the fact that a poorly designed process can not be "fixed" by controller action and accentuates the need for proper control and process integration in the design stages of a process. Poor performance relative to minimum variance control indicates that the controller is limiting process performance. It is important to note however that poor performance relative to minimum variance is not necessarily a bad thing and when this occurs more advanced measures need to be considered to confirm the initial suggestion. Remember that minimum variance is only concerned with control error variance reduction and does not consider control effort. The minimum variance benchmark is still a handy tool to indicate what the global lower bound on variance is and if controllers are operating close to minimum variance. Very little can and should be done to improve operation by controller tuning if this is the case (Huang & Shah, 1999).

4.4.2 Extended benchmarking methods

There exist numerous other benchmarking methods that will be shortly be discussed in this section. Most of the methods proposed are used to cater for the fact that the minimum variance benchmark is more often than not an unrealistic control state that reduces output variance or the control error, but does not take into account what the control effort is. The benchmarks that follow make the benchmark condition a much more realistic and attainable state. The minimum variance benchmark was used as a comparitive measure in this study, so the methods discussed here are for the sake of completeness and future implementation on the considered process discussed in chapter 5.

Historical data

The basis for historical benchmarking is that the current operation is compared to periods of operation when the plant was doing well. Researchers that have proposed methods for this type of benchmarking include Huang & Shah (1999) and Gao et al. (2003).

User specified benchmark

The minimum variance or optimal H_2 law is a global optimum for control loop assessment and is not necessarily the most desirable in practice. It can be said that if a controller's performance is close to minimum variance then the controller is performing well and performance can't be improved much, but when the controller is far from minimum variance the controller is not necessarily performing badly and a higher level of performance assessment is necessary. If, for instance, we want to put some constraints on the controller output, like specifying a minimum overshoot or a specific settling time, the minimum variance case won't necessarily be the appropriate benchmark. User-defined benchmarks are implemented for these cases where there are some practical issues involved which will prohibit the controller from achieving minimum variance. The way it is done is by specifying a desired transfer function for the closed loop response that is to be achieved. Let's consider equation 4.52 given in the book by Huang & Shah (1999).

$$y_t|_{user} = (\underbrace{f_0 + f_1 q^{-1} + ... + f_{D-2} q^{-D+2} + f_{D-1} q^{-D+1}}_{F} + q^{-D} G_R) d_t \tag{4.52}$$

Where $y_t|_{user}$ is the user specified benchmark output response and d_t is the external disturbance acting on the process. Equation 4.52 becomes the optimal minimum variance response if the last term is zero. The last term includes the transfer function, G_R, which contains the desired properties specified by the user.

Since the user specified dynamics are available, the variance of the user specified output can be compared with the actual current variance from the measured output. The performance index in equation 4.53 can then be used to do comparative performance analysis.

$$PI_{user} = \frac{\sigma^2_{user}}{\sigma^2_y} \tag{4.53}$$

Extended prediction horizon

For the minimum variance benchmark an estimate of the deadtime is needed, but in some cases it may be difficult or expensive to determine the deadtime. The deadtime is then used as the prediction horizon which determines the amount of time the deviations in the response is not affected by the controller (see equation 4.46). This is true for ideal situations but controllers have dynamics and cannot reduce variance instantaneously and exhibits some sort of dynamic behaviour before it settles out. The extended prediction horizon caters for this non-ideal case. The extended prediction horizon is based on setting an extended settling time (prediction) criterion for the closed loop response (Harris et al., 1999). Thornhill, Oettinger, and Fedenczuk (1999) proposed the minimum variance index is plotted for a number of deadtime intervals and then use the extended prediction horizon as the time for the deadtime where the index settles out and does not vary rapidly. It should be noted however that following methods like benchmarking on historical data and extended horizons relies on engineering judgements and experience which is subjective and not necessarily the optimum.

Generalised minimum variance

Minimum variance control usually goes along with aggressive control action which is bad for input saturation considerations as well as MV movements and so forth. The generalised minimum variance approach of Grimble (2002) penalises MV movements as well as unrealistic control errors to yield a more realistic and practical benchmark for performance assessment. Figure 4.13 shows a univariate feedback control loop together with the performance assessment configuration to illustrate the assessment methodology (Thornhill et al., 1999). From the figure it is clear that there are weighing factors, P_c and F_c applied to the control error and control signal to lessen the aggressiveness of the controller and to make the control more realistic. The general minimum variance controller is then defined as the controller that minimises the objective function shown

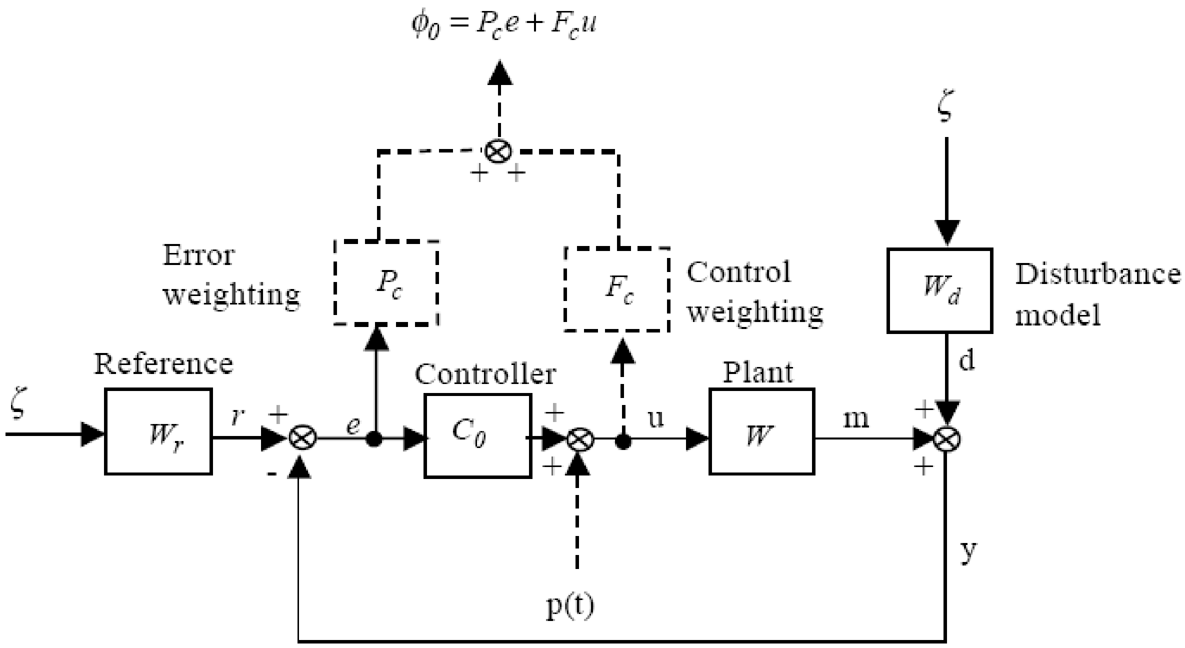


Figure 4.13: The generalised minimum variance methodology (Grimble, 2002).

in equation 4.54.

$$J = E\{(P_c e(t) + F_c u(t))^2\} \tag{4.54}$$

Linear quadratic control (LQG)

Similar to the general minimum variance technique, the LQG method of benchmarking is also a higher level of performance measurement that incorporates the amount of control effort in the performance evaluation. The method determines the minimum variance of the process output if there is an upper limit on the variance of the process input. Figure 4.14 (taken from Huang & Shah (1999)) provides a good indication of the LQG optimisation problem. The cost function that is considered to obtain the trade off curve is shown in equation 4.55.

$$J(\lambda) = E[y_t^2] + \lambda E[u_t^2] \tag{4.55}$$

If λ is varied in equation 4.14 various optimum solutions for $E[y_t^2]$ and $E[u_t^2]$ can be obtained. These solutions then represent trade-off curves that identify the achievable or acceptable operating region. Let's say we have an upper bound on the controller output (process input) variance, α, then we can identify what the realistic minimum variance of the controller input (process output) could be without violating the input constraint (see figure 4.14). This realistic minimum variance of the process output could then be used as a benchmark for comparative performance analysis.

The problem with LQG control is that it is a much more computationally intensive

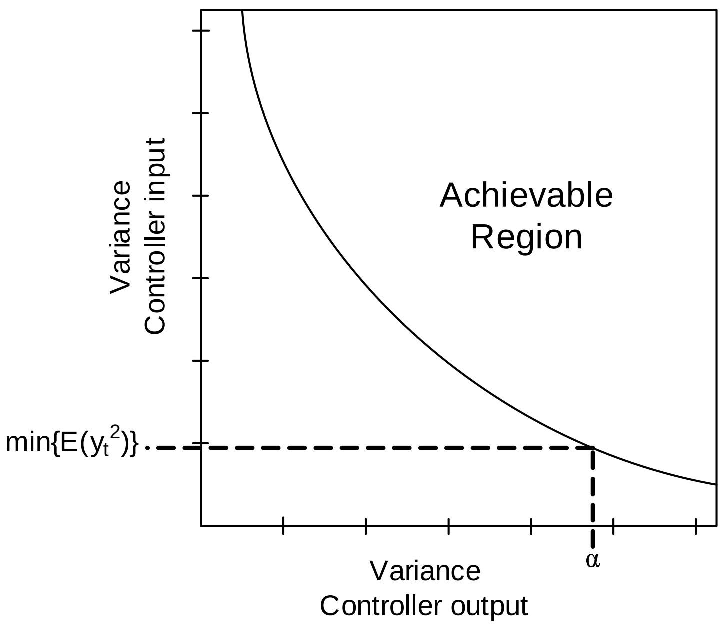


Figure 4.14: The trade off curve for the solution to the LQG problem with the benchmark condition $E[u_t^2] \leq \alpha$ and $min\{E[y_t^2]\}$ (Huang & Shah, 1999).

algorithm when compared to the afore-mentioned minimum variance extensions. The LQG technique as well as the general minimum variance technique are good options if actuator wear and MV saturation is a problem for the controller. It should be kept in mind that when implementing these techniques extra measurements are needed which means more data acquisition and more instruments which has financial implications.

Optimal PID

The method is based on minimising the process output variance by considering PID control only. To do this a disturbance model is necessary to calculate the controller optimal controller parameters. Seeing that most controllers on a chemical processing plant are PID controllers optimal parameters for them would be a great benefit, especially seeing that they are actual achievable benchmarks. The plant models used to solve the optimisation problems have to be accurate for the benchmarking to be realistic and this is not always possible. The variance for the optimal PID loop may be attainable and more functional than the minimum variance controller but seeing the ease of implementation and non-intrusive nature of the MVC technique it may be preferred.

Various techniques exist to determine what the optimal PID parameters are and some are discussed in the following literature Edgar and Ko (2004), Huang (2003) and Skogestad (2003).

4.5 Principal component analysis (PCA)

As mentioned before chemical processes are complex multidimensional systems that are very difficult to visualise. Principal component analysis (PCA) is a multivariate statistical technique to reduce the dimensionality of a particular dataset to such an extent that it can be visualised on normal 2D or 3D plots. The PCA technique reduces the number of variables to the minimum number that completely describes the process by identifying the principal drivers that govern where the process is heading. By performing a PCA we determine the true dimensionality of a process by considering only the variables (referred to as the principal components) that originate from the analysis (Zhang, Martin, and Morris, 1997). The common problem with performance analysis usually is that there is enough data but the amount of data makes it difficult to extract the right information to determine how the plant is performing. By performing a PCA we identify envelopes of normal operation on the reduced dimensional plots. Malfunction or bad performance can then be detected by plotting current operation on the reduced dimensional axes and then determining where the current operation is relative to the normal operational envelope.

4.5.1 Linear principal component analysis

As was the case with the techniques discussed in section 4.2 data can vary a lot depending on their units. To cater for this, the raw data are usually standardised to have all the data in the same units. This can be done by dividing all the variables by their standard deviation. Once the data is standardised it is projected on a lower dimensional subspace. This reduces the number of variables which makes the analysis of plant information easier. The general method of linear PCA is that a straight line is fitted through the data and residuals of the data and then projected onto these linear combinations to reduce dimensionality.

The linear combination that approximates the standardised raw original data represents the first principal component. The linear combination of the data that describes the largest amount of variability in the data defines the principal components. If the data is denoted by a matrix, x, with each column representing a single variable then the variability can be represented by equation 4.56.

$$t_1 = p_1^T x \qquad with \qquad ||p|| = 1 \tag{4.56}$$

The loading vector is p_1 and determines the direction of the linear approximation and t_1, the score, is the coordinates of a point on the linear approximation.

The second principal component can then be calculated from the residual data matrix,

E, calculated from equation 4.57.

$$E_1 = X - p_1^T t_1 \tag{4.57}$$

The second principal component can be calculated from equation 4.58 and represents the linear combination that describes the largest amount of variability for the residual matrix.

$$t_2 = p_2^T E_1 \tag{4.58}$$

The score, t_2, and the loading vector, p_2, have the same properties as for the first linear approximation only now they refer to the residual linear approximation.

To calculate further principal components the procedure is continued by obtaining the residual of the second approximation and then the residual of the third, etc. The number of principal components will be equal to the number of variables in the original data set. Usually not all the principal components are calculated seeing that the variance in the data gets reduced each time a residual is analysed. This is because there are only a few variables that are the principal drivers in the process.

4.5.2 Extensions of linear PCA

PCA is usually done by obtaining linear combinations of the raw dataset and then the residuals. Various techniques have been developed to fit a non-linear curve through the data. The analysis is then completed in the same manner as for the linear case. This technique of non-linear fits is known as non-linear PCA. The shape of the principal curve is determined by the particular dataset. The principal curve then serves as the one dimensional approximation of the multi-dimensional dataset. Non-linear PCA is applied to processes with extreme non-linear behaviour where a small number of linear principal components is not a good representation of the process dimensionality. Zhang et al. (1997) have illustrated the benefit of non-linear PCA techniques on a polymerisation reactor.

PCA is performed on the assumption that the data is statistically independent of time (not autocorrelated) and has a Gaussian distribution. This is definitely not always the case for real chemical processes due to numerous reasons that include non-linearities, instrument inaccuracy, etc. (see section 4.2.2). A technique called independent component analysis has been developed to cater for non-random datasets and is discussed by the following researchers Lee, Yoo, and Lee (2004) and Kano, Hasebe, Hashimoto, and Ohno (2004).

4.6 Plant evaluation index

Almost all of the methods discussed in this chapter have been evaluation techniques for single loop monitoring and evaluation. The problem is however that plant operation is dependent on large numbers of loops. It is a near impossible task to evaluate each and every loop individually. A holistic measure of plant operation is needed to provide a plant wide indication of performance. A plant wide evaluation index (PWI) was developed to attempt to quantify plant wide performance.

If we consider any processing plant there are four main value factors that influence the value addition of plant operation. These value factors are:

- Quantity of unrefined feed entering the plant
- Quantity of valuable products that leave the plant
- Quality of the products
- Utilities and other processing costs which allow for controlled and efficient operation

In the optimisation of the plant's operation we usually look to increase the throughput of the plant as much as possible while still maintaining product quality and keeping within the design limitations of the plant. The extra feed necessary to do this as well as the increased utility usage reduces the value addition that is obtained from extra product. From this line of reasoning we can formulate an optimisation problem as in equation 4.59.

$$J_{PWI} = Product + Quality - Feed - Utility \tag{4.59}$$

For a plant to perform at its best the objective function in equation 4.59 needs to be maximised which means that as much as possible product of good quality needs to be produced by utilising as little as possible feed and utilities. Two problems arise from this formulation of the optimisation problem.

The first is that the value factors need to be quantified in some way and if we use normal process variables like for instance flow rate, the terms may not be consistent. Suppose that the *Feed* term can be quantified by a flow rate while the *Quality* term can be something like a concentration or a composition. If this is the case then we foresee a problem because the terms do not contribute equally to the objective because of their units.

An important fact to note is that equal weight contribution to the objective is not necessarily the ultimate aim in the formulation of the objective function. This is because different processes have different control philosophies. It might be important for a certain processing facility to produce product at as high as possible a production rate, but the constraints on the quality are reasonably relaxed. For example, the product may vary between 50% and 80% purity. This means that if the product is at 50% or at 80% it doesn't matter, it is adding the same value to production. For this case the *Quality* term should add less weight to the objective when compared to the *Product* term.

The second problem is that the objective function does not compensate for constraints that exist on a real plant. Constraints like safe operation, environmental limits, smooth operation, stability, etc. need to be included in the optimisation problem. So the maximising of the objective in equation 4.59 is not sufficient for optimisation of the plant operation.

One way to cater for the inconsistency problem is to add weights to the terms to scale them in such a manner that they contribute the right amount of weight to the goal. Doing this will provide us with a new weighted objective function shown in equation 4.60.

$$J_{PWI} = w_2\ Product + w_3\ Quality - w_4\ Utility - w_1\ Feed \tag{4.60}$$

The question arises how do we determine the weights to transform the objective function terms into some kind of universal value. One way to do this is to choose the weights in terms of monetary ratios. The monetary weight is shown in equation 4.61.

$$w_i = \frac{cost}{value\ factor} \qquad \left[\frac{\$}{kg/hr}\right] \tag{4.61}$$

The weight shown in equation 4.61 will be for instance for a feed of which the cost to obtain is known and of which the flow rate is measured. The monetary weight is an effective way to scale the value factors that are represented by variables like flow rate, but for quality factors it is more difficult seeing that no real monetary value can be linked to variables like concentration or conversion. One solution is to combine factors like for instance the *Product* and *Quality* factors into one as in equation 4.62.

$$Value\ factor = C_i \times W \tag{4.62}$$

A new value factor is therefore created in equation 4.62 that will represent the production rate of the key component, i, in the product stream, W. C_i is for instance the concentration of component i in stream W.

From the above formulation of the objective function it is clear that some time scale or evaluation period needs to be defined. We cannot use single measured variables from the plant at some particular sampling instant seeing that it would in all likelihood not be representative of the normal operation of the plant. An average that is representative of normal operation over the evaluation period needs to be defined. Numerous averaging techniques exist of which some are briefly mentioned in terms of distributions in section 4.2.2. There are various possibilities to represent a norm of data and is a recommendation for future work that will continue on this study. The averaging method used in this investigation was to consider the value factors in terms of mass and energy accumulation over a predefined evaluation period.

A typical form for the proposed objective function can then be represented as in equation 4.63.

$$J_{PWI} = \sum_{i=1}^{n} w_{1_i} \int_{t_a}^{t_b} Prod_i \, dt + \sum_{i=1}^{n} w_{2_i} \bar{x}_i - \sum_{i=1}^{m} w_{3_i} \int_{t_a}^{t_b} Feed_i \, dt - \sum_{i=1}^{q} w_{4_i} \int_{t_a}^{t_b} Util_i \, dt \tag{4.63}$$

In equation 4.63 $Prod_i$, $Feed_i$ and $Util_i$ represent a particular product, feed and utility flow rate. n, m and q are the number of product, feed and utility flows entering or leaving the process respectively. The period of evaluation is defined from t_a to t_b. $\bar{x}_i$ is the average composition of a key component in product stream i with the average calculated over the evaluation period. w are weights that are assigned to scale the value factors.

The question now arises how do we use this developed cost function in the general performance monitoring structure? As was mentioned in section 2.8, the general optimisation problem and performance monitoring are closely related. If we consider the objective function in equation 4.63 we want the objective maximised for best plant performance. So if we want to quantify performance we have to set a benchmark of optimal performance and compare the actual operation for a particular evaluation period with this benchmark.

In order to set a benchmark we have to decide on an optimal operating state for the plant. Various methods can be used to do this, for instance, the design conditions of the plant, plant simulations, historical operating values, etc. The method used in this research was to solve the steady state mass balance for the process with optimal feed flow rates and feed compositions specified. Each actual operating value factor can then be compared with its optimum calculated from the plant model. The benefit of using benchmarks to evaluate performance is that the optimal state does not have to be an attainable state. We only need a constant reference point against which various periods of operation can be measured. So if a full scale plant model is not available, operator experience or any source of plant information can be used to set the benchmark. With this said it is always better to have an attainable optimal state seeing that it gives a tangible goal for process personnel to work to.

A single plant wide index (PWI) can be defined as a quantitative value for plant performance. The way that it was formulated was by comparing each value factor with its own benchmark value and making sure that the ratio of the two is between 0 and 1. The index can be defined as in equation 4.64 by utilising the objective function in equation 4.63.

$$\begin{aligned} PWI &= 100\left[w_1 \sum_{i=1}^{n} \frac{\int_{t_a}^{t_b} Prod_{i_{act}}\ dt}{\int_{t_a}^{t_b} Prod_{i_{opt}}\ dt} + w_2 \sum_{i=1}^{n} \frac{\bar{x}_{i_{act}}}{\bar{x}_{i_{opt}}} + \right. \\ & \left. w_3 \sum_{i=1}^{m} \frac{\int_{t_a}^{t_b} Feed_{i_{opt}}\ dt}{\int_{t_a}^{t_b} Feed_{i_{act}}\ dt} + w_4 \sum_{i=1}^{q} \frac{\int_{t_a}^{t_b} Util_{i_{opt}}\ dt}{\int_{t_a}^{t_b} Util_{i_{act}}\ dt}\right] \end{aligned} \quad (4.64)$$

$$1 = w_1 + w_2 + w_3 + w_4 \quad (4.65)$$

The variables and parameters used in equation 4.64 are as defined in equation 4.63. The subscript *opt* refers to the optimum operating state that is predefined and *act* refers to the actual operating state. The ratios of the values should all be less than one for normal regulatory operation and by multiplying the ratios by weights that sum to 1 will mean that the index will be between 0 and 100. It depends on the specified state but in almost all normal operating cases the optimum state accumulation or average will be larger than the actual if the value factor in the objective function (equation 4.63) is to be maximised. If the value factor (ex. *Utility*) needs to be minimised the optimum accumulation or average will be smaller than the actual. That is why factors that are to be minimised have the optimum state as the numerator and the actual state as the denominator, while factors that are to be maximised (ex. *Product*) have the actual state as the numerator and the optimum state as the denominator. We have to consider each term carefully when we are relating the value factor to its benchmark. For instance, the *Quality* value factor in equation 4.64 ($\frac{\bar{x}_{act}}{\bar{x}_{opt}}$) is defined in terms of the purity of the required valuable product, so if the composition is high the term will be close to 1. If we decided to define the value factor in terms of impurities the term will be the inverse ($\frac{\bar{x}_{opt}}{\bar{x}_{act}}$) where a low composition will be close to 1. The averaging method for x is however wrong if the optimum *Quality* is not an absolute maximum or minimum value for x. In these cases the error squared or integral of the absolute error (IAE) of x should be used.

If we look at equation 4.64 more closely we identify that the PWI in equation 4.64 is very similar to the original objective function defined in equation 4.63. The only difference is that the scaling for the objective function is done with the benchmark operating state.

An identified drawback of the PWI is that it does not cater for important differences in the same value factors. For instance, if we add ten feeds to a plant it adds equal weight to all of them by just normal summation. If we however have an extremely expensive catalyst feed for instance that enters as a feed stream, we would want to place more emphasis on its minimisation rather than something like a water feed stream to a stripper. Weighting of the individual components of a value factor is possible where the number of feed and product streams are few but can become troublesome for large scale processing plants.

The solution to this is to rather make the PWI a unit specific index. This is where the plant is divided into smaller processing units like, for instance, a single distillation column or maybe even a train or a reactor section, etc. This reduces the number of components of the value factors. If this segmented approach is followed unit performance indexes (UPI) can be standardised by evaluating the same unit specific value factors. A set of specific value factors can then be used for a processing unit that falls in the reactor category and another set for a separating unit like a distillation column. Obviously the value factors will then become unit specific like for a distillation column the standard *Quality* value factor may be the ratio of composition of a key variable in the distillate and bottoms streams, while for a reactor it may be the percentage conversion of a certain reactant, etc. The PWI can then still be calculated by weighting and summing the individual UPI values to provide one number of plant wide performance. The actual implementation and application of the PWI are shown and discussed in later chapters.

