ABSTRACT

Title of dissertation: INFERENTIAL MODEL PREDICTIVE CONTROL USING STATISTICAL TOOLS

Kedar Himanshu Dave, Doctor of Philosophy, 2005

Dissertation directed by: Professor Thomas J. Mc Avoy
Department of Chemical Engineering

With an ever increasing emphasis on reducing costs and improving quality control, the application of advanced process control in the bulk chemical and petrochemical industry is steadily rising. Two major areas of development are model-based control strategies and process sensors. This study deals with the application of multivariate statistical techniques for developing soft-sensors in an inferential model predictive control framework. McAvoy [31] has proposed model predictive statistical process control (MP-SPC), a principal component (PC) score control methodology. MP-SPC was found to be very effective in reducing the variability in the quality variables without using any real-time, on-line quality or disturbance measurements. This work extends McAvoy's [31] formulation to incorporate multiple manipulated variables and demonstrates the controller's performance under different disturbance scenarios and for an additional case study. Moreover, implementation issues critical to the success of the formulations considered such as controller tuning, measurement selection and model identification are also studied. A key feature is the emphasis on confirming the consistency of the cross-correlation between the selected measurements and the quality variable before on-line implementation and that between the scores and the quality variables after on-line implementation.

An analysis of the controller's performance in dealing with disturbances of different

frequencies, sizes and directions, as well as non-stationarities in the disturbance, reveals the robustness of the approach. The penalty on manipulated variable moves is the most effective tuning parameter. A unique scheme, developed in this study, takes advantage of the information contained in historical databases combined with plant testing to generate collinear PC score models. The proposed measurement selection algorithm ranks measurements that have a consistent cross-correlation with the quality variable according to their cross-correlation coefficient and lead time. Higher ranked variables are chosen as long as they make sufficiently large contributions to the PC score model. Several approaches for identifying dynamic score models are proposed. All approaches put greater emphasis on short term predictions. Two approaches utilize the statistics associated with the PC score models. The Hotelling's T^2 statistic and the Q-residual information may be used to remove outliers during pre-processing or may be incorporated as sample weights.

The process dynamics and controller performance results presented in this study are simulations based on well-known, industrially benchmarked, test-bed models: the Tennessee Eastman challenge process and the azeotropic distillation tower of the Vinyl Acetate monomer process.

INFERENTIAL MODEL PREDICTIVE CONTROL USING STATISTICAL TOOLS

by

Kedar Himanshu Dave

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Commmittee:

Professor Thomas J. Mc Avoy, Chair/Advisor
Associate Professor Raymond A. Adomaitis
Associate Professor Evanghelos Zafiriou
Professor Timothy A. Barbari
Professor William S. Levine

ACKNOWLEDGMENTS

I fear that my whole-hearted acknowledgement of my thesis advisor might sound cliched, but I do want to take this opportunity to express my gratitude for his help nevertheless. Prof. Mc Avoy has seen me through a lot of ups and downs and I hope I have lived, and continue to live, up to his expectations. He has been very patient and understanding even during times of personal tragedy. His straight-forward attitude and emphasis on practical applicability have been his hallmark. I was very comfortable with his managing style and I can confidently say that all his former students have a similar sentiment.

Dr. Evanghelos Zafiriou has also guided me through the last few years and he has helped me make the right decisions. I am also grateful to Prof. Barbari (the former chair of the department), Dr. Adomaitis (the present chair) and Dr. Sheryl Ehrman (the graduate advisor) for having helped me see through troubled times. Many thanks to Prof. William S. Levine for serving on my committee.

I would like to acknowledge the following friends for their companionship: Gokul & Anu Kalur, Krishna Kaipa, Len Jones, Nilesh & Ashwini Bhat, Priya Ranjan, Pushkar Godkar, Rajiv & Tejas Gandhi and Ramesh & Ramya Gopalan. Thanks are also due to my lab mates Rong Chen and Wendy Abi Assali.

Last, but not the least, I am thankful to my parents and my brothers for their continuing long-distance support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

This chapter gives a broad overview of current process control and process measurement approaches. These approaches provide the basic premise for this study. Literature related to the proposed approach to process control is briefly reviewed. Research philosophy, research goals and an outline of the thesis are also given.

1.1 Motivation

Control theory has applications in many diverse fields. Eventhough each discipline has it's own subtleties, the underlying principles are fundamentally similar. When comparing the challenges faced in the control of bulk chemical processes against those faced when controlling electrical or mechanical systems, chemical process control deals with predominantly regulator problems (as opposed to servo problems), relatively slower dynamics, feedback effects and strong interactions in a multivariate setting, lack of adequate measurements and the need for simplicity in implementation.

In recent times, the chemical process industries have witnessed an ever increasing push to produce higher quality products, to reduce product rejection rates (by decreasing off-specification products), to minimize costs, to improve energy efficiency, to increase production rates, and to satisfy increasingly stringent safety and environmental regulations. Improved process control strategies and/or improved process measurement are means of meeting these higher standards. Since the primary goal of chemical process control is load disturbance rejection, it is a good idea to gain an understanding of the nature of