CHAPTER 5

The process

This chapter provides insight to the process that the performance assessment technique was applied to. The process setup is explained together with the operating software. Data transfer and storage is also discussed. If reference is made to programs or files refer to appendix E for more detail.

5.1 Process Description

A process flow diagram (PFD) of the process that was used in the investigation is attached in Appendix A. The process consists of a glass fractional distillation column with ten plates and an option to feed on plate 3 or 6 from the bottom. The column will, for purposes of this investigation only separate binary mixtures of ethanol and water.

The feed is sub-cooled and its temperature is dependent on the amount of cooling that is provided by a bottoms cooler , HX-02. The column uses a total condenser, HX-03. A total steam reboiler, HX-01, is used that provides a saturated vapour stream back into the column.

The process is a closed system where the product streams (distillate and bottoms) return back to the feed drum, DM-02, to be re-distilled. Material therefore cycles through the rig without any take-off of streams except for sampling purposes.

The column uses two steam kettle boilers as heat source that provide medium pressure steam to the column reboiler (about 1 MPa saturated).

5.2 Regulatory control philosophy

The control structure that is currently used was set-up to ensure good separation and product of a good quality. The reason for this is to ensure good separation to illustrate mass transfer principles for undergraduate students. Production rate is therefore not of utmost importance. The control structure can be summarised as follows:

- Firstly, the feed flow to the column will be controlled at a certain rate. This means that a constant flow of liquid enters the column. The feed temperature is controlled with the bottoms cooler by adjusting the cooling water. This will only work if bottoms flow rate is high enough, so the loop will typically only be set to auto when the column is drawing off product.

- Then, to maintain the mass balance in the column, the level in the reboiler and the reflux drum is kept constant by adjusting the bottoms and the top product flow rates. This illustrates the fact that constant product flow rate will not always be achieved.

- Quality control will then be done by controlling the bottom and top plate temperatures at specific set-points. The bottom plate temperature will be controlled by means of a cascade controller to the steam pressure loop. The reason for the cascade loop is cater for disturbances in the steam supply from the boilers. The reflux flow rate back into the column will be used to control the top plate temperature. An extra loop was implemented to keep the reflux temperature constant by adjusting the cooling water flow to the condenser. This ensures that product of a constant composition should be achieved. Composition which is inferred from temperature is not always accurate, but sufficient especially seeing that a binary mixture is used.

The regulatory control strategy therefore ensures a constant feed flow and temperature. The levels in the system are kept constant to maintain mass balance. Quality control is done by controlling the top and bottom plate temperatures.

5.3 Process instruments

Instrument communication forms an integral part of the performance evaluation structure. It forms the basis of data capturing and some recent technological advances have made field value measurement extremely functional. These advances include SMART instruments that have made instruments multi-functional in the sense that they can provide more information on the measurement than only the measured value alone. Some of this multi-functionality include remote calibration, sensor health, etc. In order to take full advantage of SMART instruments the right communication protocol needs to be implemented. HART and Foundation Fieldbus are two of these data transfer protocols. HART uses the standard $4 - 20\ mA$ analog signal with a digital signal superimposed on it. This allows for conventional analog measurement and A/D conversion but the digital signal superimposed on it provides the extra functional information of the measurement.

The advantage of the extra digital signal is that it provides two way communication with the instrument. Foundation Fieldbus is similar to the HART technology except that all communication is completely digital. When using Foundation Fieldbus, wiring costs are reduced because numerous instruments are connected via the same wire.

Table 5.1 provides information on the measurement technology that is available on the distillation column. It is important to know the capability of the instruments to help with fault detection, diagnosis and maintenance. Al instruments in the column area are intrinsically safe.

The measuring devices and their communication technology are shown in table 5.1.

Table 5.1: Measuring device communication the distillation column.

Process Variable	Tag	Device	Communication
Temperature	-	Thermocouples	analog
	TT-02 to TT-09	Transmitters	Foundation Fieldbus
	TT-01, TT-10 to TT-19	Transmitters	HART
Mass flow rate	-	flow meters	analog
	FT-01 to FT-08	Flow transmitters	HART
Pressure	PT-02	Pressure transmitter	HART
	PT-01	Pressure transmitter	Foundation Fieldbus
Level	LT-02	DP cell	HART
	LT-01	DP cell	Foundation Fieldbus

As can be seen from table 5.1, regular analog, HART enabled and Foundation Fieldbus instruments are available on the column.

5.4 Digital communication

The digital communication with the process consists of the $DeltaV^{TM}$ operating system and hardware. The $DeltaV^{TM}$ system makes the process fully automated by providing the link between field instruments and a computer workstation. All signal routing happens by means of a central junction box which feeds data to and from the field as well as to and from workstations. Figure 5.1 provides an overview of the $DeltaV^{TM}$ system (Fischer-Rosemount, 2003). How the configuration in figure 5.1 works is that firstly the link to the field happens through the I/O subsystem. This is where D/A or A/D conversion takes place. The I/O subsystem then also communicates with a digital controller module. A power supply is mounted on the same board to provide power to the controller modules and I/O cards. The controller module is a digital controller that contains all the controller information needed to operate the plant (algorithms, parameters, set-points, alarms, etc.). It is therefore not necessary to have a computer workstation running to control the plant seeing that all the control action is done via the controller module. The question is,

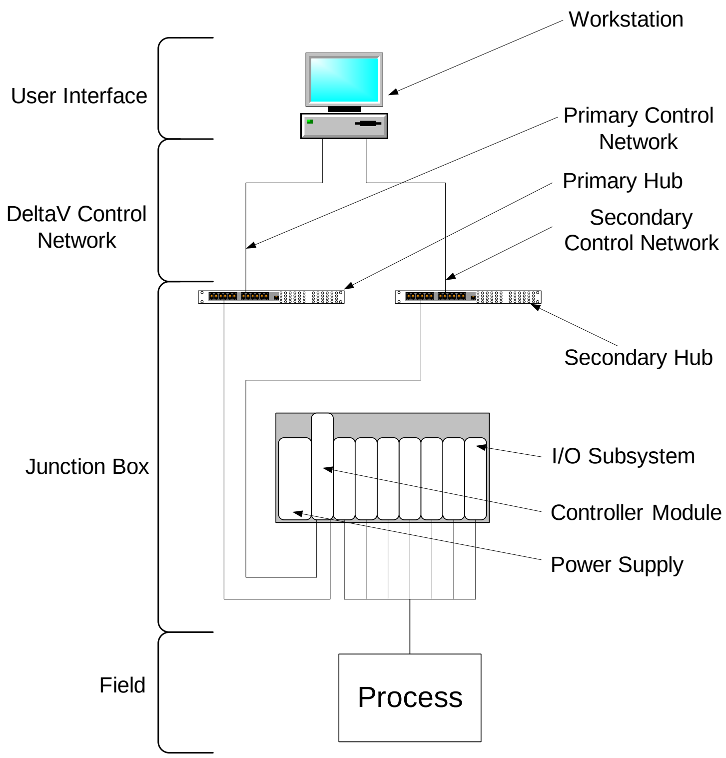


Figure 5.1: Overview of the $DeltaV^{TM}$ system.

how does the controller module obtain the control information? This is done via the $DeltaV^{TM}$ control (secondary/primary) network by which information is downloaded to the controller modules from multiple workstations. There are various workstation types on the control network and this is discussed in section 5.4.1. Each type of workstation has a different functionality together with a set of users with various access rights. Only specific users on certain workstations have permission to download new information to the controller.

5.4.1 Operating software

The workstations on the $DeltaV^{TM}$ control network have different functionalities depending on the software and licensing installed on the machine. The $DeltaV^{TM}$ network at the University of Pretoria has three $DeltaV^{TM}$ workstations. The three stations as well as their functionalities are:

- Professional plus station (PROplus - Configuration, operation and Database configuration.

- Application station - Run-time database plus user selected applications. For the purpose of this investigation the 3rd party OPC application is important and is discussed further in section 5.4.2. The application station is the only station that is not only on the $DeltaV^{TM}$ control network but also on the local area network

(LAN) of the university.

- Operator station - For plant operation alone.

5.4.2 Data capturing

The application station has a data historian that runs continuously to capture data according to time stamps. This allows users to access plant data at any time in the past up to the present. The tags that are captured need to be activated and downloaded on the PROplus station. The application station has third party OPC functionality which enables an OPC server to run continuously. This is extremely functional seeing that OPC clients that are not on the *DeltaV*^{TM} control network can have access to the plant operating data because the application station is on the university LAN. The only application that is needed on the client machines is OPC remote which is distributed with the *DeltaV*^{TM} installation software.

Data for this investigation was captured from the OPC server, real-time and logged on the OPC client machine by using the MATLAB environment. MATLAB is a technical computing environment with OPC capability enabled by the OPC Toolbox. This is not the optimal method for data capturing seeing that the data logging has to be triggered by the client before data is available (not continuous). The OPC Toolbox is not designed to replace continuous historians and a better method will be to obtain data directly from the continuous historian via OPC (Mathworks, 2005). This is a recommendation for future work on performance assessment of the distillation column.

Although the desired data capturing method was not applied, the OPC Toolbox logging was sufficient to illustrate the performance interface functionality discussed in chapter 6. It is important to identify the key variables to be logged in order to obtain all applicable information with respect to process operation. The following variables were identified as functional to performance assessment and their application in the assessment interface is discussed in chapter 6.

- Controller set-points
- Actual CV values
- Controller outputs (valve position)
- Controller modes (auto/manual)
- Product flow rates
- Utility flow rates and temperatures

Table B.1 in the appendices shows all the tag names as well as the process variables they represent. A template was created in the OPC Toolbox interface for convenience for other MATLAB users that want to connect to the column via OPC. The *osf* file is found on the CD accompanying this document.

The tag names were assigned in the *DeltaV*^{TM} environment upon commissioning of the column. Problems were experienced with name space browsing within the OPC toolbox graphical interface. The MATLAB function *getnamespace* was used to retrieve the server name space in the MATLAB command line. The name space was then available as a structure in the MATLAB workspace. The structure is incredibly big with numerous fields and substructures which contain the entire *DeltaV*^{TM} control structure which made browsing for tags very difficult seeing that one gets lost in the tree. To make the browsing of the structure more functional it was converted to a xml file which has a tree format similar to file system explorer software. This made browsing for tags easier seeing that the user can track its position in the tree. A MATLAB function called *struct2xml* was used to convert the name space tree structure to an xml file (Sandrock, 2005). The *struct2xml* function as well as the xml files are available on the CD that accompanies this document. The xml file is very functional seeing that can be used to identify tags to all possible variables that can be obtained from the OPC server. Numerous applications can be used to view xml files. Microsoft Excel was used in this project.

It is possible to connect to the continuous historian by means of Microsoft Excel. This is done by using the PI data server add-in that is included in the *DeltaV*^{TM} product. This allows data retrieval according to timestamps in the past. This is functional seeing that the logging does not have to be triggered to record data, but seeing that all the programming to analyse data was done in the more powerful MATLAB environment this method was not considered. MATLAB does however allow for data importing from Excel so it is a possible route that could be considered if direct MATLAB-continuous historian communication is not possible.

A problem was experienced with the data timestamps logged on the client machine. Every sampling instant the client machine sends a data packet containing the tags which need to be sampled, to the OPC server. The server then responds with another packet containing the values of the tags. Every packet however contains its own timestamp that is the same as the local time on that workstation. The result is that the same value for a tag gets logged at two different times. If the client machine and the machine that runs the server are at different times, two sets of exactly the same data is captured. To cater for this, the time on both workstations (client and server) was synchronised with a time server on campus. This ensured that sending and receiving of data happened synchronously on both machines.

A detail that should be remembered when logging with the OPC toolbox is that data only gets logged when the value of that tag changes from one sampling instant to the next. So if during the period of logging the set-point of a controller for example does not change the actual set-point value is not stored. This is not a problem for measured process variables from the field like temperature or flow rate seeing that they change regularly but set-points and controller modes are things that should be noted. This was catered for by changing all the controller modes and then immediately turning them back to their original state. This allows for a quick change in the set-point and obviously would also be picked up in the mode value. It should be ensured that this changing of modes is longer than the sampling period of the data logging, otherwise the change won't be recorded by the OPC client. If values have a quality string of value *repeat* it means that the value hasn't changed and the value takes on the last changed value. If there is no previous value for the specific tag, the value takes on the *NAN* string which indicates not a number.

CHAPTER 6

Structure implementation

This chapter discusses the implementation of a performance monitoring structure on the process discussed in chapter 5. The implementation was done with the aid of two performance evaluation user interfaces. The structure comprises of the following sequential approach:

- Data acquisition
- Plant wide assessment
- Single loop assessment
- Adaptation of the control philosophy
- Archiving

This chapter discusses the implementation of the performance monitoring structure and in chapter 7 the implemented structure is applied to real periods of operation on the column.

The implementation programming is available on the CD that accompanies this document. The files that are available on the CD are discussed in appendix E.

The performance structure was developed with an initial aim to make it completely general, transparent and real-time. This meant that it could be used on any process and should be understandable to all parties who take part in plant operation decisions. This includes management, operations and the design office. This was partially achieved as will be seen in the sections that follow. A good foundation has been set and work that is to follow on this research will have a clear view of the procedures and structures that are necessary to have an efficient performance monitoring, assessment and diagnosis system.

The structure implementation can be summarised by the flow diagram shown in figure 6.1. Two performance interfaces form the basis of the implemented performance

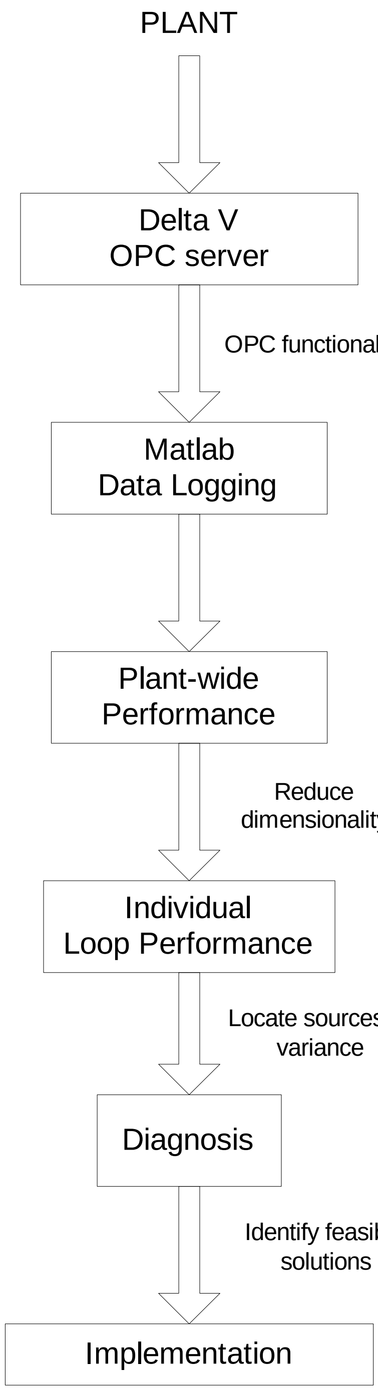


Figure 6.1: The flow diagram of the implemented performance monitoring structure.

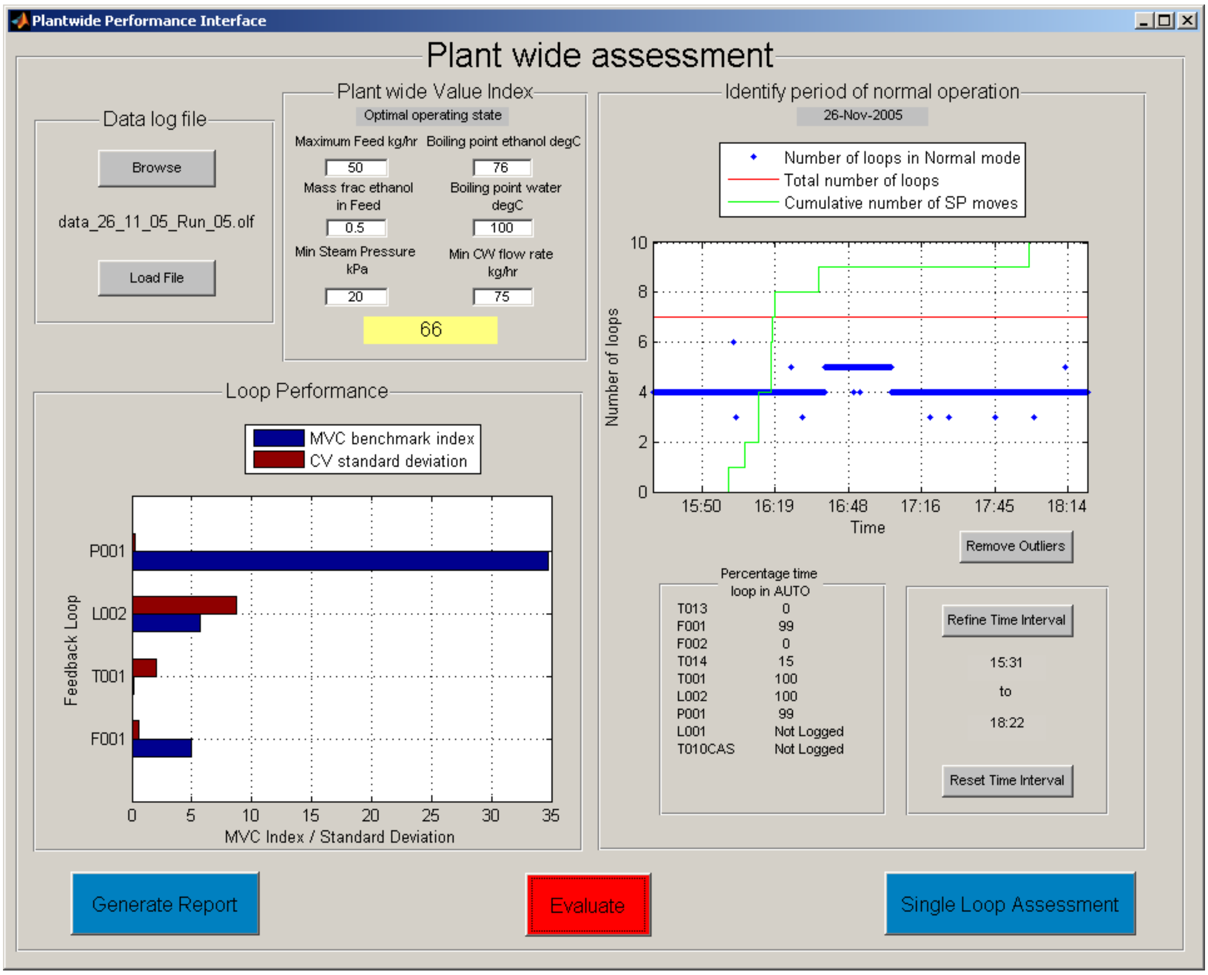


Figure 6.2: The graphical user interface used for plant wide performance evaluation

structure. The first interface gives a holistic view of process operation and control system performance and is discussed in section 6.1. After the plant wide interface is considered possible problematic control loops should be evident and the individual loop assessment interface (discussed in section 6.2) is then applied. Possible improvements in the control structure should be clear after single loop assessment which can then be implemented.

6.1 Plant wide performance interface

The interface that was developed to give an overall plant wide idea of performance is shown in figure 6.2. The interface has a sequential approach to its operation. First the data log file should be located and loaded to make the process variable structure available in the workspace. After the variables are loaded the *Evaluate* pushbutton is activated which initiates calculations. The calculation results appear as two plots on two separate axes and as a plant wide index (figure 6.2). The first results of the calculation is for an evaluation period that has the same time span as the data that was logged. The whole period is not necessarily a period of normal operation and periods of normal operation need to be specified as evaluation periods. We define normal operation as periods where as little as possible set-point changes occur with as many as possible control loops on AUTO. These periods can be identified by considering the first plot which shows the number of loops on AUTO as well as the cumulative number of set-point changes for the considered evaluation period. The normal operating periods can then be isolated and individually evaluated by using the *Refine Time Interval* pushbutton. Once the evaluation period has been set we can look at specific sources of variability due to poor performing loops. At the end of the evaluation a HTML plant wide performance report is generated and archived. The single loop interface is then used to investigate the possible poor performing loops identified by the plant wide interface. The individual sections in the sequential approach are discussed in more detail in the sections that follow.

6.1.1 Data acquisition

The data logging procedure by means of the OPC toolbox has been discussed in section 5.4.2. This is definitely not the optimal way to retrieve data for analysis seeing that it is not continuous. A method has to be devised to import data into MATLAB directly from the DeltaV continuous historian.

If data has been logged with the OPC toolbox it needs to be imported and reworked into a format that can easily be manipulated by statistical methods into useful information. The way this was achieved was by importing the logged data as a structure with the built-in function *opcread*. The imported structure was created with the following four fields:

- ID - A vector of the tag names on the OPC server that was logged.
- $Value$ - A matrix of the actual values of the identified tags.
- $Qual$ - A matrix of the signal quality of the identified tags.
- $Tstamp$ - A matrix of timestamps that identifies the sampling instants.

The ID field is a cell array which consists of a single column that contains strings of the tag names provided by the OPC server identified in table B.1 in the appendices. The $Value$ and $Tstamp$ fields are double arrays that have the same number of columns as the number of tags while the rows indicate the sampling instants. The $Qual$ field is a cell array with the same dimensions as the $Value$ and $Tstamp$ fields. A specific variable's data can then easily be retrieved by recalling the sampling instants corresponding to the column number that corresponds to the position of the tag in the ID field. The same can be done to retrieve the variable values and signal quality.

The time stamp value is available as a serial date number which is the number of days from the reference starting point of *1-Jan-0000*. The serial date number is a standard general way to represent time and used to prevent formatting errors when specifying dates and times. The signal quality strings are part of the inherent diagnostic capabilities of the plant instruments and software (SMART functionality). It provides the user with extra information with respect to the measurements taken. The allows for extra measurement information like the read value lies outside its calibration range by indicating in string format that the measurement was "high limited" or "low limited" depending if the measurement is above or below its calibration limits. The OPC foundation has a specific method to report this data so that it can be uniformly understood by OPC enabled software. Refer to Mathworks (2005) for quality string definitions.

In the plantwide interface the following procedure is followed to create the structure in the workspace. Firstly, the local hard drive is browsed to locate the $*.olf$ file that was created by logging the data in the OPC toolbox interface. Once the file is located the $load\ file$ pushbutton is clicked and the callback function then imports the structure into the workspace. Now with the data in the workspace further analysis is possible. Figure 6.3 shows the corresponding load file panel for data acquisition from the data log file.

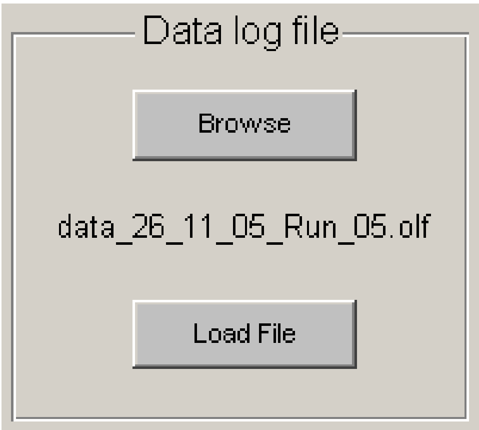


Figure 6.3: The data log file pane.

6.1.2 Periods of pure regulatory control

In order to evaluate regulatory performance a proper evaluation period needs to be identified. For regulatory control as little as possible set-point changes have to occur. Also, for the column to be performing well in terms of disturbance rejection, as many as possible of the loops defined in the base layer control philosophy need to be in operation. The way periods of regulatory control are identified is by considering figure 6.4 in the plant wide interface. From figure 6.4 we get an overview of how many SP changes have been made as well as the number of loops that were on auto. The method that was used to detect set-point changes was to consider the change for consecutive samples. If the change in

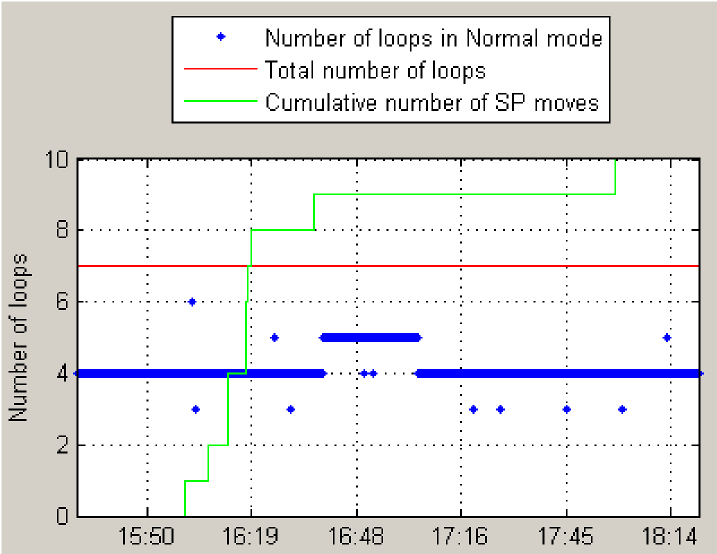


Figure 6.4: The number of loops on auto as well as the cumulative set-point changes.

set-point is larger than 1%, a SP change is said to have occurred. The cumulative sum of the set-point changes are plotted, so when the set-point plot trend flattens out and does not change according to time one knows that the column is operating under regulatory control. Only the set-point changes for the loops that were actively functioning for more than 50% of the time is plotted. Seeing that when the loop is set on MANUAL the SP and the actual value is the same and will therefore mean lots of set-point changes. The plots may look unrealistic because of large set-point change numbers being plotted. This is not necessarily the case seeing that in periods of shut-down and start-up large variable movement occurs and if the loop is on MANUAL the set-point moves will be large. The *refine time* pushbutton should just be implemented to locate proper evaluation times. The actual numeric percentage time on AUTO for the loops are reported in the *Time on AUTO* panel shown in figure 6.5.

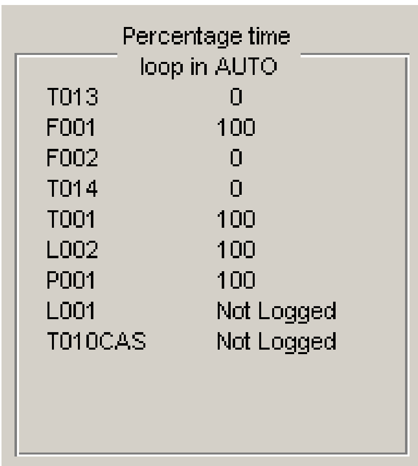


Figure 6.5: The percentage time on AUTO for all control loops.

The *Refine Time Interval* pushbutton can now be used to change the evaluation period to the desired period of normal operation. By activating the *Refine Time Interval* the mouse pointer becomes an active coordinate selector. The mouse pointer is then moved to the starting time for the evaluation period and then clicked. After setting the starting time the pointer still remains a coordinate selector and the end time can then be specified. After the end time has been selected the time strings change to the selected ones. The evaluation period has now been set and the *Evaluate* pushbutton can be selected to perform calculations for this period of operation.

6.1.3 Plant wide Value Index

The plant wide index discussed in section 4.6 gives an overall impression of how well the plant is performing. High values (close to 100) mean that the plant is performing close to the specified optimum operating state. The panel that shows the plant wide performance is shown in figure 6.6. As can be seen from figure 6.6 the optimal operating

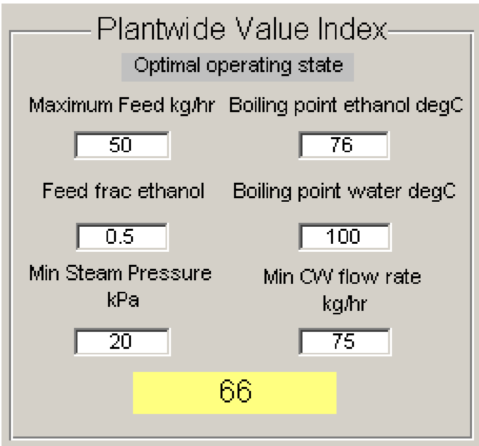


Figure 6.6: The plant wide performance panel.

state can be specified by the user. The specified parameters do not always result in an attainable steady operating state. This is not necessary seeing that the optimum operating state can be used as a fictitious benchmark or reference and current operation is only evaluated comparatively. To have a realistic optimum operating state, the steady-state mass balance needs to be solved for a specific feed and separation which is a fairly simple task. This makes the plant wide index a very flexible as well as realistic measure of overall performance. The index can be compared to historical values of the index to make judgements on performance. It should be kept in mind that the historical comparisons will only be useful if the benchmark optimum state is the same. Equation 6.1 was used to calculate the index for the considered process.

$$PWI = \frac{1}{6}\left[\frac{\int F_{act}dt}{\int F_{opt}dt} + \frac{\int Q_{act}dt}{\int Q_{opt}dt} + \frac{\int U_{opt}dt}{\int U_{act}dt} + \frac{\bar{P}_{opt}}{\bar{P}_{act}} + \frac{\bar{T}_{bot_{act}}}{\bar{T}_{bot_{opt}}} + \frac{\bar{T}_{top_{opt}}}{\bar{T}_{top_{act}}}\right]100 \qquad (6.1)$$

The time intervals for integrals in equation 6.1 are for the selected evaluation period. In equation 6.1 F, Q and U represent the flow rates of the feed, products and cooling water in kg/hr respectively. $\bar{P}$ and $\bar{T}$ refer to the average of the steam pressure and plate temperature over the evaluation period. As can be seen from equation 6.1 equal weights have been assigned to the terms. The subscripts *opt* and *act* refer to the optimal and actual operating points respectively. The temperatures in equation 6.1 refer to the bottom and top plate temperatures.

Seeing that the flow rate of the steam is not measured explicitly on the column the steam pressure, P, was used as an indication of steam usage. There is also no on-line composition measurement on the column. The bottom and top plate temperatures were used as an indication of the quality of the separation. This method of inferred composition is not very good especially when the column is in a transient state as well as when more than two components are distilled.

6.1.4 Sources of variability

Now that we have identified a normal period of operation we can reduce the dimensionality of the performance problem by identifying possible sources of process variability. Firstly we have to identify possible problems with the data acquisition process. The *Time on AUTO* pane can be used for this. Locate the loops that show data "not logged". These loops cannot be considered within the performance interfaces and need to be evaluated through inspection of data on the continuous historian. The reason for the lack of logging needs to be identified and rectified. Secondly identify loops that were not on AUTO for the period of evaluation. Identify reasons for this by looking at information supplied in the single loop interface as well as in the Delta V operating system configuration. After the "not logged" loops and the MANUAL loops have been identified single loop analysis of functioning control loops can be considered. This can be done with the aid of the plots on the *Loop Performance* pane shown in figure 6.7. Figure 6.7 shows the MVC benchmark index as well as the standard deviation of the CV for the loops that are on AUTO for more than 50% of the evaluation time. Good tight control (good disturbance rejection) will usually mean a large MVC index value and a small standard deviation of the CV. By considering this fact possible poor loop performance can be identified. It should be noted however that a poor performing loop according to the MVC index will not necessarily mean bad plant performance. Historical values of the index should definitely be considered to get a feel for what a good value for the index should be. On the other hand if the MVC index is high (close to a 100) the loop is by default performing well. The bar graph is a starting point for variance location and further analysis. In figure 6.7 it can be seen that all the loops seem to be doing well except for the level control on the

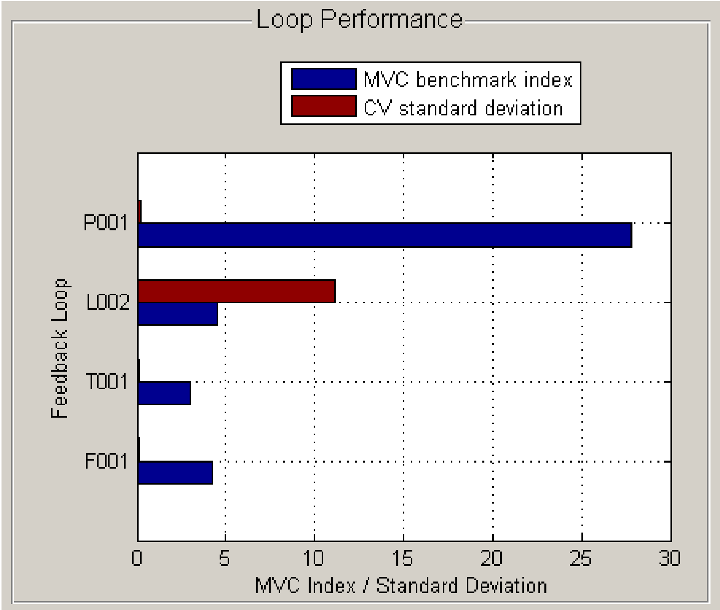


Figure 6.7: The Loop Performance axis in the plant wide interface.

reflux drum, loop $L002$. The single loop assessment interface should then be utilised first for the suspected level loop to locate the source of variance and then for the other loops that seem to be doing well (section 6.2).

6.1.5 Plant wide performance report

Reports in both the interfaces are automatically generated with the MATLAB built-in report generator tool. The report generation of the results in the plant wide interface is activated by the *Generate Report* pushbutton. The callback function that is called by the pushbutton creates a summary of the results in HTML format. An example of the plant wide report is shown in appendix C.

The reporting function of the interface was added to help set up a data base of past periods of operation. This is useful because most of the techniques and measures should be compared with past periods of operation to aid evaluation.

The run-time for the automatic report generation is long (several minutes). This is due to the variable availability in the workspace of MATLAB. The current workspace containing all the variables needed for evaluation first has to be saved. Then, when the report script is executed, the file gets loaded and makes the variables available to the report generator to insert it into the report. A recommendation for future work is to investigate the possibility for the report generator to import variables directly from the interface handles structure. This will cut out the saving and loading of the variable structure which takes long due to the size of the datasets.

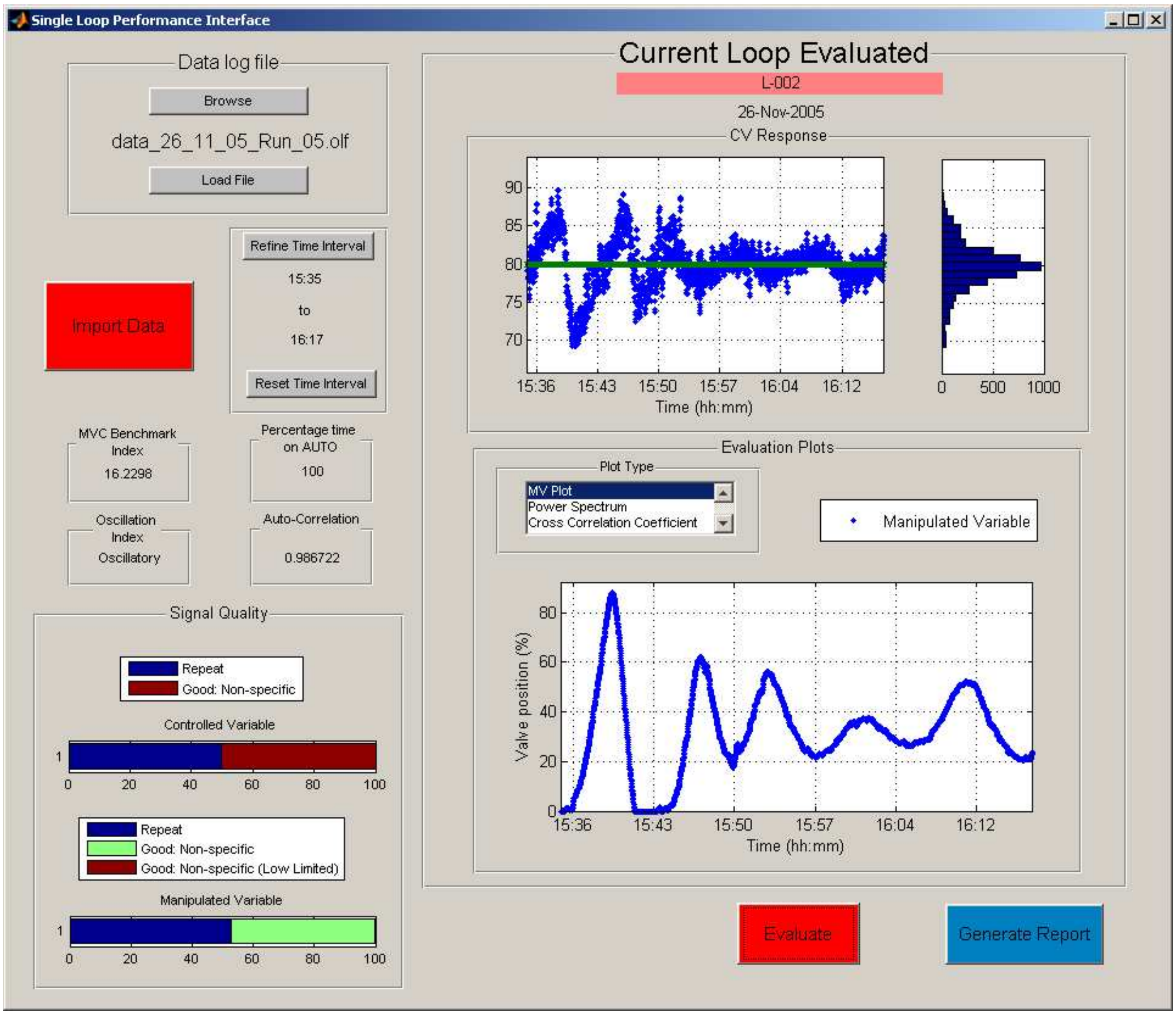


Figure 6.8: The graphical user interface used for single loop performance evaluation

6.2 Single loop performance interface

The single loop performance interface should be seen as a tool for locating and confirming suspicions of variance in the control structure identified by the plant wide interface. The loop performance interface is shown in figure 6.8. An important fact to note is that the two interfaces operate independently from each other, which means that the plantwide interface does not have to be activated or running for the single loop interface to work. This enables single loop assessment of the data without even considering the plantwide interface which is handy if the poor performing loops are known from a source other than the plantwide interface. The single loop interface can however be launched from the plant wide interface by activating the *Single Loop Assessment* pushbutton. To use the single loop interface, much the same sequential approach is implemented than for the plant wide interface. The $*.olf$ log file first has to be loaded. The data are then dimensionally reduced to only include the information of the loop under consideration. This is done by clicking the *Import Data* pushbutton which displays a window of all the loops that can be considered. When the loop information is available in the workspace the *Evaluate* pushbutton can be clicked to perform the statistical analysis of the data and to provide the results in the interface. Now once again periods of normal regulatory control operation can be identified by locating periods where the loop was on AUTO with as few as possible set-point changes. The evaluation period can be adjusted by the *Refine Time Interval* pushbutton. After the time interval has been refined the *Evaluate* pushbutton needs to be activated again to redo the calculation. A lot of information is then available and discussed in the subsequent sections. After the single loop analysis has been completed, a report can be generated and archived for future reference. The various functions of the single loop interface are discussed in more detail in the sections that follow.

6.2.1 Data acquisition

The data acquisition for the single loop interface is exactly the same as for the plant wide interface case. The only difference is that only the data needed for the considered loop is kept in the current directory to speed up calculation time. This is done by having an extra callback function which is executed with the *Import Data* pushbutton. The pushbutton opens a window with all the available loops on the column. When a loop is selected the relevant data is extracted from the large structure containing all logged data that was discussed in section 6.1.1. Calling the *Import Data* callback function creates a smaller structure with only the relevant information. If one loop is evaluated and another needs to be considered, the *Import Data* pushbutton is simply clicked again and the data structure gets replaced by the new loop's relevant data.

6.2.2 Time series plots and distributions

The CV Response pane shows the CV response versus time as well as a histogram of the data. This is simple information that provides significant information on the performance of the loop. The response provides an initial qualitative impression of the performance. The traditional methods of performance monitoring discussed in section 4.1.1 can be applied to this time series plot. The actual value of the CV as well as the set-point is shown. The histogram is a handy tool to detect some key characteristics of the control action. For instance distributions that have tall, narrow peaks are associated with good control while short, wide peaks show bad performance. Also skew distributions indicate characteristics like stiction of valves, non-linearities, constraints, etc. Examples of distributions that are performing poorly are shown in chapter 7.

6.2.3 Signal quality

The signal quality strings that are logged in the OPC foundation format are summarised in the *signal quality* pane in the interface. It displays the quality for the MV signal (controller output) as well as the CV signal (controller input) for the considered loop. The values are displayed as stacked bar plots with the length of the bars the percentage of the samples that took on that specific quality status. It is used to identify whether the data that was captured is accurate and trustworthy. Various labels are assigned to samples and the user should refer to Mathworks (2005) for the explanations of the string meanings. The OPC foundation has defined three statuses for defining quality. The first is the major status display whether the measurement was *bad*, *good*, *uncertain*, etc. The second provides an explanation for the major status which is something like *measurement failure*, *comm failure*, etc. The last division provides an indication of how the previous status info will limit the measurement value, like for instance *high limited*. A typical signal quality string will be, *Good: Non-specific (Low Limited)*. This will mean that the instrument is limited to its calibration range or is saturated, etc.

6.2.4 Refining the evaluation period

As was the case in the plant wide interface the *Refine time interval* pushbutton is used to identify periods where only load disturbances were acting on the process and the controller was switched on AUTO. The percentage time the loop was on AUTO can be seen from the time series plot and is also displayed as the actual number in the interface. Set-point changes can also be located on the time series plot by considering the set-point trend. It should be remembered however that the evaluation period is now for the considered loop and other loops may be changing set-point and interacting with the considered loop. It is therefore handy to consult the plantwide interface or HTML report to determine if other loops were changing set-point or mode that may affect the performance of the loop currently under consideration. The control loop interaction will show up on the cross correlation plots in the single loop interface and it is important to know if the interaction was operator induced or because of the control configuration. As was the case with the plantwide interface the evaluation period can at any time be reset to the original time span of the original data structure.

6.2.5 MVC benchmark index

The MVC benchmark index is a single quantity that indicates how well a particular loop is performing. The value varies between 0 and 100 and it indicates how good the control loop is to minimise controllable variance in the CV. The index compares current operation to the optimal case of minimum variance control. It is however in most cases not possible to reach this optimal case and the index is therefore small. This does not mean the controller is performing badly; it may just be the best that it can do. That is why the index should be compared to historical values for the loop. The theory behind the index is discussed in section 4.4.1.

6.2.6 Evaluation type

A drop down list labelled, *Plot Type*, was added to the interface to reduce the amount of information that gets displayed on the interface. This allows the user to choose what plots need to be performed. It also reduces the computation time for evaluations seeing that all the calculations are not done for the same evaluation executions. It currently has three options which include:

- MV plot - Plot of the manipulated variable of the considered loop. This is useful to identify excessive MV movements, MV saturation, etc.
- Power spectrum - Plot of the frequency distribution of the CV. This is implemented in oscillation detection.
- Cross correlation coefficient - Plot of the correlation coefficient between the CV and the MV of the same loop as well as the MV of some other specified loop.

The implementation of these plots are discussed in more detail in the following sections

MV Saturation

MV saturation is a common problem in normal feedback control structures. Two indicators of MV saturation are available in the interface. The first is to selected the MV plot in the *Plot Type* drop down list and to then visually inspect the MV response vs. time. The second indication is to consider the bar graph of the signal quality to see if the MV signal was high or low limited during the evaluation period.

Oscillation detection

Two quantities are used to locate oscillations in the considered control loop. One is the method proposed by Hägglund (1995) discussed in section 4.3.2. It was implemented as a yes/no type indication of oscillatory behaviour. If, according to the algorithm, the response is oscillatory the string *Oscillatory* is displayed and when the response is not, the string *Non-oscillatory* gets displayed. A few user defined parameters need to be specified to execute the algorithm. The default parameters for all the loops are shown in table 6.1. These values are chosen and cannot be edited in the interface itself. It should be edited in the MATLAB m-file that contains the interface code. ω_u is the ultimate frequency of the CV and is obviously different for each control loop. This makes the method not as

Table 6.1: The user defined parameters for the Hägglund (1995) algorithm

Parameter	Value	units
a	0.01	-
ω_u	10	seconds
n_{lim}	10	-

accurate that it could be seeing that the ultimate frequency for a flow loop will be much larger than for the temperature loops. Once again the result of the algorithm should be considered for past periods of operation to get a feel for typical results for responses. To make the Hägglund (1995) method loop specific is a recommendation for future work.

To compensate for inaccuracies of the Hägglund (1995) method the power spectral plot of the CV is a handy tool (see section 4.3.1). The power spectrum shows the frequency components of the signal. If large peaks occur in the low frequency range it indicates that there is slow dynamic oscillating behaviour which is bad seeing that the controller should be able to compensate for this. High frequency peaks are usually too quick for the controller to react to and are deemed uncontrollable variance. To perform the power spectrum calculation the FFT must be performed. Before the FFT was performed the data was reworked to the residual values of the CV that was obtained by subtracting the mean of the CV. Another important factor that should be considered when performing the FFT is the resolution. The resolution should be large enough to cover the entire specified frequency range.

For the PSD plots done in the interface it was decided to perform the FFT for a resolution of 2^{16}. This was found to provide sufficient resolution to gain the relevant dynamic behaviour for most of the CV's. Remember that the algorithm for the FFT calculation is the quickest for resolutions that are in factors of 2. The useful frequency range are for all frequencies smaller than the sampling frequency. The sampling frequency for most of the data structures are 0.5 *seconds*. The plot was therefore performed for frequencies up to 2 Hz.

Cross correlation plot

The last option in the *Plot Type* list is the cross correlation coefficient. This plot will provide an indication of two loops control interaction or interference. Two cross correlation coefficients are calculated. The one is for the considered loops CV and its MV while the other is for the considered CV and the MV of an other loop specified by the user. This is a good way to locate controller interaction and to identify possibilities for advanced control applications like decoupling or MPC. The correlation method is not complete however, disturbance-CV and CV-CV correlations still need to be implemented. This will provide information on what causes variance in the control loop. The auto-correlation coefficient for a lag of one is also displayed as an explicit value in the interface. This value provides an indication of the randomness of the CV signal. Values close to 0 indicate randomness. If the auto-correlation coefficient is zero then the controller is performing excellently seeing that only random noise is contained in the CV. This means that all predictable trends have been removed by the controller.

6.2.7 Report generating

Similar to the plant wide case a report can be generated for the single loop evaluation case by activating the *Generate Report* pushbutton. The callback function that is called by the pushbutton creates a summary of the single loop results in HTML format. An example of the single loop assessment report is shown in appendix D.

The single loop report generator works on exactly the same method as for the plant wide case and has the same problems with long run-times. Improvements can be made with the single loop evaluation reporting seeing that it has not been optimally configured. Information like signal quality is for instance not displayed in the report. Also the person that does the interpretation of the interfaces should be able to add comments to the report which is not currently catered for. The same can be said for the plant wide reporting.

CHAPTER 7

Structure application

This chapter shows typical examples of how the implemented structure can be applied to monitor process performance. All the data used for the illustrative examples were obtained from the laboratory distillation column discussed in chapter 5.

7.1 Evaluating a period of operation

In this section the implemented performance structure will be applied to a period of operation of the considered process. Data was captured for a period from 15:22 to 18:29 on 26 November 2005.

7.1.1 Plant wide evaluation

The unrefined evaluation of the data by the plant wide interface is shown in figure 7.1. The first step in the evaluation methodology will be to locate logging issues as well as to determine why some of the loops were not switched to AUTO at all. This is done in the sections that follow.

Poor data acquisition

From the initial unrefined plant wide interface evaluation shown in figure 7.1 it is apparent that not all the data from loop, $L001$ and $T010CAS$, were logged. Reasons for this need to be identified and rectified.

The $T010CAS$ loop is a cascade loop that controls the bottom plate temperature by changing the set-point of the steam pressure loop, $P001$. The reason for this loop not being logged is that it has not been commissioned yet. This means the controller

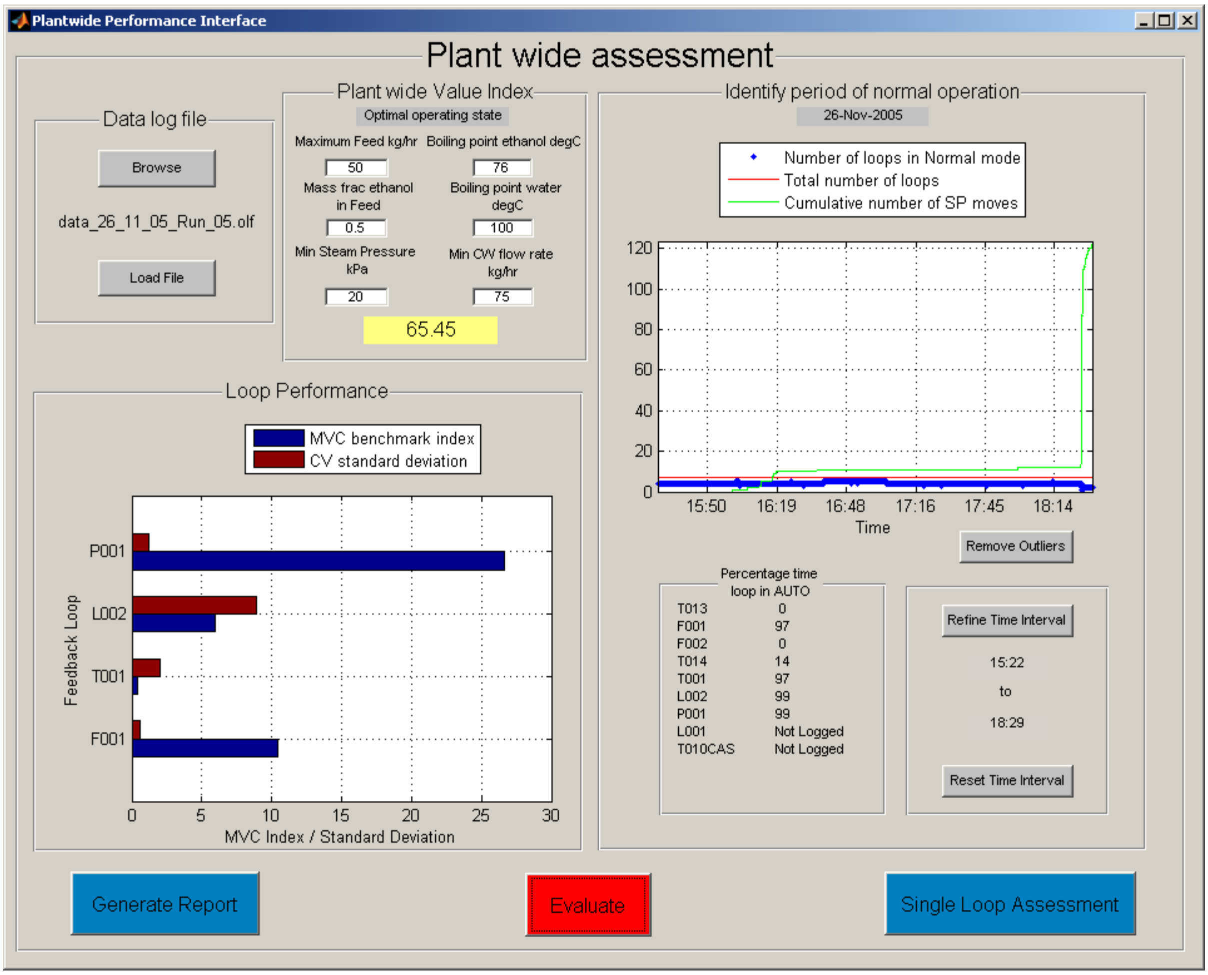


Figure 7.1: The unrefined data evaluation by the plant wide interface.

function block has not been properly set-up on the DeltaV system. Tags have however been created.

The boiler level loop, $L001$, was in operation for the evaluation period but was not logged into MATLAB. The problem was that the set-point did not change over the entire period of evaluation. This was due to the logging nature of the OPC toolbox as mentioned in section 5.4.2. So the SP was not logged seeing that the value didn't change and no previous value for the SP was logged in the specific data set. It should be noted that all the set-points that need to be logged have to be changed and quickly set back to the original value to enable logging.

Loops not in normal mode

As can be seen from the plant wide interface loops, $T013$ and $F002$, were on MANUAL for the entire evaluation period. The reasons for this need to be identified and rectified if necessary.

$F002$ is the second feed to the column. It is not necessary for purpose of this investigation to use this feed seeing that it is more relevant to distillation configuration studies. So the loop was switched to MANUAL and the valve closed. Flow loop $F002$ is therefore not a problem and we continue with the evaluation.

The feed temperature loop, $T013$, is the temperature loop that can be used to control feed temperature into the column. This loop was also set to MANUAL and the CW valve closed seeing that this loop is not critical in the successful operation of the column. This loop is there to fulfil a feedforward type action to dampen fluctuations in feed temperature to ensure smoother operation. It was decided not to use this loop for now seeing that other loops on the column are more critical to performance and first needed to be optimised before advanced control applications like this feedforward loop need to be considered.

Initial refinement of the evaluation period

To refine the evaluation period for the particular dataset the period of normal operation pane shown in figure 7.1 can be considered. As one can see, the current evaluation period (15:22 to 18:29) includes part of a shutdown period of the column seeing that the number of set-point changes becomes large to the end of the period. This means that some of the controllers that were on AUTO were set to MANUAL during shutdown and the set-point followed the trend of the CV. If we refine the time interval to exclude the shutdown period the plantwide interface provides an evaluation shown in figure 7.2. The evaluation period is now from 15:31 to 18:24. The period now excludes a nine minute interval at the start of the original time interval as well as the a five minute interval at the end. The nine minute interval at the beginning won't make a big difference to the overall performance

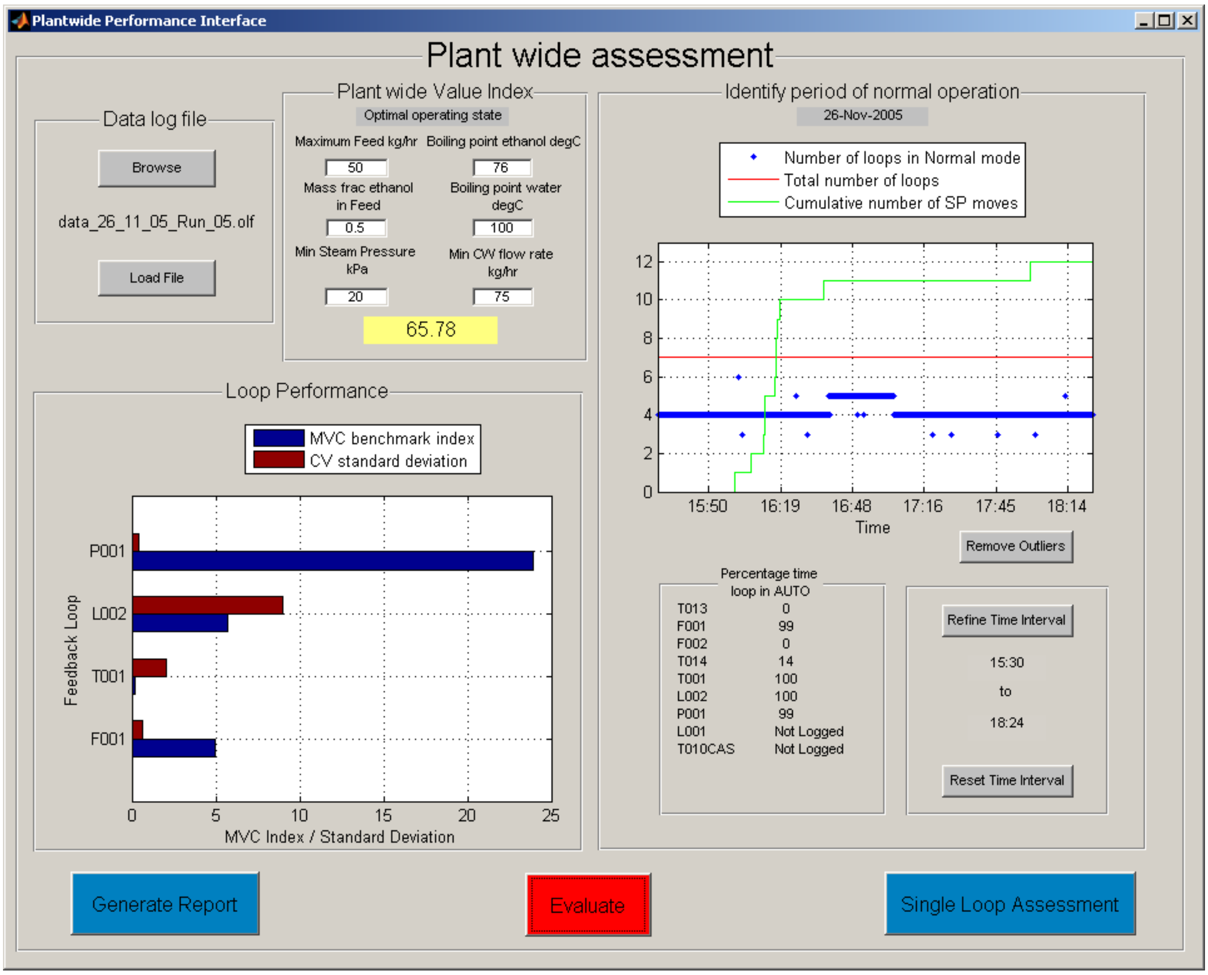


Figure 7.2: The plant wide interface evaluation for the period excluding shutdown.

of the plant seeing that no set-point changes occurred and the number of loops on AUTO also was constant. So whatever happened in those 9 minutes was probably continuing to happen in the period up to the first set-point change. The five minute interval at the end however had an effect seeing that the PWI increased from 65.45 to 65.78 due to the exclusion of only five minutes of shutdown that was originally included.

With the new evaluation period set, three distinct operating regions can be identified. In the first period four loops were in operation and then went over to a period where five loops were in operation and then back to the same loops that were originally operating. Table 7.1 shows the loops that were in operation. Refer to the PFD in appendix A for clarity on loop orientation. As can be seen from table 7.1, loop T014 was turned on

Table 7.1: The loops in operation during the three operating periods

Period	Loops in operation	Description
15:31 to 16:40	F001	Feed flow loop
	T001	Top plate temperature
	L002	Distillate drum level
	P001	Steam pressure control
16:40 to 17:05	F001	Feed flow loop
	T001	Top plate temperature
	L002	Distillate drum level
	P001	Steam pressure control
	T014	Reflux temperature
17:05 to 18:24	F001	Feed flow loop
	T001	Top plate temperature
	L002	Distillate drum level
	P001	Steam pressure control

AUTO for only 14% of the evaluation period. To determine why this is the case the single loop interface needs to be considered and this is done in section 7.1.2.

Regulatory performance assessment

To do a plantwide regulatory performance evaluation, the three periods shown in table 7.1 will be considered as three separate periods. The plant wide performance interface for the first period is shown in figure 7.3. As can be seen from figure 7.3, numerous set-point changes occurred during the first evaluation period. The set-point changes occurred for the top plate temperature loop as will be seen in the single loop evaluation of the loop in section 7.1.2. Some of the set-point changes are due to outliers in the set-point signal which need to be ignored. The PWI for the evaluation period was equal to 66.45. This value does not mean much at this stage seeing that this value has to be

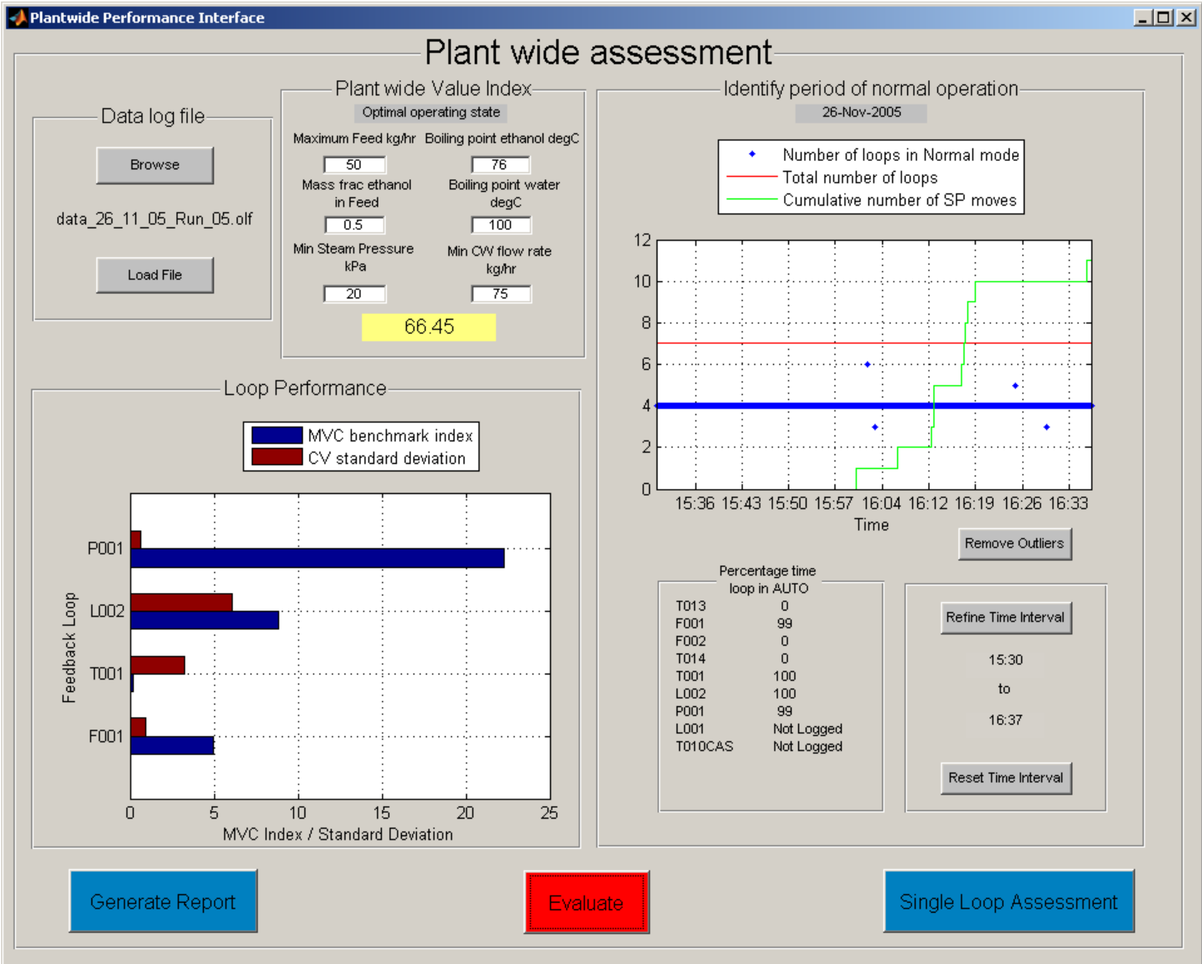


Figure 7.3: Plant wide regulatory assessment for the period from 15:31 to 16:40.

compared to other evaluation periods. Useful information is displayed however in the *Loop Performance* pane. We can see that the pressure and feed loops (P001 and F001) looked as though they were performing well seeing that they have a small CV standard deviation and a comparatively large MVC index. The temperature and level loops (T001 and L002) on the other hand seem suspicious seeing that their CV standard deviation is larger and the MVC index for the temperature loop is very small. Further single loop analysis is necessary but the initial feel is that loop L002 and T001 needs to be looked at. The seemingly poor performance of the level and temperature loop may be due to the non-regulatory period considered, this suspicion needs to be confirmed in the single loop assessment.

Next we consider the middle period of operation where five loops were in operation. The period is shown in figure 7.4. We can see that no set-point changes were made during

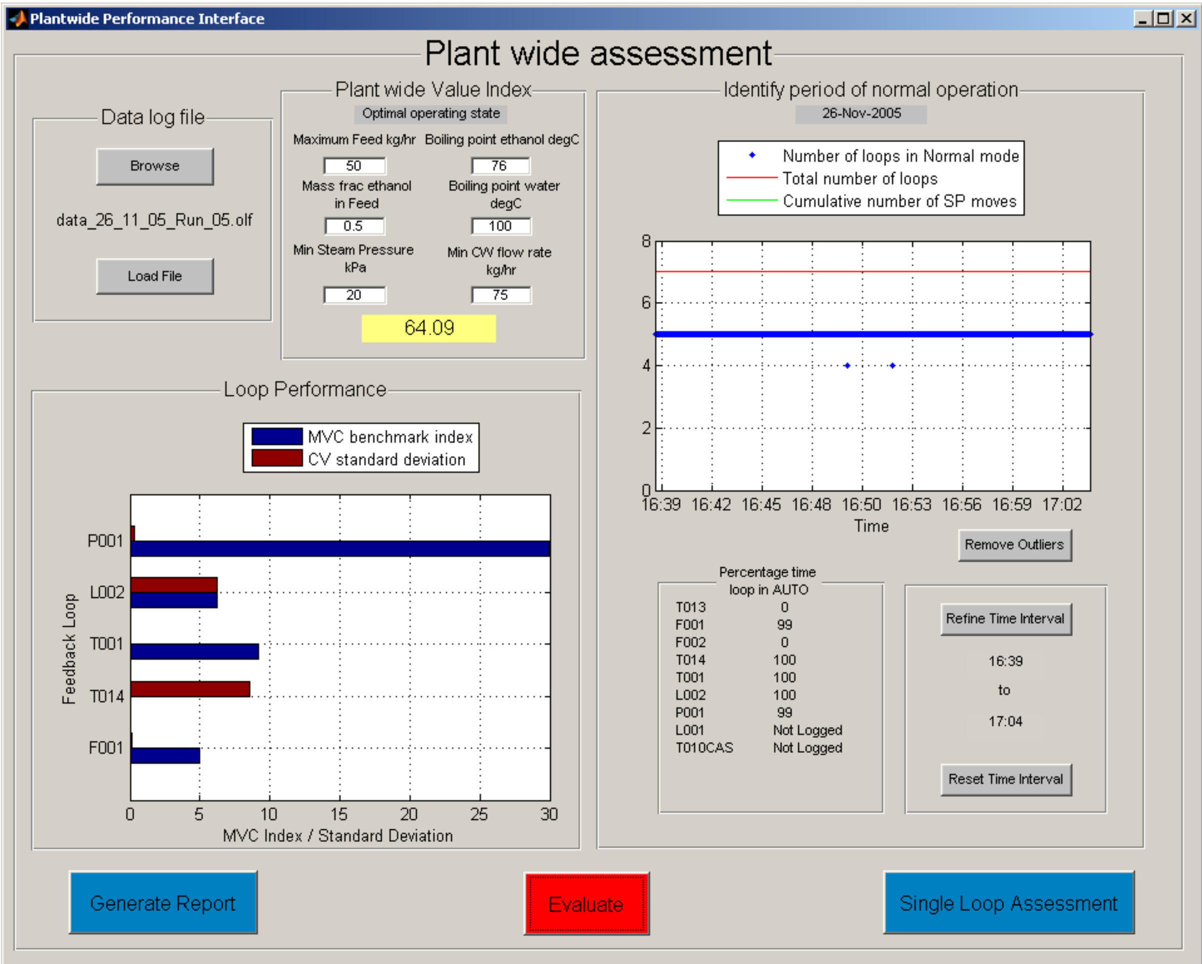


Figure 7.4: Plant wide regulatory assessment for the period from 16:40 to 17:05.

the evaluation period which makes it a pure regulatory control evaluation analysis. For this period the *PWI* for the evaluation period is equal to 64.09. This is less than for the initial period which means the plant performance has dropped for the second period. Once again similar to the first period, P001 and F001 seem to be doing well if we consider the *Loop Performance* pane. The top plate temperature, T001, also seems to be doing well seeing that the variance of the CV is near zero. This adds to the suspicion that the set-point changes that occurred in the loop for the first period was the source of variance, especially seeing that no set-point changes occurred for this period. The reflux temperature loop , T014, seem to be performing very poorly. The level loop, L002, seems to be performing a little worse when compared to the first period.

For the third period the evaluation is shown in figure 7.5. As can be seen from

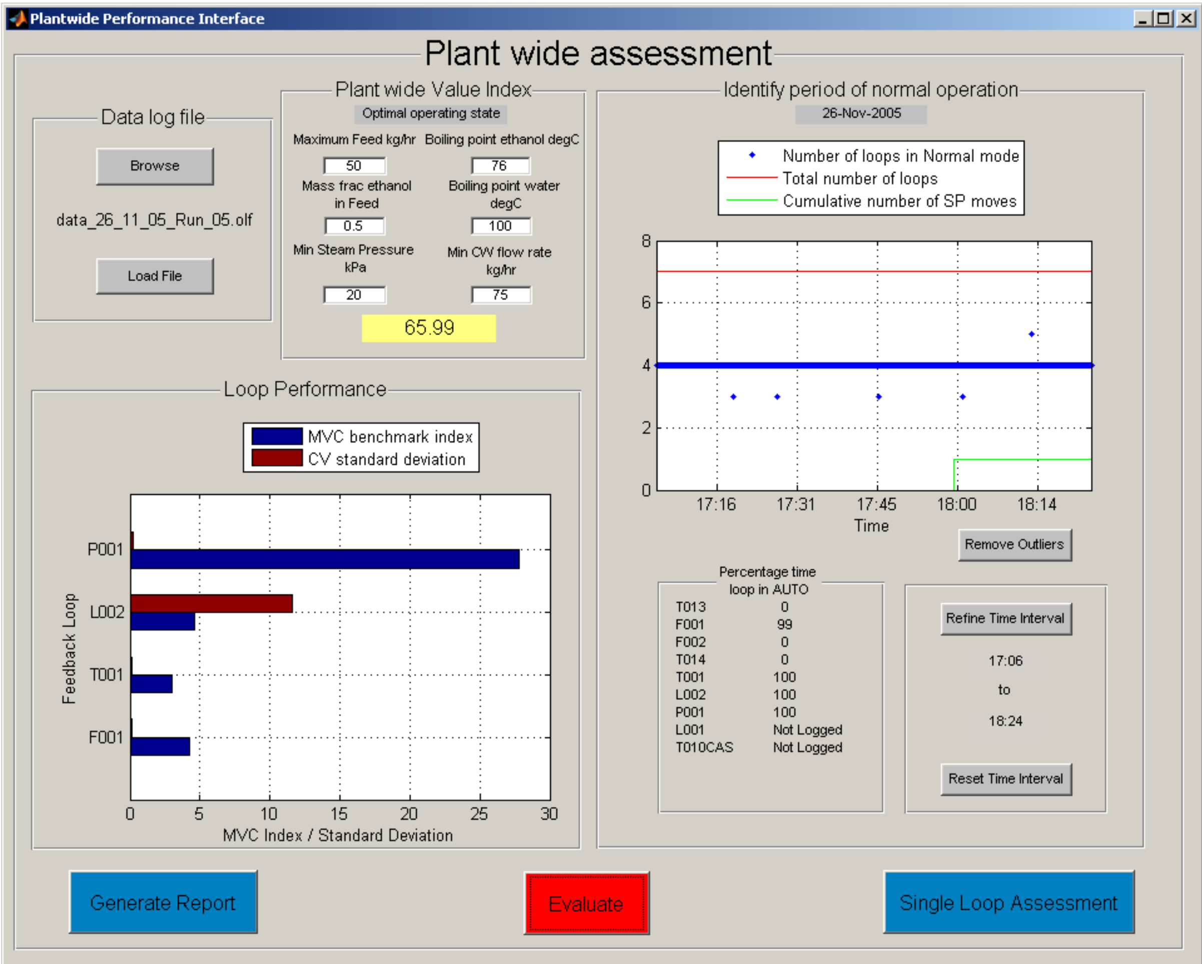


Figure 7.5: Plant wide regulatory assessment for the period from 17:06 to 18:24.

figure 7.5 only one set-point change occurred during the evaluation period. The *PWI* for the period is 65.99 which is better than the second period but still not as good as the first period. If the single loop performance is considered in the *Loop Performance* pane we see once again that the feed flow and steam pressure seems to be doing well. The top plate temperature, T001, perform reasonably but apparently worse than the second period according to the MVC index. The distillate level is suspected to have performed the worst in this period compared to the others.

In all the plots on the *Normal operation* pane there are some clear outliers. These outliers occur due to the determination of the AUTO or MANUAL mode calculation.

The mode is determined by the difference of the CV and SP and if they are exactly the same value (4 decimals) the loop is said to be on MANUAL. This method works seeing that the SP tag value takes on the CV value if the loop is on MANUAL. The outliers therefore occur if the loop is on AUTO and the CV and SP are exactly the same at a sampling instant. This occurs very rarely as can be seen from figure 7.5 where it occurred 5 times out of a sample set of 9349. These outliers can be removed by the *Remove Outliers* pushbutton. The number of outliers small and would have a minuscule effect on the calculations, so they were ignored.

Summary of plant wide results

Before the single loop interface is considered it is handy to summarise what insight has been gained into the performance of the process for the initial evaluation period. Possible problematic loops have been identified and their suboptimal performance need to be confirmed by using the single loop interface. The following can be concluded from the plant wide evaluation:

- The operation for the first evaluation period was the best according to the PWI.
- The reflux temperature loop, T014, performed badly for the periods on AUTO.
- The feed flow and steam pressure (F001 and P001) loops performed well for all the three periods.
- The top plate temperature, T001, performed badly in the first period and well after that.
- The level loop, L001, operated with large variance with best performance in the first evaluation period.
- The source of bad performance for the loops, T001 and L001, may be due to loop interaction of the two loops seeing that both loop MV's feed from the distillate drum.
- The first period of operation contained numerous set-point moves which could have affected the accuracy of the plantwide performance analysis in the first period.

This information now aids as a starting point for individual loop assessment. If most of the suspicions in the plantwide evaluation is confirmed by single loop assessment we have successfully reduced the dimensionality of the performance problem seeing that we have successfully identified problem cases by means of a plantwide evaluation.

7.1.2 Single loop evaluation

We now consider the serious cases picked up in the plant wide assessment first. As mentioned before from the data acquisition analysis the T010CAS loop needs to be commissioned. This will improve bottoms product quality considerably. The boiler level loop, L001, was evaluated through inspection of the data from the continuous historian. The loop performed moderately with some MV saturation. These loops need to be considered in more detail. The loops left on MANUAL are not critical in the column operation and are not considered further. The worst performing loop identified by the plantwide analysis is considered first.

Reflux temperature loop, T014

Feedback loop T014 was only in AUTO mode for 14% of the original evaluation time. To find out why this was the case, the single loop interface for the period on AUTO (figure 7.6) is considered. We immediately see from figure 7.6 that the loop is not at all

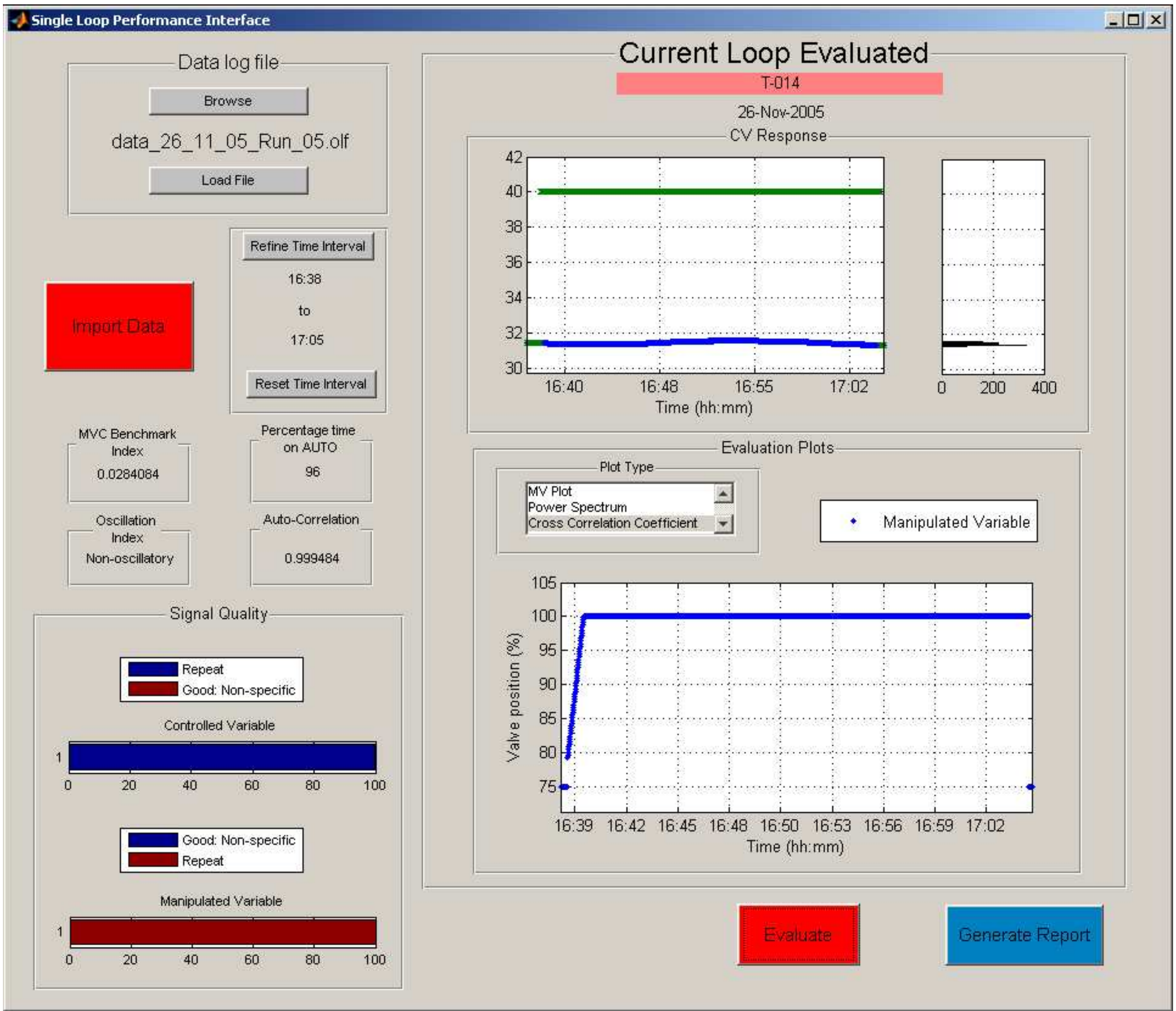


Figure 7.6: The single loop performance for T014 for the period it was on AUTO.

performing well. The control error is huge. The MVC benchmark index confirms this with a very low value of 0.028. It is seen from the evaluation plot that the MV became saturated which provides us with a clue as to why the performance is so bad. The MV is saturated on the high side which means the valve on the cooling water line is fully open. This is strange because the reflux temperature is too low and is below its set-point. The valve therefore should not be opening, it should be closing. The controller gain therefore has the wrong sign. The diagnosis is therefore that the controller tuning is insufficient and needs to be retuned.

Top plate temperature, T001

To consider the loop performance of the top plate temperature the single loop interface was implemented and is shown in figure 7.7. As can be seen from figure 7.7 the loop

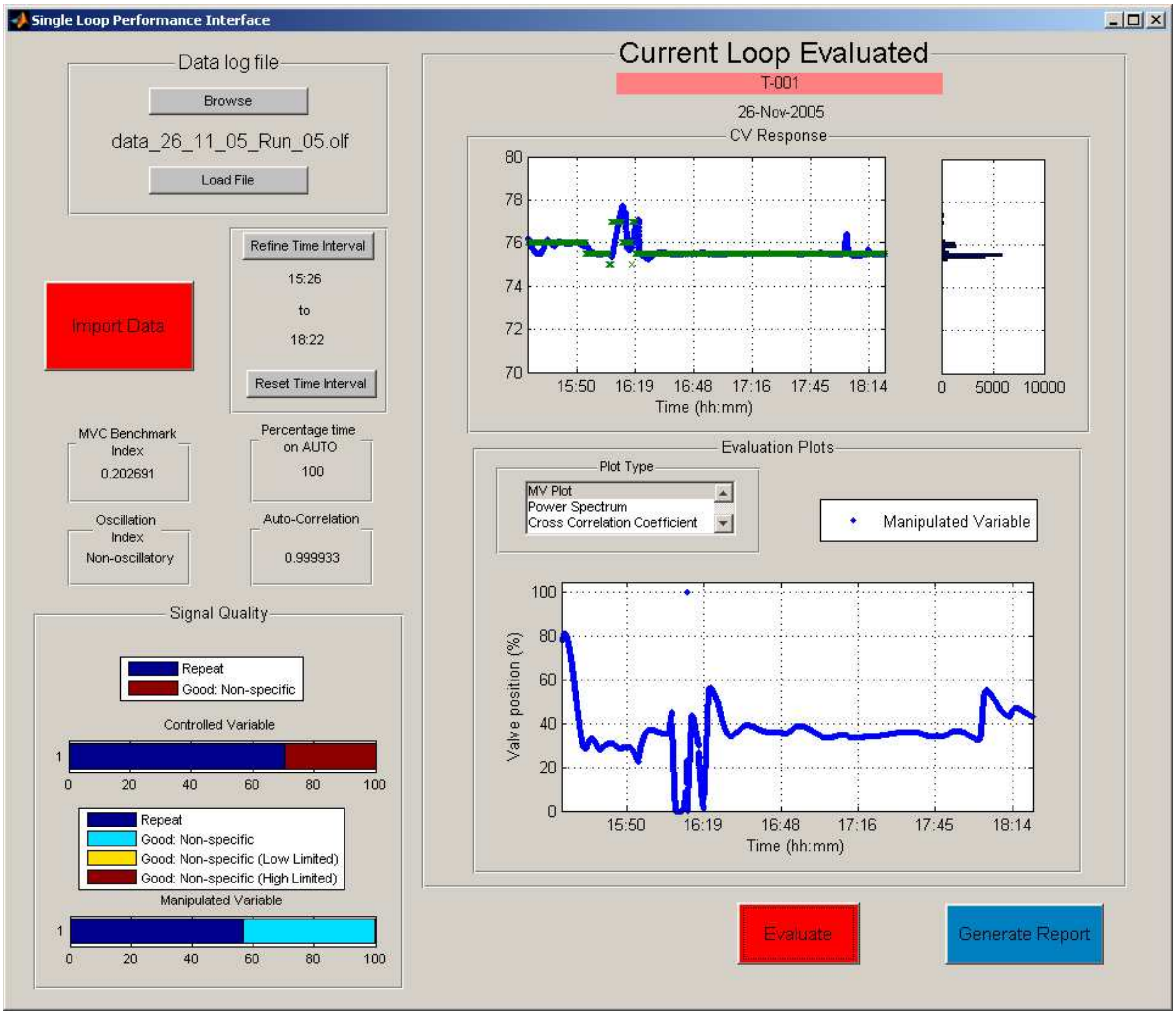


Figure 7.7: The single loop peformance for the top plate temperature, T001, over an evaluation period of 15:26 to 18:22.

performed well for the whole period of evaluation. The initial plant wide evaluation showed that the performance may be bad in the initial period of 15:30 to 16:37. We can now confirm that performance was not due to bad regulatory performance but rather due to set-point changes made to the loop. This identifies what happened in the first period of evaluation in the plant wide interface. The variance was therefore caused by set-point movements. The set-point movements also caused some MV saturation, but saturation during normal operation was not a problem. The MVC index also successfully located the set-point influence. If we consider the index for the specific loop over the three periods considered in the plant wide interface we get the result shown in table 7.2. From table 7.1

Table 7.2: The MVC index for the three evaluation periods applied in the plant wide analysis.

Evaluation period	MVC Index
15:27 to 18:37	0.17
16:39 to 17:05	11.13
17:05 to 18:20	4.14

it is clear that the loop performed best over the middle evaluation period and we know the poor performance in the first period was due to set-point movements. We need to identify why the last period of operation showed poorer performance according to the MVC index. To identify this we consider the single loop evaluation for the last period which is shown in figure 7.8. From figure 7.8 it is clear that some external disturbance affected the CV during this period which is also reflected in the MVC benchmark in table 7.2. This made the MV jump to a new state at around 18:00. Nothing seemingly changed in the configuration of the loop. The operating conditions during the three operating periods stayed relatively constant except for operator intervention through set-point changes. So we suspect that a set-point change occurred somewhere in the plant which affected the top plate temperature performance. This suspicion is confirmed by considering the last evaluation period in the plant wide interface (figure 7.5) where we saw a set-point change just before 18:00. The time from when the step disturbance occurred and when it was noted in the in the top plate temperature is in the range of three minutes. So if we have to locate the set-point disturbance source we consider the correlation coefficient plots at a lag of three minutes. If we look at the process configuration on the PFD in figure A.1 we immediately should suspect the level loop, L002, seeing that both MV's feed from the reflux drum. To confirm that the set-point indeed did occur in distillate drum level set-point we look at the L002 single loop evaluation in figure 7.9. It is clear from figure 7.9 that the set-point change did indeed occur in the distillate drum level controller. The set-point was changed at 17:59 from a value of 80 to 65. The degradation in performance was therefore due to a change in the set-point in the level loop that has interacted with the temperature loop. The interaction happened when the level decreased the head for the reflux flow decreased which meant a reduction in flow. To compensate for this the

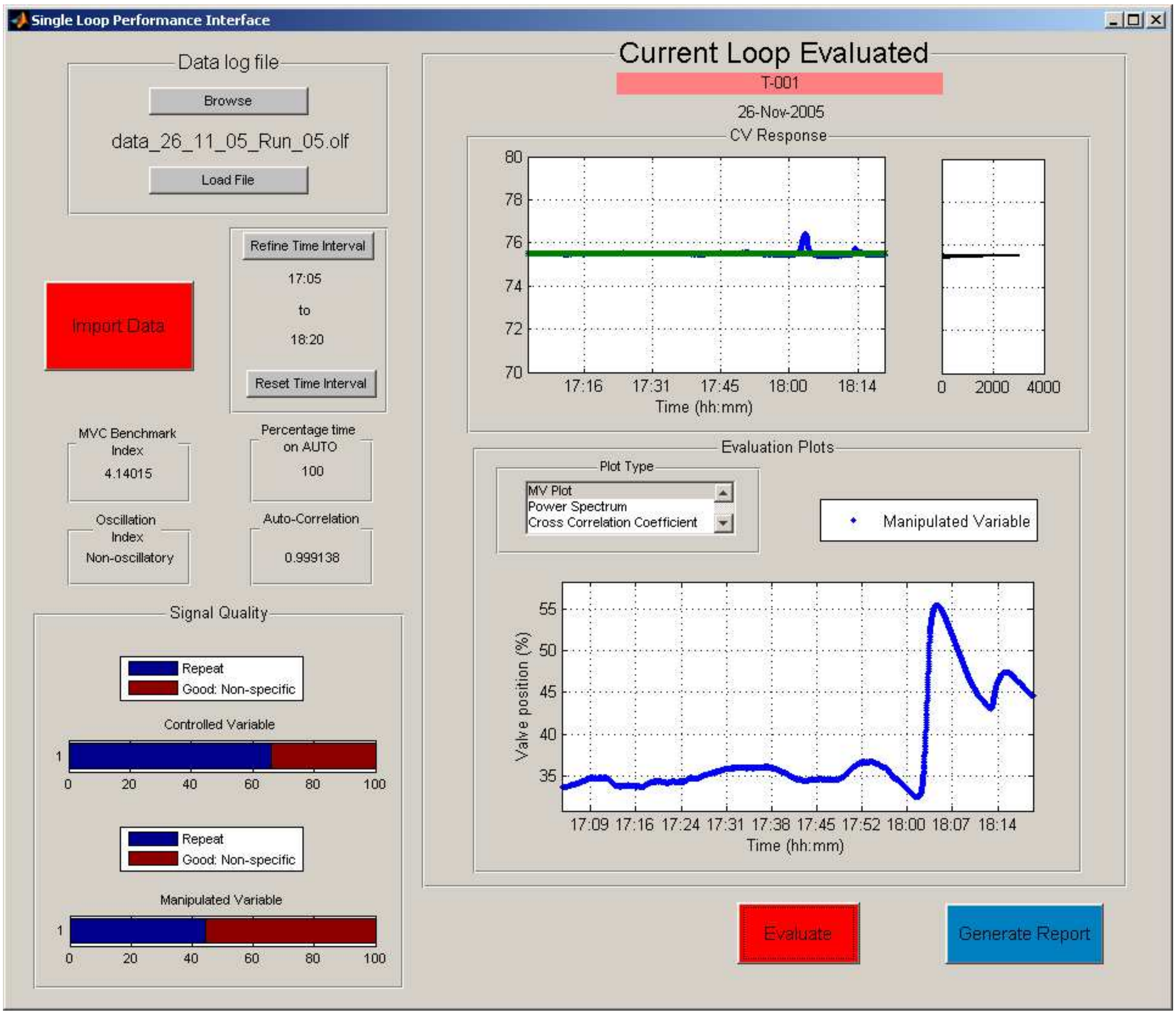


Figure 7.8: The single loop assessment for T001 for the last period of evaluation.

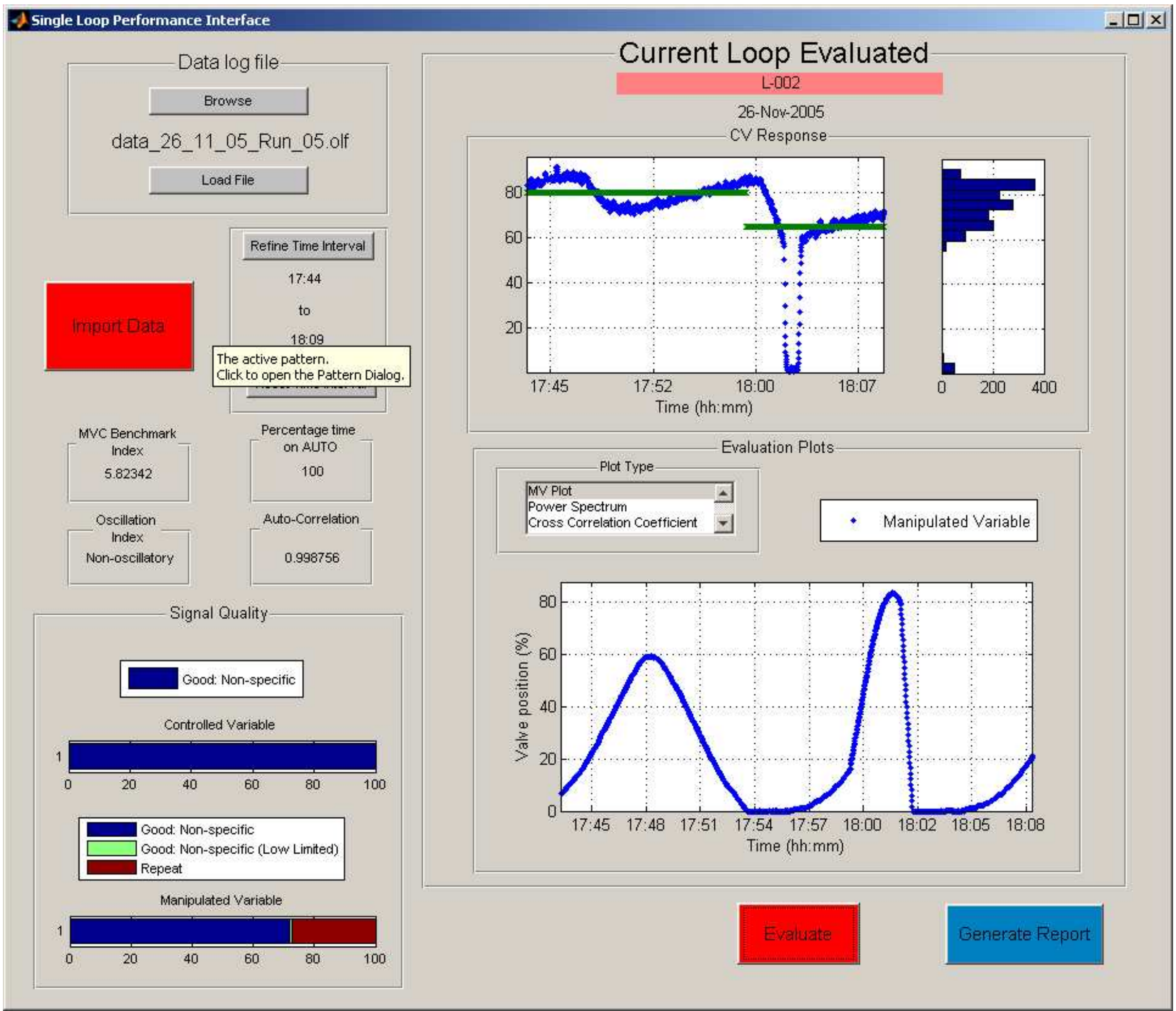


Figure 7.9: The set-point change in level loop, L002, can be seen in the time series plot. This set-point change caused the disturbance in the top plate temperature.

temperature valve was opened by the T001 controller to maintain temperature set-point (see figure 7.8 for MV movement).

Distillate drum level, L002

It is handy to consider the actual distillate drum configuration in figure 7.10 while we consider the level controller performance. Two important facts should be noted when we

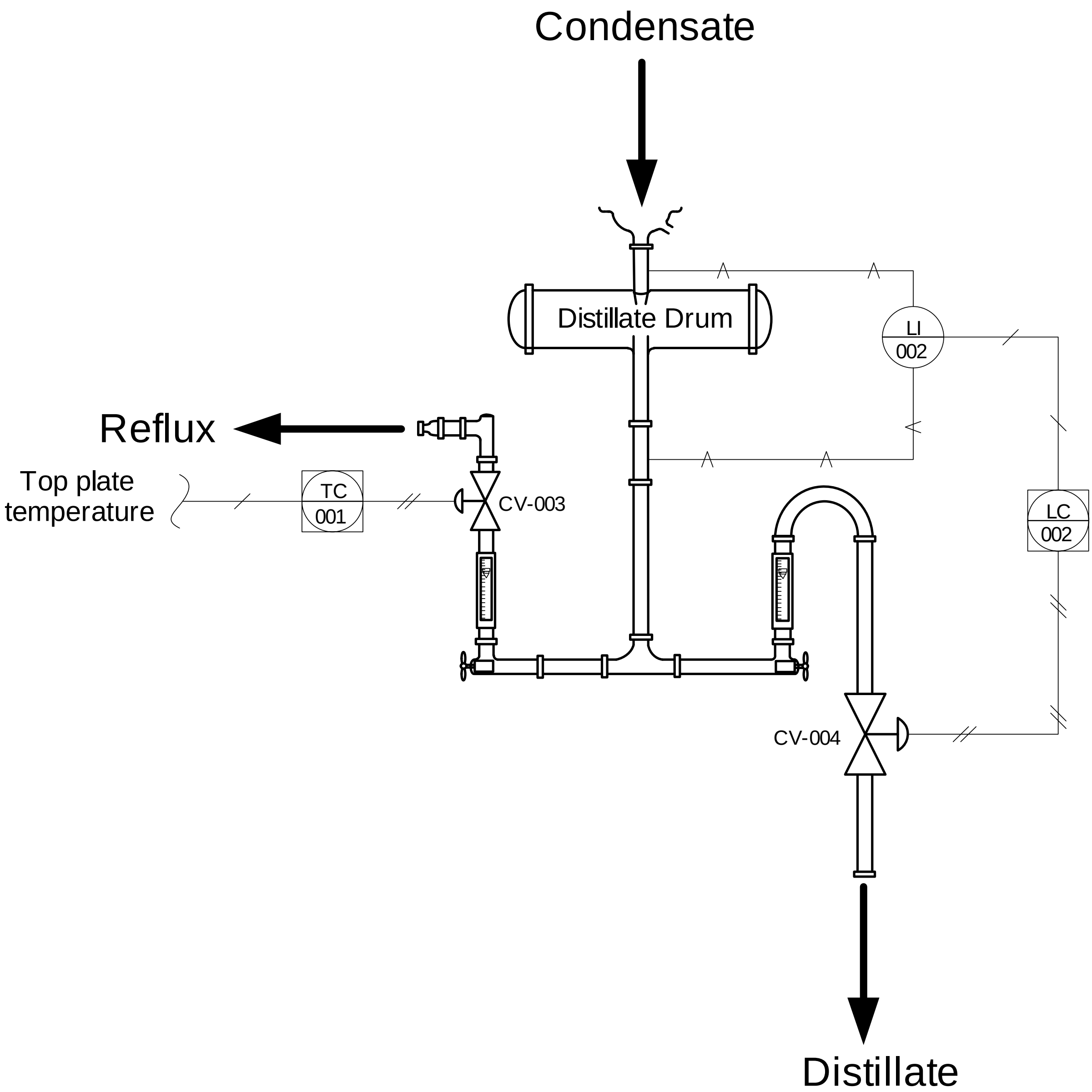


Figure 7.10: The distillate drum configuration on the distillation column set-up.

consider the configuration in figure 7.10. One is that the distillate drum level provides the static head for flow back into the column as reflux, and for flow out of the system as distillate product. This causes interaction of the temperature and level loops as was noted in the previous section. The second important fact is the differential pressure measurement ports over the drum. The ports are above and below the drum with the distances on the figure representative of the actual distances on the rig. This means that the changes in the differential pressure readings will not be linear with changes in the real level. The shape of the drum is cylindrical confirming the non-linearity of the system. All the controllers on the column are linear feedback controllers and their performance are expected to bad for non-linear systems.

The single loop evaluation is shown in figure 7.11. From figure 7.11 we can start

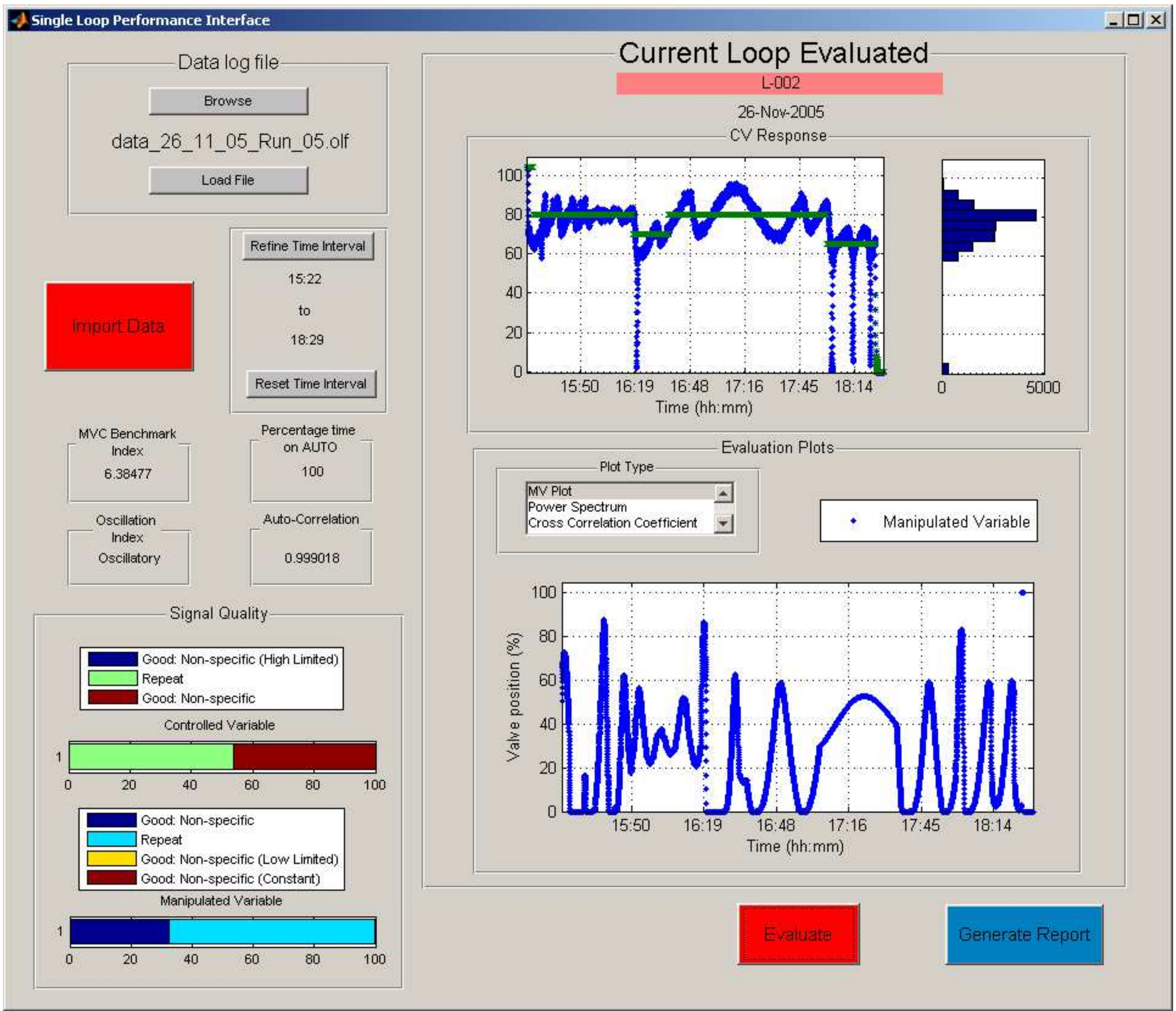


Figure 7.11: The single loop evaluation of the distillate drum level controller, L002.

making a qualitative performance assessment. First we see a lot variance in the CV which is not very good as was predicted by the plant wide assessment. Secondly we see a lot of MV movement with considerable saturation. Thirdly we see that the dynamics of the CV between 60% and a 100% is completely different from the dynamics between 0% and 60%. When the level measurement is between 60% and a 100%, the response is slow and oscillating but when the level goes below 60%, it almost immediately drops to empty and then quickly recovers again to 60%. This is due to the non-linearity caused by the DP-cell port placement shown in figure 7.10. When the level changes quickly (below 60%) it means that no level exists in the drum and level only exists in the pipe that connects the drum at the bottom. The CV tracks the set-point well when no MV saturation happens as can be seen from the first part of the evaluation period up to the first set-point change.

Two periods of regulatory operation will now be considered. The first part is shown in figure 7.12. The evaluation in figure 7.12 shows reasonable control with the CV settling

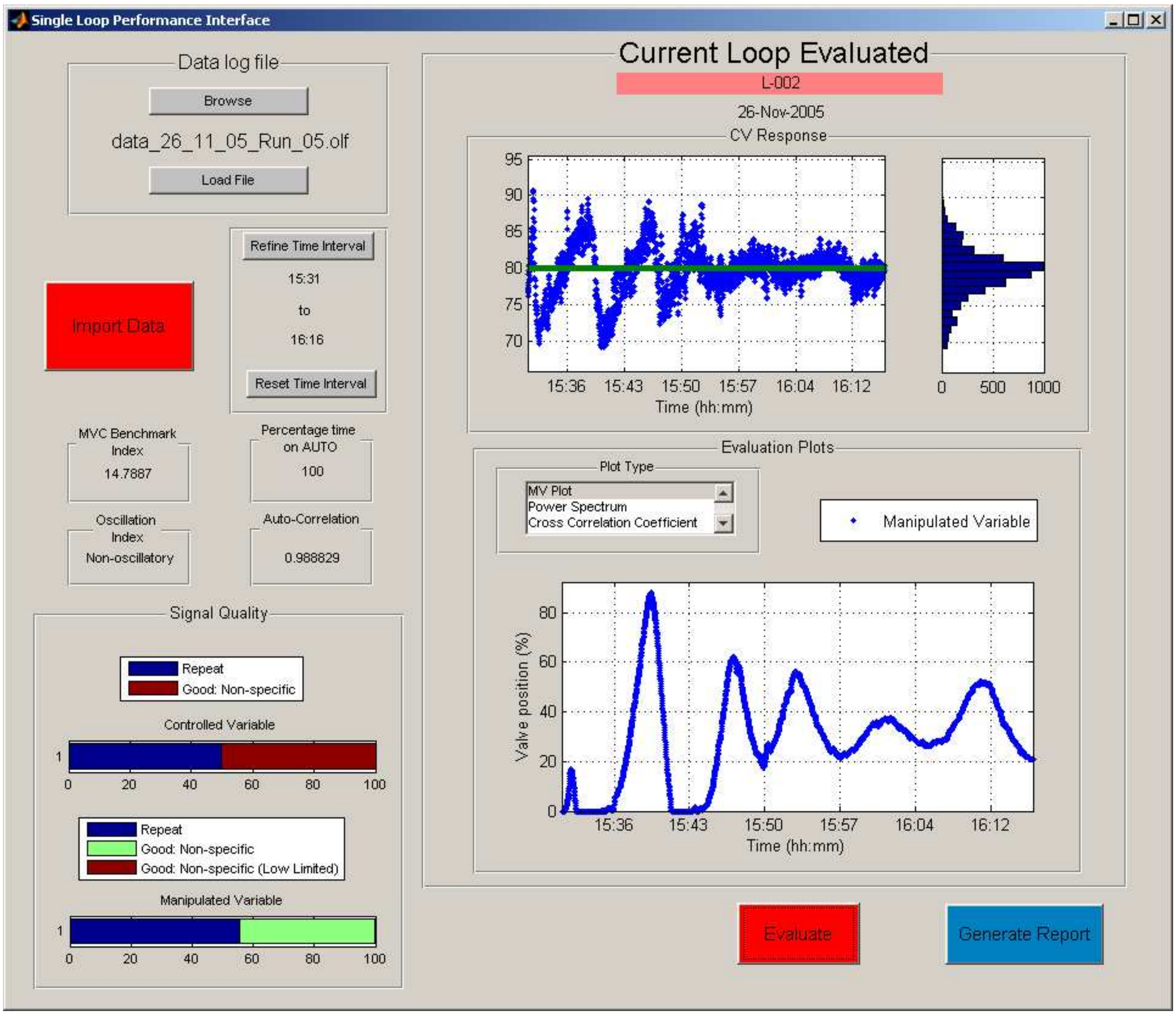


Figure 7.12: The single loop evaluation of L002 from 15:31 to 16:16.

nicely to the set-point. The distribution plot also shows a narrow tall peak which is a good sign. The MV showed only minimal saturation. The MVC benchmark index was 14.79 for the period. If we compare this period of operation with the period shown in figure 7.13 we clearly see that the performance has considerably deteriorated. The MVC benchmark reflects this with a value of 7.11, half of the previous period.

The oscillation index in figure 7.13 shows "Non-Oscillatory" while the response shows clear oscillations. This is due to the fact that the period of evaluation is too short. The limit on the number of consecutive loads is 10 (section 6.2.6) and there is clearly only 6 loads that occurred with 7 control sign changes. Further investigation was done on why

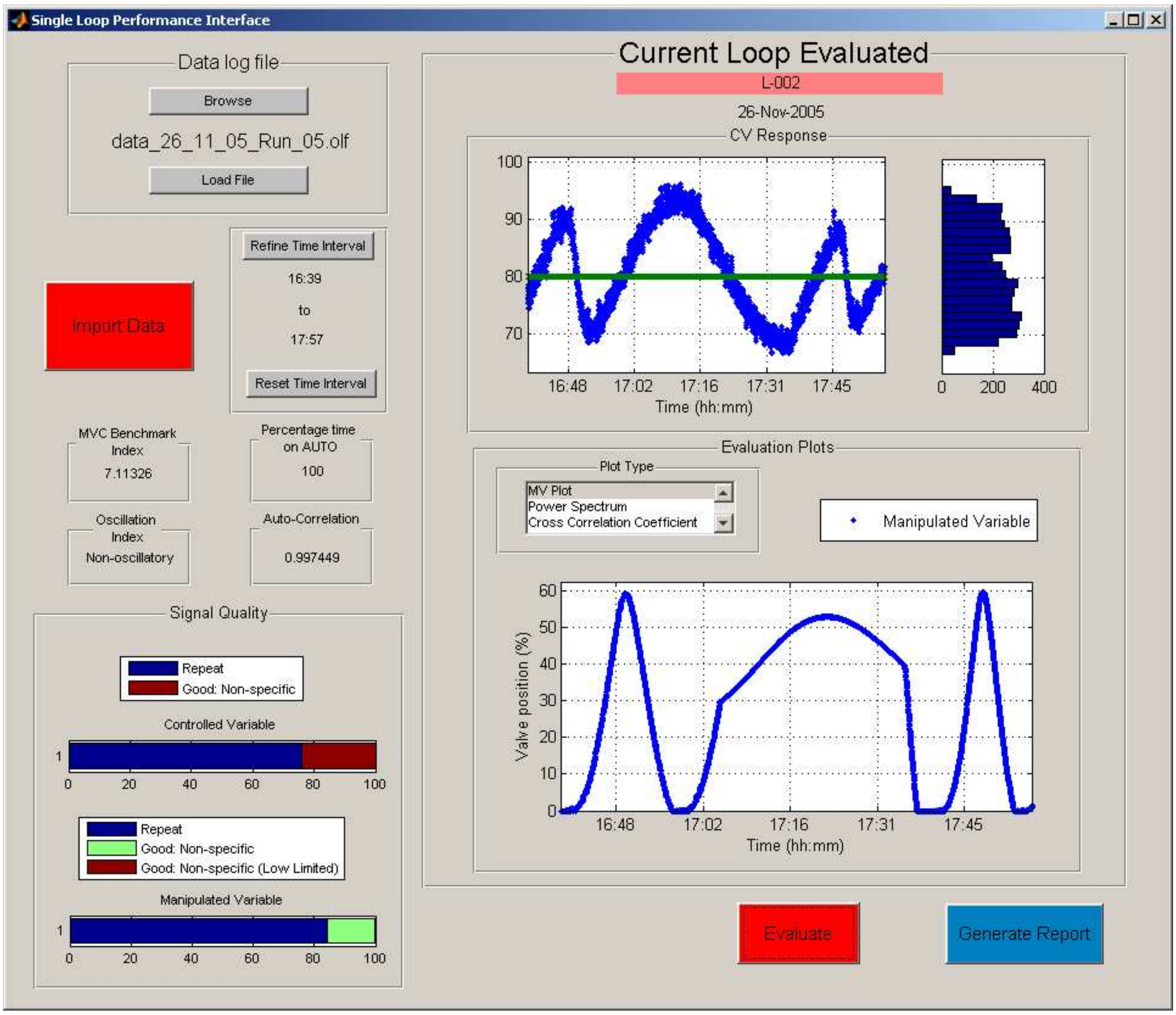


Figure 7.13: The single loop evaluation of L002 from 16:39 to 17:57.

the performance degraded for the second period of evaluation (figure 7.13). The initial thought was that loop interaction from set-point changes to the top plate temperature loop was the cause, but seeing that the first part of operation actually contained all the temperature set-point changes (see figure 7.7) this could not have been seeing that in the period of set-point changes the performance was better than when no set-point changes occurred in loop T001. This still doesn't mean that the poor performance is not due to interaction, consider the two evaluation periods again in figures 7.14 and 7.15.

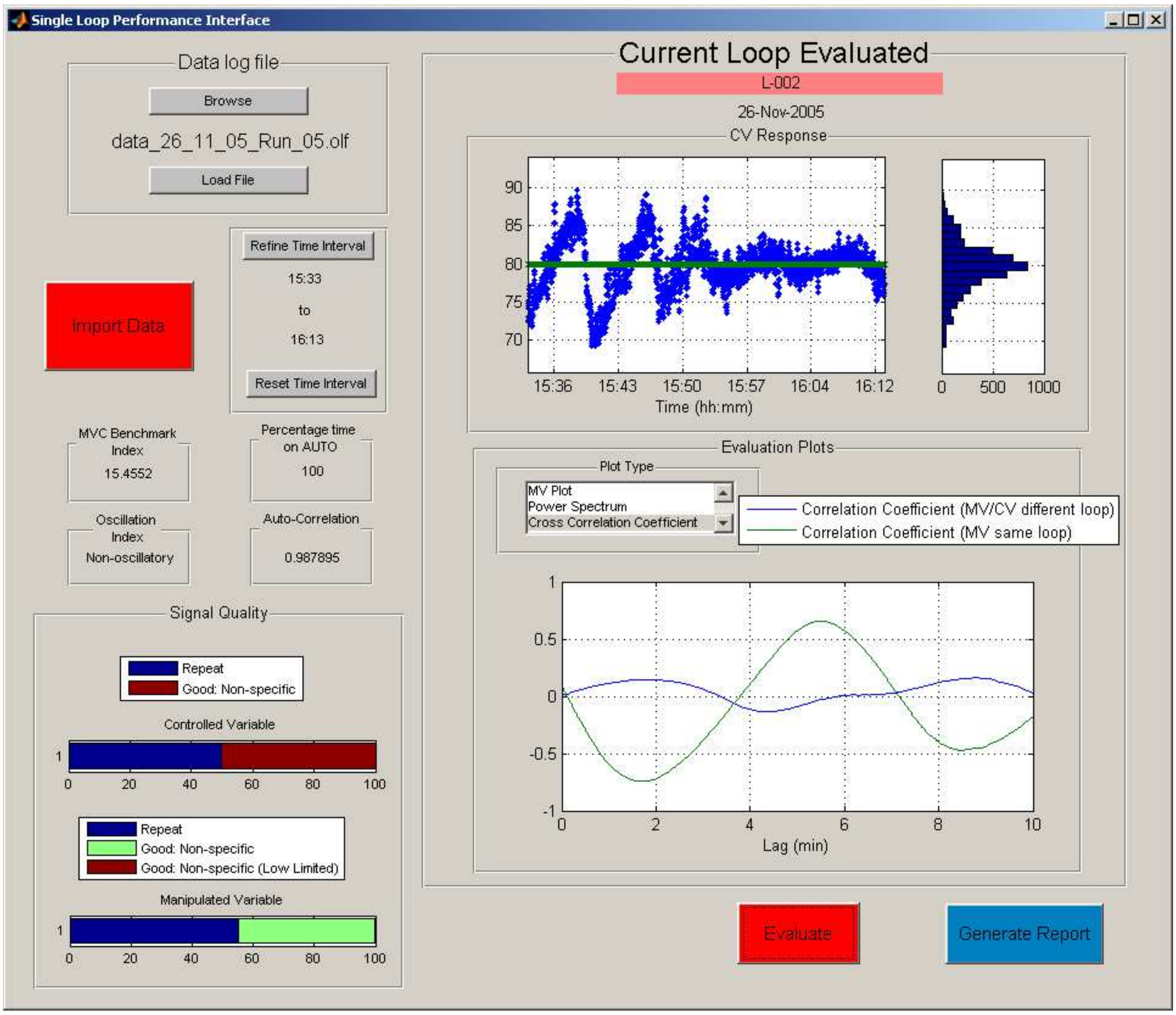


Figure 7.14: The single loop evaluation of L002 from 15:33 to 16:13 with the cross correlation coefficients.

The figures 7.14 and 7.15 show the correlation between the level and the distillate valve position (green trend) as well as the level and the reflux valve position (blue trend). From the figures it can clearly be noted that in the first evaluation period the interaction of the temperature loop was minimal because the blue trend is close to zero for all the considered lags. The second period shown in figure 7.15 has a lot of interaction. Further investigation showed that a number of controller parameters have been used during the evaluation periods. This is evident in the MV and CV time series plots shown

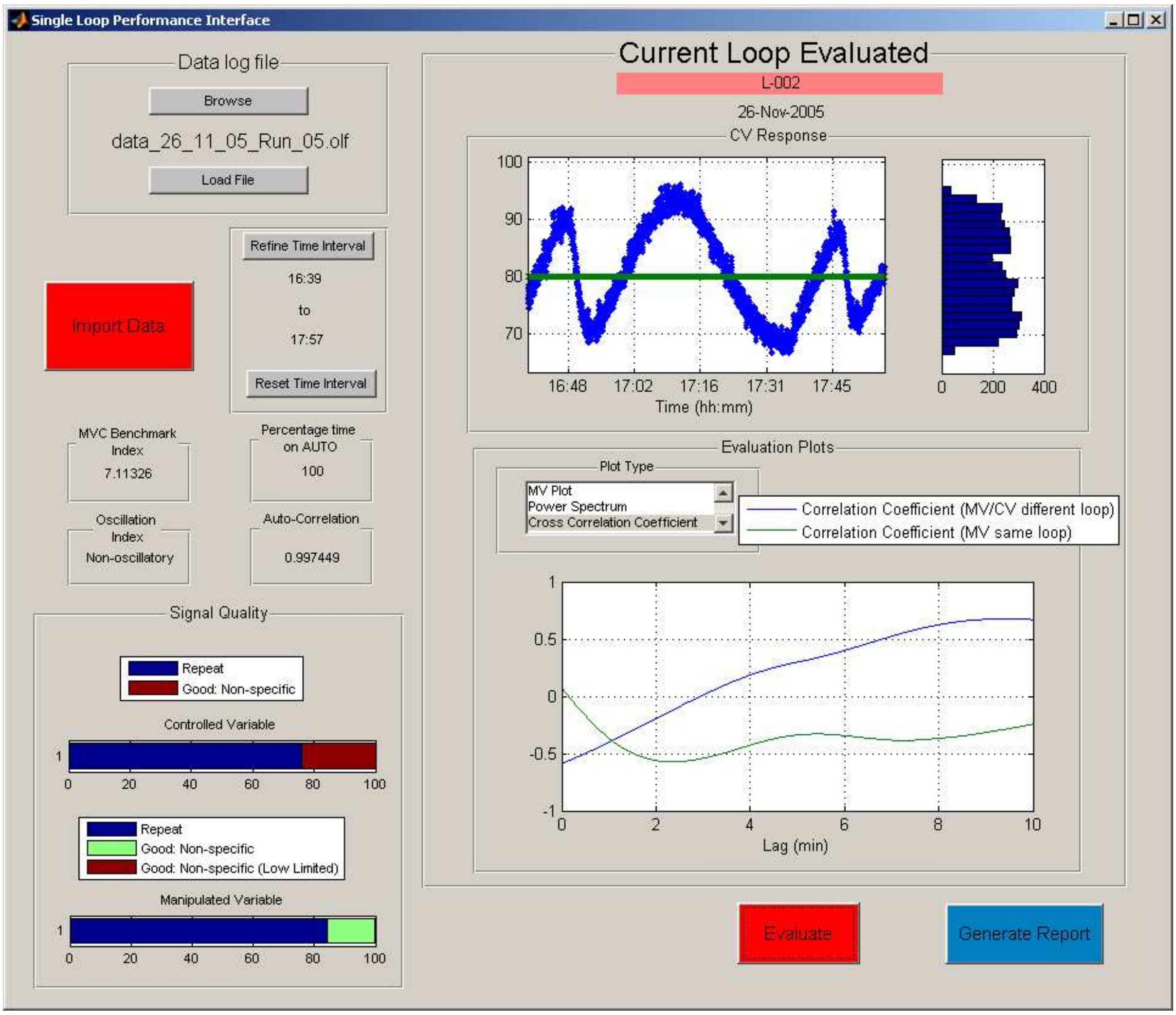


Figure 7.15: The single loop evaluation of L002 from 15:33 to 16:13 with the cross correlation coefficients.

in figure 7.11. Table 7.3 shows the controller parameters that have been in operation. From the controller parameters and the figures it can be seen that the addition of the

Table 7.3: The controller parameters for the distillate drum level controller, L002.

Period	Proportional Gain	Integral time (sec)
15:22 to 16:33	2	0
16:33 to 17:05	4	10
17:05 to 17:35	0.5	10
17:35 to 18:30	2	2

integral time to the control strategy has slowed down the controller action. The MV didn't move as quickly as without the integral time. This also increased the variable interaction seeing that the effect of the MV on consecutive controller executions has been reduced. It is recommended that the controller be properly tuned to make sure that the distillate flow rate is the governing variable that determines level in the drum.

Feed flow rate, F001

The single loop evaluation of the feed flow rate is shown in figure 7.16. It can be seen from the single loop evaluation that the performance of feed flow rate loop is very good. There is some variance in the initial period when the set-point was 35 kg/hr but stabilised well when the set-point was changed to 30 kg/hr. The initial set-point of 35 kg/hr may be the reason why the PWI performed so well in the initial evaluation period considered in section 6.1. At the end of the evaluation period there a short period where the column was in shut down mode. This single loop evaluation confirms the predicted good performance by the plant wide interface evaluation.

Steam pressure, P001

The single loop evaluation of the steam pressure loop is shown in figure 7.17. From figure 7.17 we can see some oscillating behaviour but not large enough to be detected by the Hägglund (1995) algorithm. The pressure stays close to the set-point for the entire evaluation period. In the initial periods two instances occurred where the steam kettles tripped and no steam supply was available to the reboiler. This occurred at the start of the evaluation period and reasonable quality steam supply was delivered after that. The oscillating nature of the supply is due to the on and off switching of the kettle. The kettle produces steam semi-continuously. It boils-up for a certain amount of time then switches off which makes the supply drop then when the pressure drops below a certain limit it switches on again and so the process continues. The controller compensates for these pressure changes in the supply and provides steam with minuscule deviations (± 1 kPa)

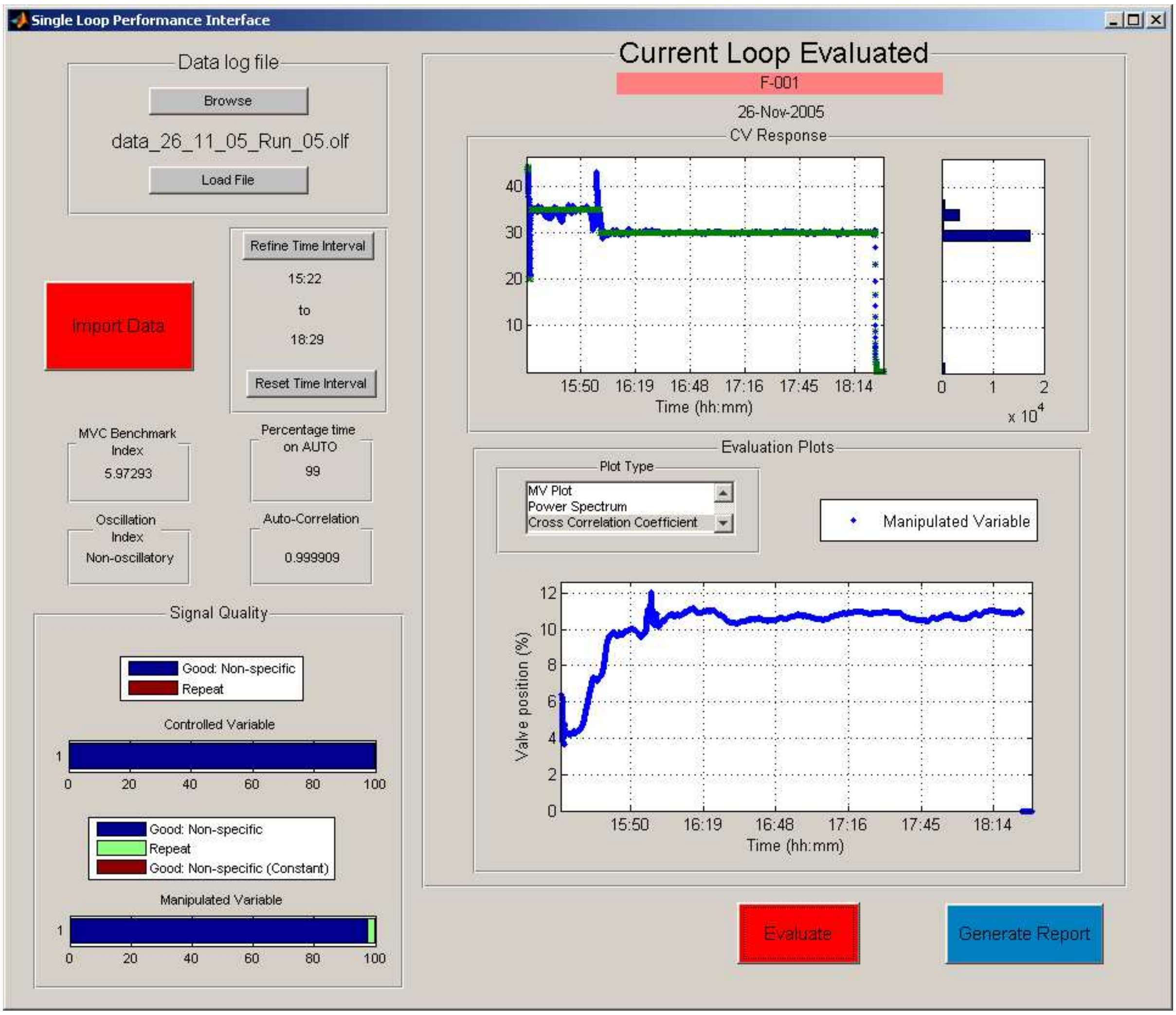


Figure 7.16: The feed flow rate over the entire initial eavluation period.

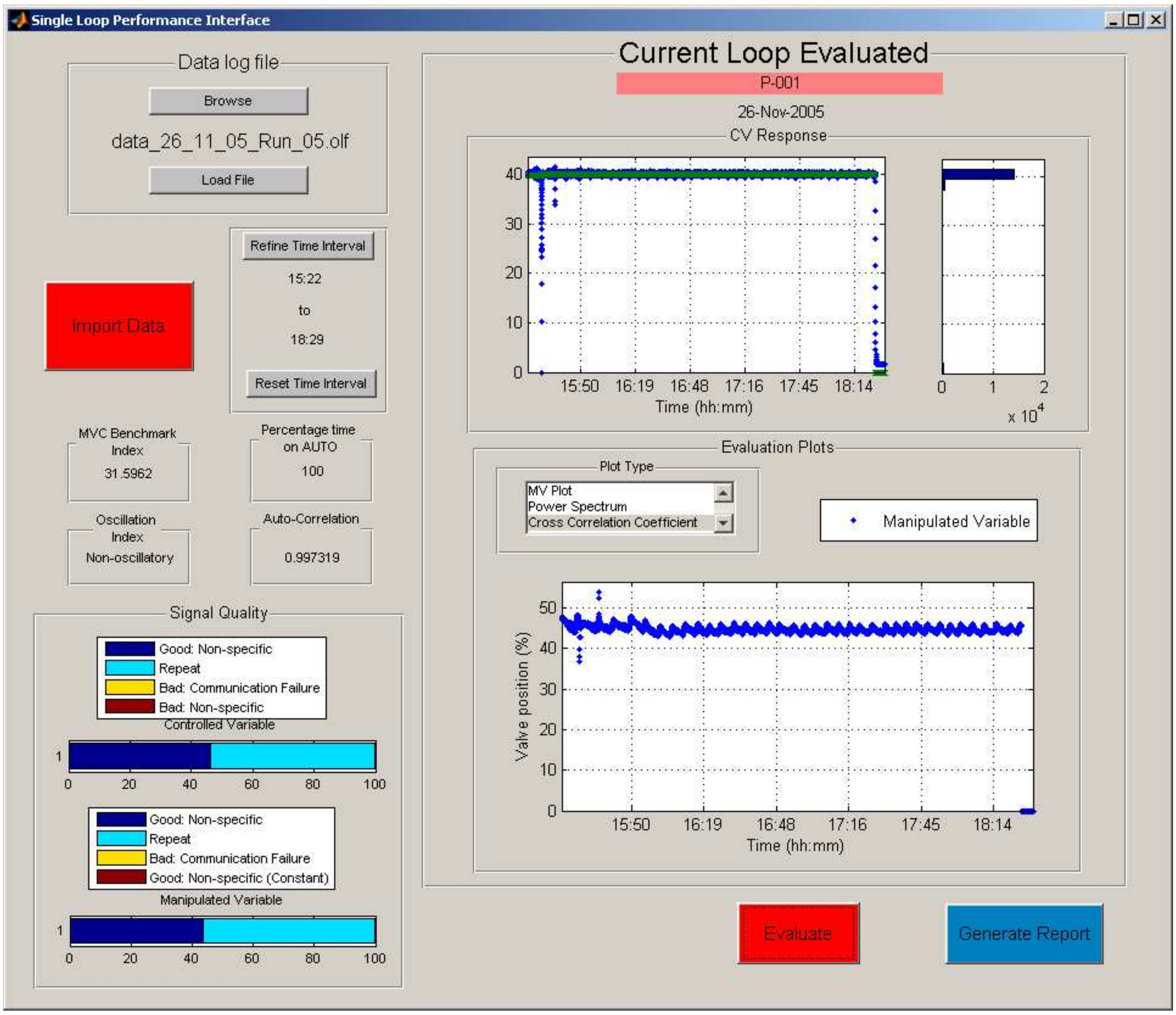


Figure 7.17: The steam pressure supply to reboiler single loop evaluation.

to the reboiler which indicates good performance. Actuator wear may become a problem for the steam valve seeing that it is constantly doing quick adjustments but they are small, however.

Just to show the application of the power spectral density plot consider figure 7.18. One can see several large peaks in the frequency region between 3×10^{-3} s^{-1} and $1.5 \times$

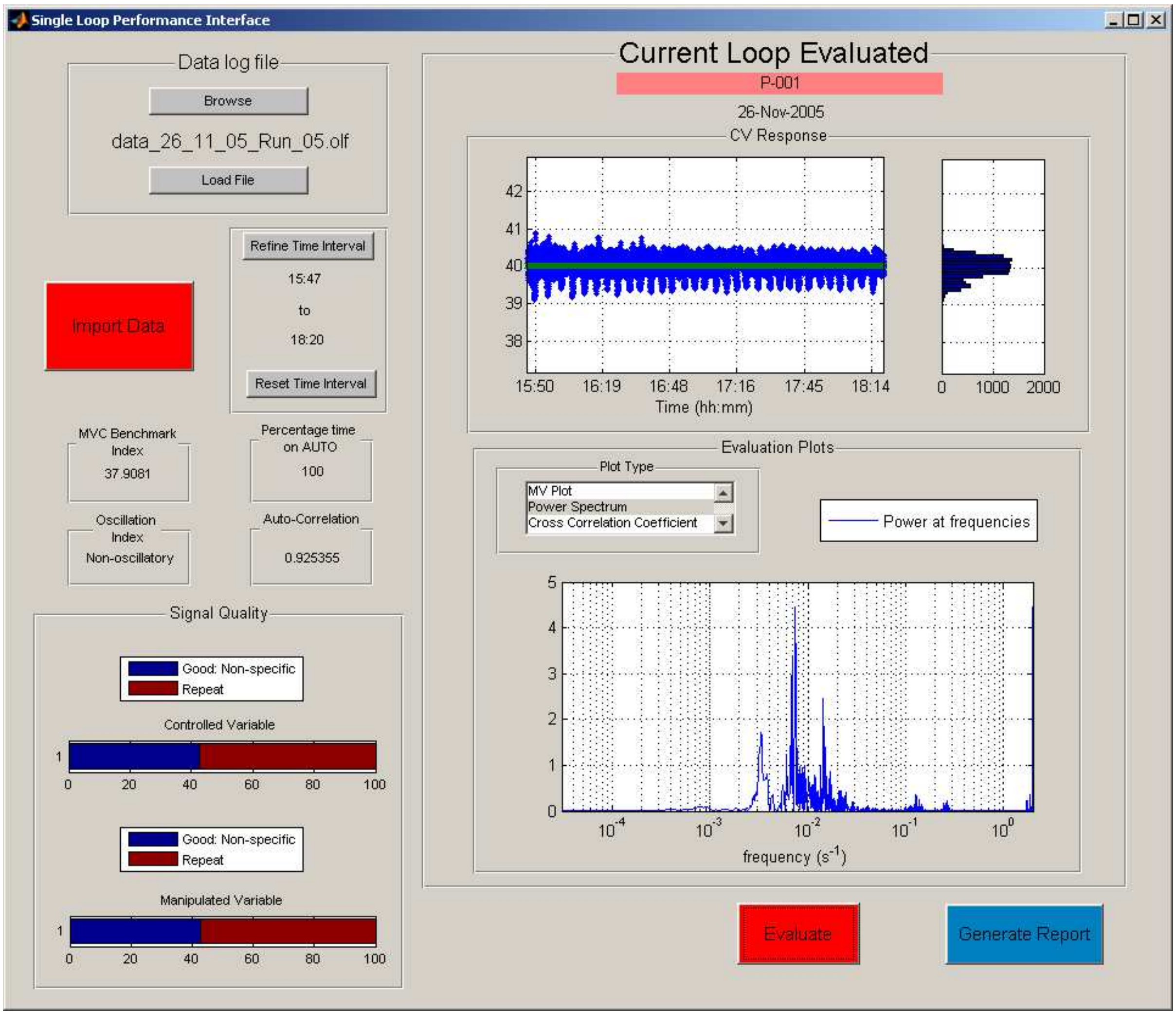


Figure 7.18: The single loop evaluation of the pressure supply to the reboiler with a PSD evaluation plot for the period from 15:47 to 18:20.

10^{-2} s^{-1}. This corresponds to a time period of 5.5 to 1.1 minutes. This is typically periods of oscillation that we don't want in the response but looking at the magnitudes of the PSD as well as the actual deviations from set-point the oscillations can be tolerated. To prove that the periods of oscillation provided by the PSD are correct, consider the refined time period of the steam pressure time series plot in figure 7.19. From figure 7.19 we can see the period where dips in the steam pressure supply occur is roughly 4 minutes apart which is in the range identified by the original PSD plot shown in figure 7.18.

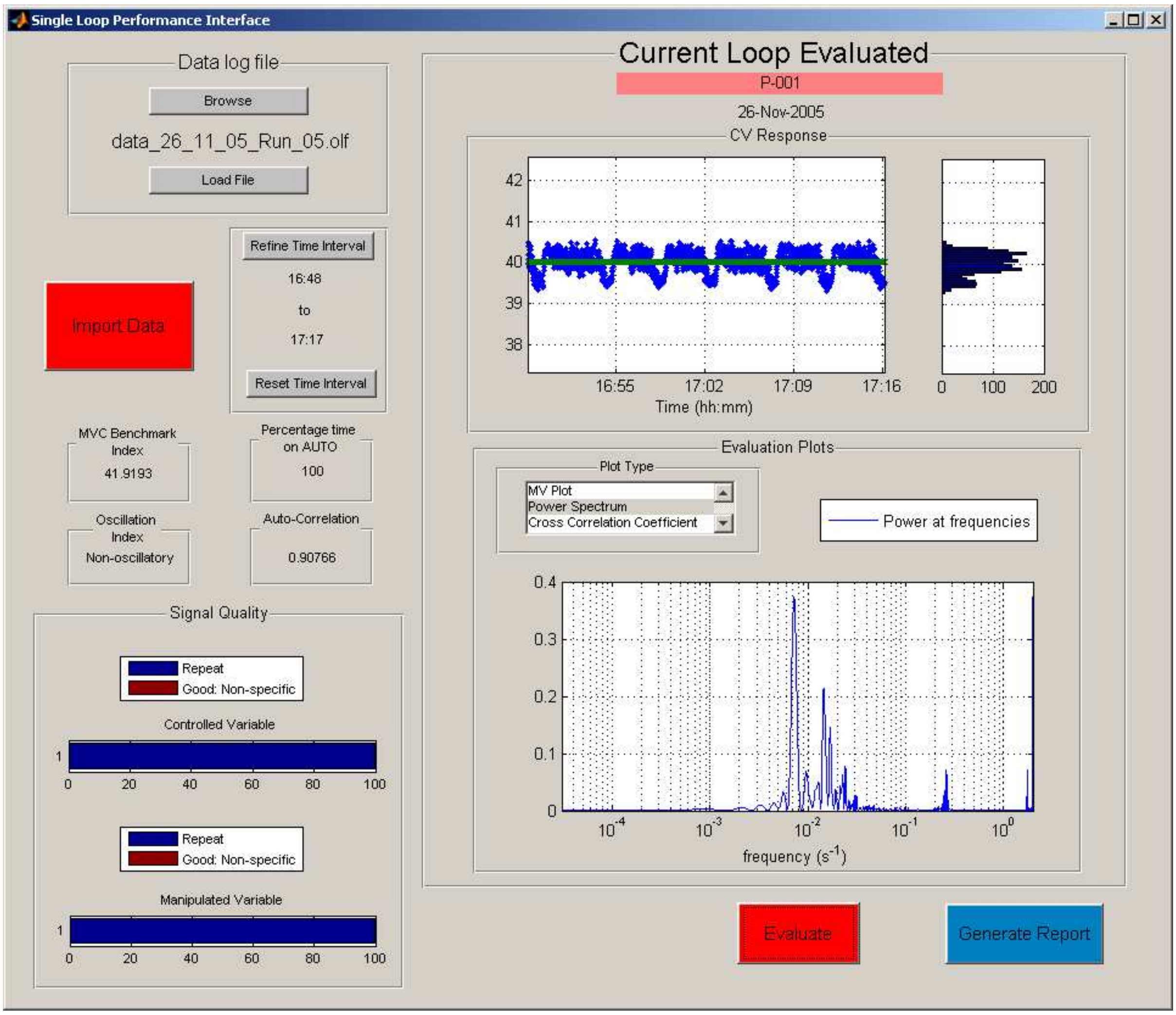


Figure 7.19: The time series plot of the steam pressure for a 29 minute period from 16:48 to 17:17.

CHAPTER 8

Conclusions and recommendations

This chapter is a summary of the project outputs and provides recommendations on future work.

8.1 Monitoring structure development

A performance monitoring structure was developed and implemented on an industrial standard lab scale distillation column. The process structure developed was based on a generic methodology that can be applied to any processing facility. The structure consists of a plant wide interface that is used to give a plant wide indication of performance and indicates possible areas of improvement. A single loop evaluation interface is used to investigate the identified sources of poor performance in the plant wide interface.

The structure should be used as a tool for normal regulatory performance assessment. It indicates where sources of bad performance are located in the process as well as possible causes for this. The structure is not a specific and direct indication of performance, it is a tool that should be interpreted by a person familiar to the relevant processing environment.

8.2 Monitoring structure application

The developed structure was applied to an industry representative lab-scale distillation column. It was found that the structure can evaluate performance in a quantitative way by performing comparative evaluation against previous periods of operation as well as against optimal benchmarks. The data capturing method was found to be insufficient due to its semi-continuous nature. The programming done was kept as generic as possible but in some instances, like referencing to tag names, it is not.

8.3 Future work

The following items have been identified as areas for possible future work:

- Data capturing - Data is available from the DeltaV continuous historian, so technically no other method of data capturing should be necessary. The problem is, however, that no method exists to import data directly from the historian to the MATLAB environment. A means should be developed to retrieve data from the continuous historian directly to replace the current data retrieval method used by the performance interfaces. It should be remembered however that the continuous historian compresses data and some dynamic behaviour may be lost.

- Automated monitoring - The performance structure developed could be adapted into an automated evaluation structure. This will mean that the performance evaluation does not have to be triggered and is performed automatically.

- Multivariate analysis - Extra multivariate analysis techniques can be added to the interface to supplement the cross correlation methods used in this research.

- Data filtering - No data filtering except for some scaling and de-trending was used in this investigation. Filtering techniques should be investigated and implemented to make calculations less computationally intensive.

- Non-linearity - Most of the techniques implemented are linear techniques and techniques for non-linear analysis should be looked at.

- Non-regulatory performance - The structure could be extended to include start-up and shut-down performance evaluation.

- APC application - If APC techniques become part of the normal control philosophy the performance structure should be extended to include this. This structure can then aid in APC justification.

- Real time optimisation - Capability to write information back to the DCS to enable changes to the control structure to improve performance on an automated basis (especially on the regulatory level).

- Commercial software - There are a number of regulatory performance software packages available on the market. They can be implemented on the process and critically evaluated to show functionality and shortcomings.

BIBLIOGRAPHY

Albright, S. C.; Winston, W. L. and Zappe, C. (2002) *Data analysis and decision making with Microsoft Excel*, Duxbury, Pacific Grove.

Åström, K. J. and Wittenmark, B. (1984) *Computer Controlled Systems*, Prentice-Hall, New York.

Biegler, L. T. and Grossmann, I. E. (2004) "Retrospective on optimization", *Computers and Chemical Engineering, Article in Press.*

Blevins, T. L.; McMillan, G. K.; Wojsznis, W. K. and Brown, M. W. (2003) *Advanced Control Unleashed*, ISA, Research Triangle Park, N.C.

Box, G. E. P.; Jenkins, G. M. and Reinsel, G. C. (1994) *Time series analysis: Forecasting and control*, Prentice Hall, third edition.

Buckley, P. S. (1964) *Techniques of Process Control*, Wiley and Sons, New York.

Clegg, A.; Xia, H. and Uduehi, D. "Controller benchmarking: From single loops to plant-wide economic assessment", Presentation - internet website: http://www.isc-ltd.com/benchmark/learning_centre/ Date accessed: 30/12/05 (2005).

do Carmo C. de Vargas, V.; Lopes, L. F. D. and Souza, A. M. (2004) "Comparative study of the performance of the cusum and ewma control charts", *Computers and Industrial Engineering, 46*, 707–724.

Duval, R. "Introductory quantitative methods (ps601) part I", Presentation - internet website: www.polsci.wvu.edu/duval/PS601/Notes/601Notes.ppt Date accessed: 30/12/05 (2005).

Edgar, T. F. and Ko, B.-S. (2004) "PID control performance assessment: the single-loop case", *American Institute of Chemical Engineering Journal, 50*, 1211–1218.

Expertune "Advanced version loop optimization tools (required for DCS systems)", Internet website: http://www.expertune.com/advanced.html Date accessed: 30/12/05 (2005).

Fischer-Rosemount *Installing your DeltaV ^{TM} Automation System*, seventh edition (2003).

Gao, J.; Patwardhan, R. S.; Akamutso, K.; Hashimoto, Y.; Emoto, G.; Shah, S. L. and Huang, B. (2003) "Performance evaluation of two industrial mpc controllers", *Control Engineering Practice,* (11), 1371–1387.

Georgakis, C.; Uztürk, D.; Subramanian, S. and Vinson, D. R. (2003) "On the operability of continuous processes", *Control Engineering Practice, 11*, 859–869.

Grimble, M. J. (2002) "Controller performance benchmarking and tuning using generalised minimum variance control", *Automatica, 38*, 2111–2119.

Hägglund, T. (1995) "A control-loop performance monitor", *Control Engineering Practice, 3* (11), 1543–1551.

Harris, T. J.; Boudreau, F. and MacGregor, J. F. (1996) "Performance assessment of multivariable feedback controllers", *Automatica, 32* (11), 1505–1518.

Harris, T. J. (1989) "Assessment of control loop performance", *Canadian Journal of Chemical Engineering, 67*, 856–861.

Harris, T. J.; Seppala, C. T. and Desborough, L. D. (1999) "A review of process monitoring and assessment techniques for univariate and multivariate control systems", *Journal of Process Control, 9*, 1–17.

Hong, X. "Lecture notes: Minimum variance control", Internet website: www.personal.reading.ac.uk/ sis01xh 28 December (2005).

Huang, B. (2003) "A pragmatic approach towards assessment of control loop performance", *International Journal of Adaptive Control and Signal Processing, 17*, 489–608.

Huang, B.; Ding, S. X. and Thornhill, N. (2004) "Practical solutions to multivariate feedback control performance assessment problem: reduced a priori knowledge of interactor matrices", *Journal of Process Control, Article in press.*

Huang, B. and Shah, S. (1999) *Performance assessment of control loops*, Springer, London.

Kano, M.; Hasebe, S.; Hashimoto and Ohno, H. (2004) "Evolution of multivariate statistical process control: application of independent component analysis and external analysis", *Computers and chemical engineering, 28*, 1157–1166.

Kramer, H. and Schmid, W. (1997) "Control charts for time series", *Nonlinear Analysis Theory, Methods and Applications, 30* (7), 4007–4016.

Lee, J.-M.; Yoo, C. and Lee, I.-B. (2004) "Statistical process monitoring with independent component analysis", *Journal of Process Control, 14*, 467–485.

Luyben, W. L. (1990) *Process modelling, simulation and control for chemical engineers*, McGraw-Hill, New York.

Marlin, T. E. (1995) *Process Control*, McGraw-Hill, New York.

Mathworks *Matlab Help (Release 7 sp 2)*, January (2005).

Montgomery, D. C. (1985) *Introduction to statistical quality control*, John Wiley & Sons, New York.

Mosca, E. and Agnoloni, T. (2003) "Closed-loop monitoring for early detection of performance losses in feedback-control systems", *Automatica,* (39), 2071–2084.

NIST-SEMATECH "e-Handbook of Statistical Methods", Internet website: http://www.itl.nist.gov/div898/handbook/ Date accessed: 30/12/05 (2005).

Perry, R. H. and Green, D. W. (1998) *Perry's Chemical Engineers' Handbook*, McGraw-Hill, seventh edition.

Salsbury, T. I. (2004) "A practical method for assessing the performance of control loops subject to random load changes", *Journal of Process Control, Article in Press.*

Sandrock, C. "MATLAB structure conversion to xml", Personal communication, University of Pretoria (2005).

Schäfer, J. and Cinar, A. (2004) "Multivariable mpc system performance assessment, monitoring, and diagnosis", *Journal of Process Control, 14*, 113–129.

Seborg, D. E.; Edgar, T. F. and Mellichamp, D. A. (2004) *Process Dynamics and Control*, Wiley, New York.

Shewhart, W. A. (1931) *Economic control of manufactured product*, Van Nostrand, New York.

Shunta, J. P. (1995) *Achieving world class manufacturing through process control*, Prentice-Hall, Englewood Cliffs, N.J.

Skogestad, S. (2003) "Simple analytic rules for model reduction and pid controller tuning", *Journal of Process Control,* (13), 291–309.

Stapenhurst, T. (2005) *Mastering statistical process control: A handbook for performance improvement using cases*, Elsevier, Burlington.

Stephanopolous, G. (1984) *Chemical Process Control: An Introduction to Theory and Practice*, Prentice-Hall, Englewood Cliffs, N.J.

Subramanian, S. and Georgakis, C. (2001) "Steady-state operabilty characteristics of idealized reactors", *Chemical Engineering Science, 56*, 5111–5130.

Swartz, C. L. E. (1996) "A computational framework for dynamic operability assessment", *Computers and Chemical Engineering, 20* (4), 365–371.

Thornhill, N.; Oettinger, M. and Fedenczuk, P. (1999) "Refinery-wide control loop performance assessment", *Journal of Process Control, 9*, 109–124.

Uztürk, D. and Georgakis, C. (2002) "Inherent dynamic operability of processes: General definitions and analysis of siso cases", *Industrial and Engineering Chemistry Research, 41*, 421–432.

Vinson, D. R. and Georgakis, C. (2000) "A new measure of process output controllability", *Journal of Process Control, 10*, 185–194.

Xia, C. and Howell, J. (2003) "Loop status monitoring and fault location", *Journal of Process Control, 13*, 679–691.

Yang, J. and Makis, V. (1997) "On the performance of classical control charts applied to process residuals", *Computers and Industrial Engineering, 33* (1-2), 121–124.

Zhang, J.; Martin, E. B. and Morris, A. J. (1997) "Process monitoring using non-linear statistical techniques", *Chemical Engineering Journal,* (67), 181–189.

Zheng, A. and Mahajanam, R. V. "A quantative controllability index and its applications", University of Massachussetts (1999).

APPENDIX A

The process flow diagram

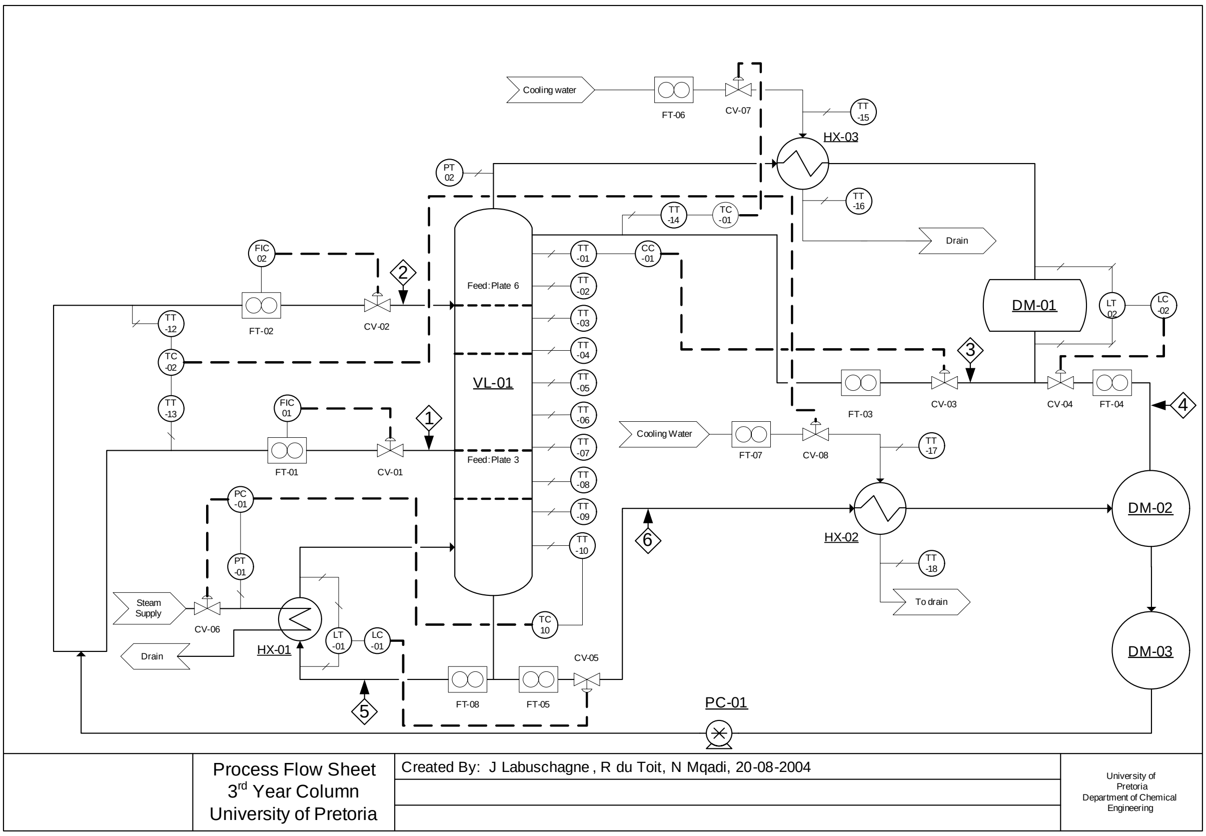


Figure A.1: The Process Flow Diagram

APPENDIX B

The process information logged via OPC

Table B.1: The process variables logged in MATLAB with their corresponding tag numbers.

Process Variable	Information	Tag number on server
Feed on plate 3	Actual flow rate	RLSA004F001/PID1/PV.CV
(PI controlled)	Set-point	RLSA004F001/PID1/SP.CV
	Controller mode	RLSA004F001/PID1/MODE.ACTUAL
	Valve position	RLSA004F001/PID1/OUT.CV
Feed on plate 6	Actual flow rate	RLSA004F002/PID1/PV.CV
(PI controlled)	Set-point	RLSA004F002/PID1/SP.CV
	Controller mode	RLSA004F002/PID1/MODE.ACTUAL
	Valve position	RLSA004F002/PID1/OUT.CV
Steam supply pressure	Actual pressure	RLSA004P001/PID1/PV.CV
(Primary cascade controlled)	Set-point	RLSA004P001/PID1/SP.CV
	Controller mode	RLSA004P001/PID1/MODE.ACTUAL
	Valve position	RLSA004P001/PID1/OUT.CV
Reboiler Level	Actual level	RLSA004L001/PID1/PV.CV
(PI controlled)	Set-point	RLSA004L001/PID1/SP.CV
	Controller mode	RLSA004L001/PID1/MODE.ACTUAL
	Valve position	RLSA004L001/PID1/OUT.CV
Condenser Level	Actual level	RLSA004L002/PID1/PV.CV
(PI controlled)	Set-point	RLSA004L002/PID1/SP.CV
	Controller mode	RLSA004L002/PID1/MODE.ACTUAL
	Valve position	RLSA004L002/PID1/OUT.CV
Top plate temperature	Actual temperature	RLSA004T001/PID1/PV.CV
(PI controlled)	Set-point	RLSA004T001/PID1/SP.CV
	Controller mode	RLSA004T001/PID1/MODE.ACTUAL
	Valve position	RLSA004T001/PID1/OUT.CV
Bottom plate temperature	Actual temperature	RLSA004T010CAS/PID1/PV.CV
(Secondary cascade controlled)	Set-point	RLSA004T010CAS/PID1/SP.CV
	Controller mode	RLSA004T010CAS/PID1/MODE.ACTUAL
	Valve position	RLSA004T010CAS/PID1/OUT.CV
Reflux temperature	Actual temperature	RLSA004T014/PID1/PV.CV
(PI controlled)	Set-point	RLSA004T014/PID1/SP.CV
	Controller mode	RLSA004T014/PID1/MODE.ACTUAL
	Valve position	RLSA004T014/PID1/OUT.CV
Feed temperature	Actual temperature	RLSA004T013/PID1/PV.CV
(PI controlled)	Set-point	RLSA004T013/PID1/SP.CV
	Controller mode	RLSA004T013/PID1/MODE.ACTUAL
	Valve position	RLSA004T013/PID1/OUT.CV
Reflux flow rate	Actual flow rate	RLSA004F003/AI1/PV.CV
Distillate flow rate	Actual flow rate	RLSA004F004/AI1/PV.CV
Bottoms flow rate	Actual flow rate	RLSA004F005/AI1/PV.CV
CW flow rate (condenser)	Actual flow rate	RLSA004F006/AI1/PV.CV
CW temperature (inlet condenser)	Actual temperature	RLSA004T015/AI1/PV.CV
CW temperature (outlet condenser)	Actual temperature	RLSA004T016/AI1/PV.CV
CW flow rate (feed cooler)	Actual flow rate	RLSA004F007/AI1/PV.CV
CW temperature (inlet feed cooler)	Actual temperature	RLSA004T017/AI1/PV.CV
CW temperature (outlet feed cooler)	Actual temperature	RLSA004T017/AI1/PV.CV

APPENDIX C

Plant wide evaluation report

Plant Wide Performance Report

Ruan du Toit

13-Dec-2005 14:32:44

Table of Contents

Chapter 1. Performance Plots

`The evaluation period`

Begin_time. 26-Nov-2005 17:08:00

End_time. 26-Nov-2005 17:55:00

Figure C.1: Plant wide evaluation report - page 1

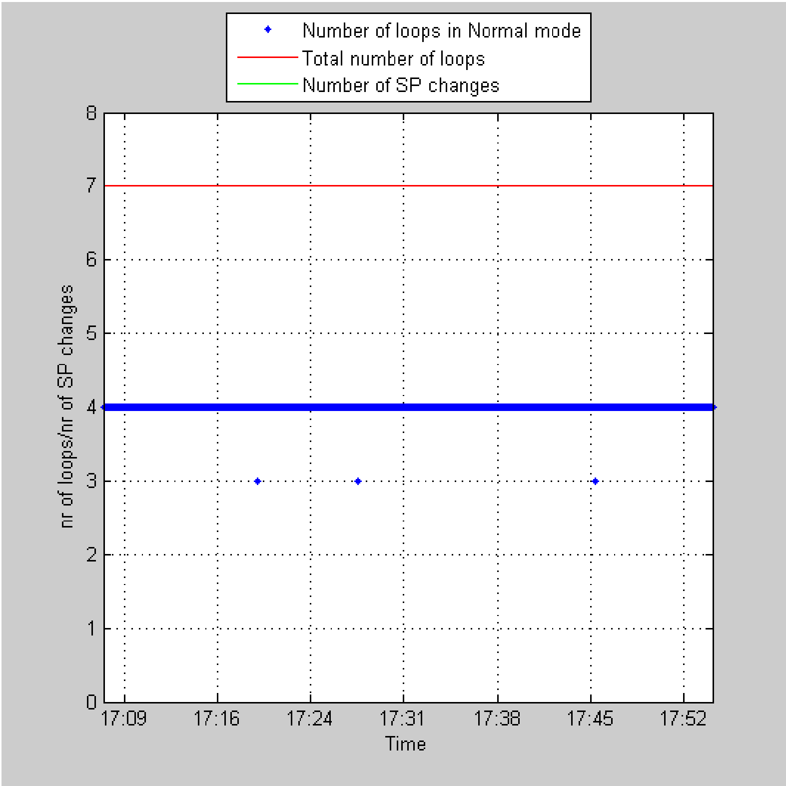
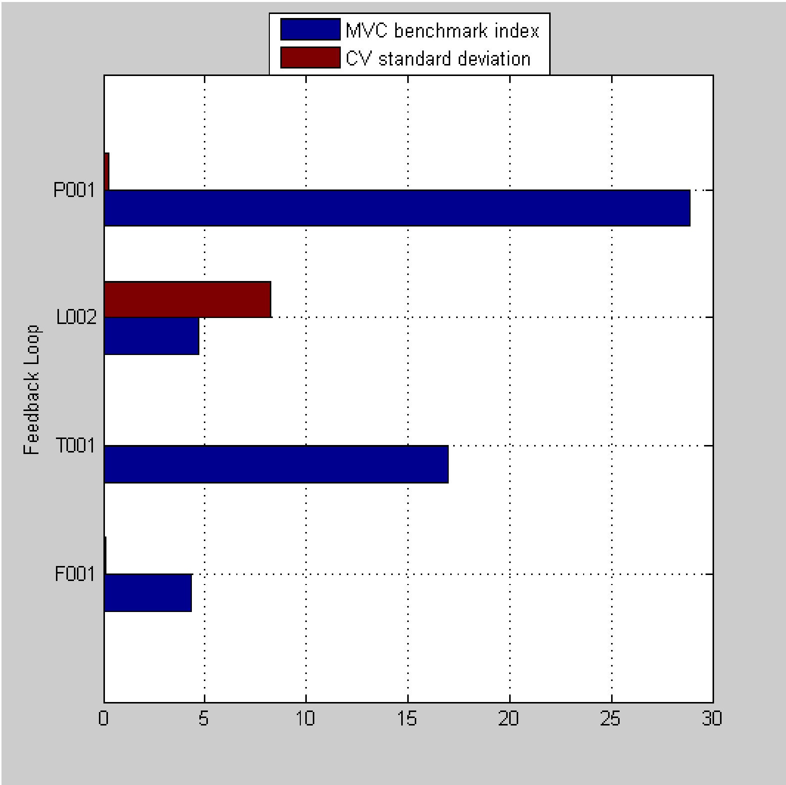


Figure C.2: Plant wide evaluation report - page 2



Chapter 2. Plant wide Index

Plant_Wide_Index. 66.11

Figure C.3: Plant wide evaluation report - page 3

APPENDIX D

Single loop evaluation report

Single Loop Performance Report

Ruan du Toit

13-Dec-2005 17:05:44

Table of Contents

List of Tables

Chapter 1. Single loop evaluation plots

`Loop being evaluated`

Table 1.1. Tag

RLSA004F001/PID1/PV.CV

Figure D.1: Single loop evaluation report - page 1

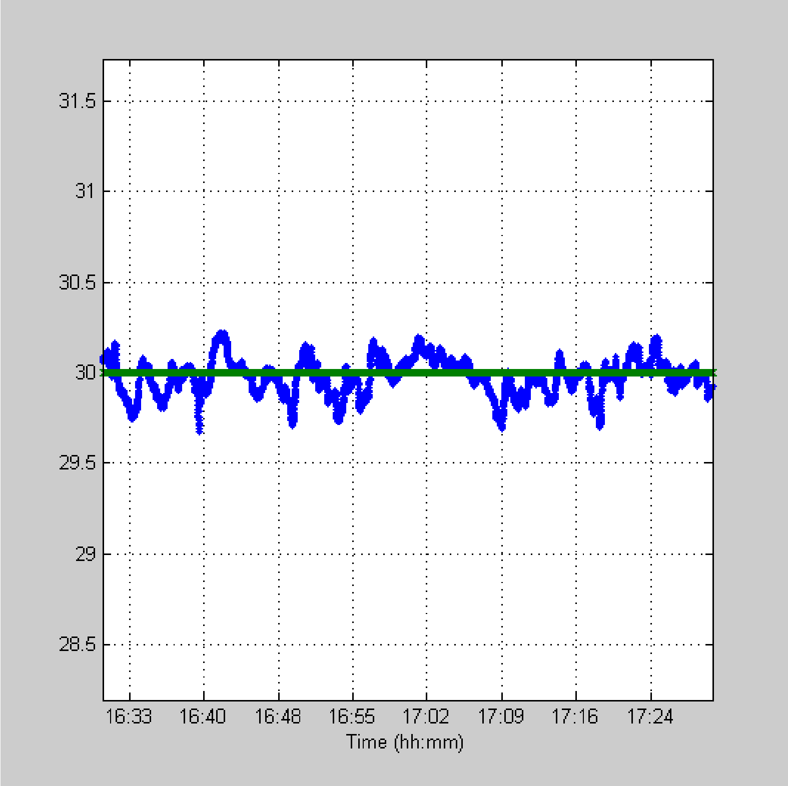


Figure D.2: Single loop evaluation report - page 2

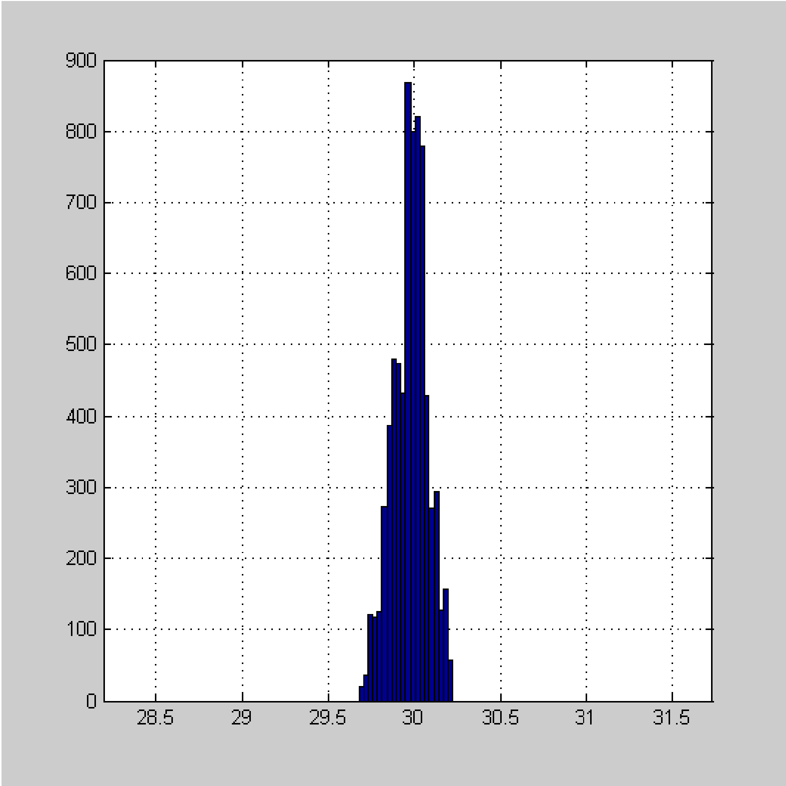
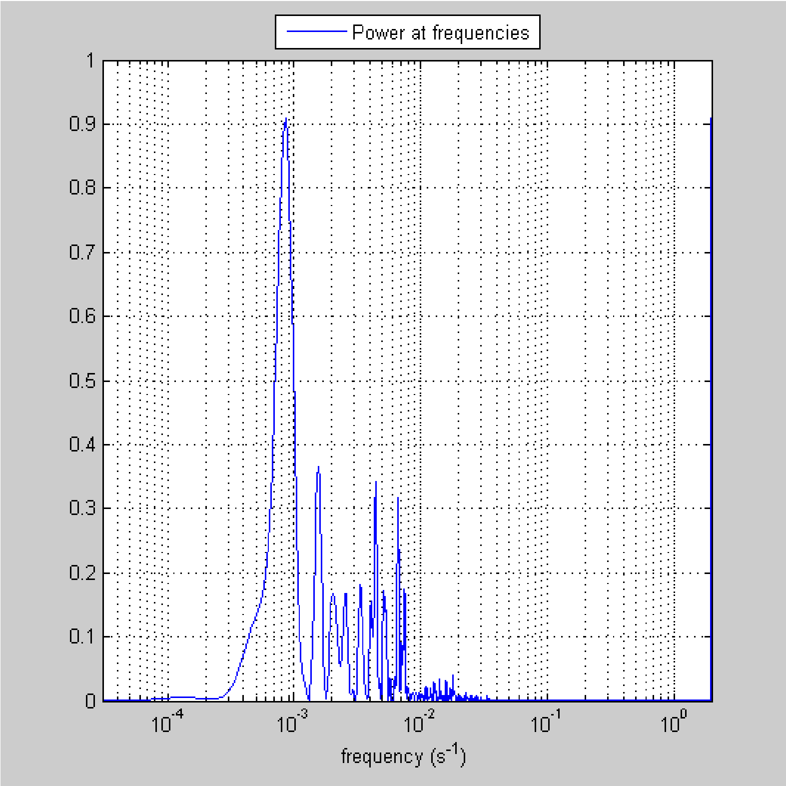


Figure D.3: Single loop evaluation report - page 3



Chapter 2. Performance Metrics

Oscillation. Non-oscillatory

Performance_Index. 4.9644

Figure D.4: Single loop evaluation report - page 4

APPENDIX E

Programming and files

This appendix discusses the files that are used in the performance monitoring structure. All the files discussed are available on the CD that accompanies this dissertation.

- **Column_Tag_Structure.mat** - The MATLAB structure that contains all the tags on the column. This structure is used to identify tags to be logged.
- **struct2xml.m** - The MATLAB M-file used to convert the MATLAB structure to xml format.
- **Column_Tags.xml** - The xml file containing the column tag namespace. This file was created by executing the *struct2xml.m* M-file.
- **Tags_interface.osf** - The OPC Toolbox file that contains all the tags for data logging via the OPC Toolbox. The tags are shown in table B.1.
- ***.olf** - Files with the *.olf* extension is operating data logged with the OPC Toolbox. The filenames are the dates on which the data were logged (for example *data_26_11_05_Run_05.olf*).
- **auto_corr_coeff.m** - Function that computes the auto correlation coefficient for a lag of 1.
- **corr_coeff.m** - Function that computes the cross correlation coefficient for various lags.
- **Data_Dimen.m** - Function that reshapes any two vectors to be of equal dimensions.

- **Data_retrieve.m** - Function that imports the OPC log file ($*.olf$) into the MATLAB current directory.
- **element_removal.m** - Function that removes specified elements from signal vectors.
- **harris.m** - Function that computes the minimum variance index according to Fellner's formula.
- **NAN_removal.m** - Function that removes elements that have the string value NAN.
- **opcread1.m** - Function that imports creates the tag structure.
- **oscil.m** - Function that executes H*ä*gglund's algorithm for oscillation detection.
- **perf_gui.m** - The M-file that launches the single loop performance interface.
- **Plant_wide_GUI.m** - The M-file that launches the plant wide performance interface.
- **plantwide_harris.m** - Function that computes the minimum variance index for a number of loops.
- **Plots_Data.m** - Function that plots the relevant graphs for report generation.
- **PowerSpec.m** - Function that calculates the PSD.
- **PW_report.m** - The report generation file used to generate the plantwide report in html format.
- **refine_time.m** - Function that sets the evaluation period.
- **sigqual.m** - Function that sets the quality strings to a format that can be plotted.
- **Single_loop_report.m** - The report generation file used to generate the single loop performance report in html format.
- **t_auto.m** - Function that calculates the percentage time a loop was on AUTO.