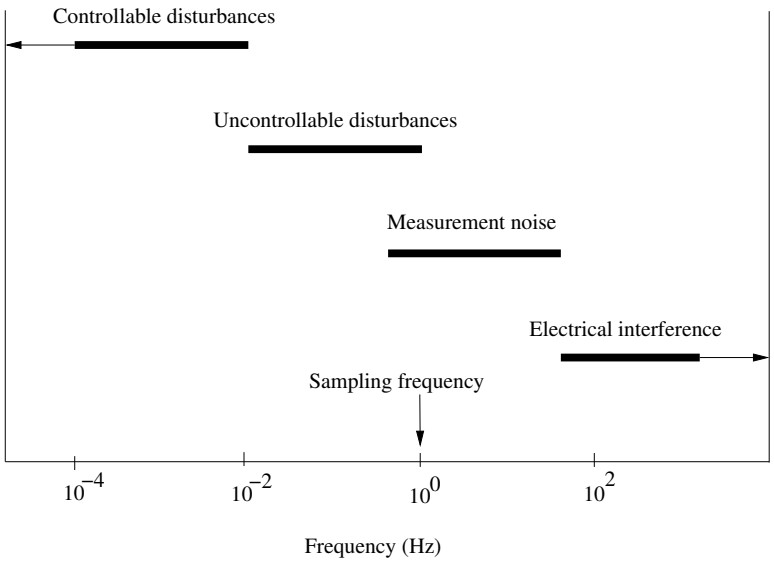

Figure 1.1: Nature of upsets encountered in bulk chemical processes. Figure adapted from Marlin [30].

fluctuations typically encountered in chemical process plants. Process upsets usually have a significant random component in them and so may be treated as random processes. A random process where all of its statistical properties (i.e. mean, correlation, variance, etc.) do not vary with time is referred to as a *Stationary process* while a process whose statistical properties do change is referred to as *Non-stationary*. In this study however, these definitions are not applied in the strictest sense. In this study, a random fluctuation whose mean does not change with time is referred to as a Stationary fluctuation while a fluctuation whose mean does change (i.e. a step-like upset) is referred to as Non-stationary. Although the upsets encountered in continuous process plants have a wide range of frequencies, as shown in Figure 1.1, process control systems target only those fluctuations that have frequencies in the range 10^{-4} to 10^{-2} cycles/sec (i.e. Hertz).

A quantitative means of justifying improvements in process control is by estimating it's direct economic impact. Consider Figure 1.2. Mean values ($\bar{X}_B$) and variances (σ_B^2) for the key economic controlled variables for the base operation are determined from the

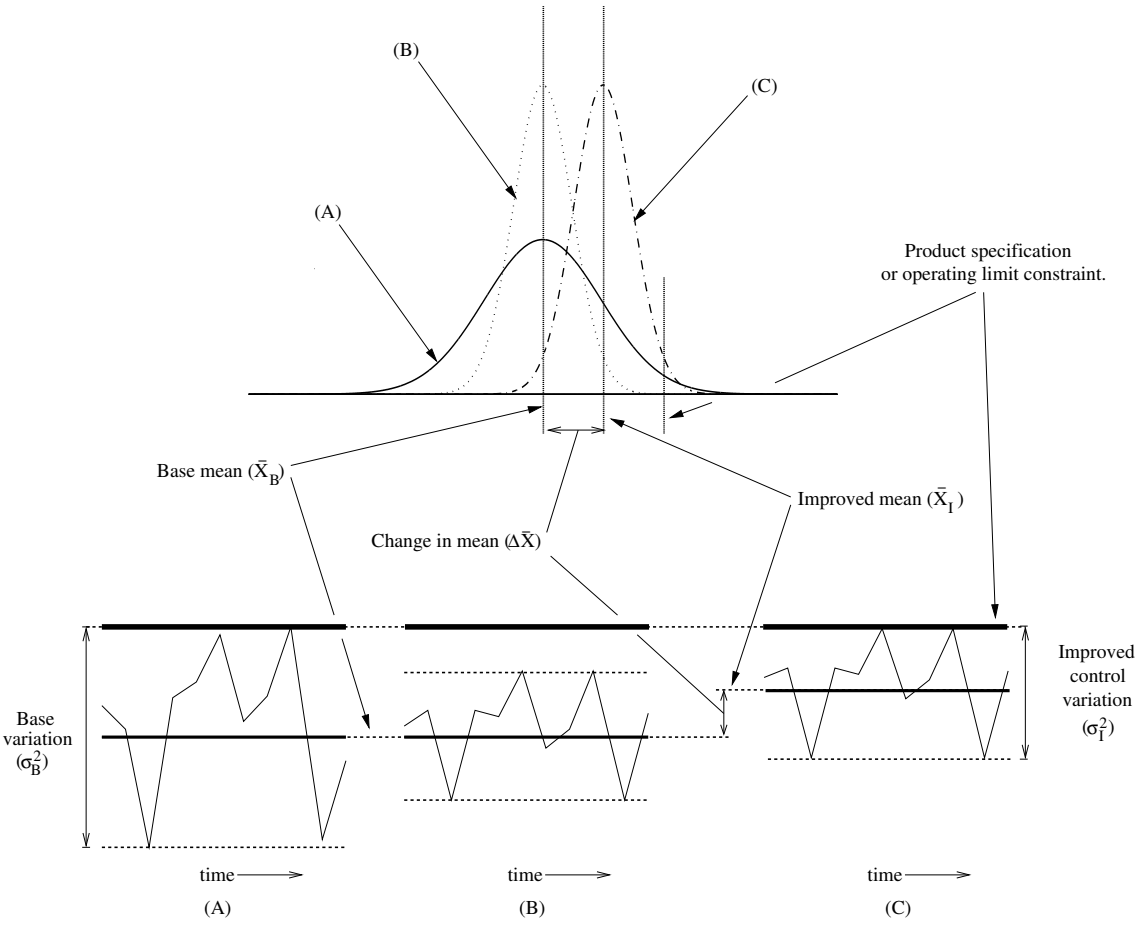

Figure 1.2: Quantifying the incentives for improvements in process control: (A) Base control (poor); (B) Improved control; and (C) Change in mean (moving closer to the constraint) as a result of improved control. Figure adapted from Muske [36] and Jones [22].

Table 1.1: A rudimentary classification of chemical process plant upsets.

	Disturbance	Noise
APC's perspective	deterministic	stochastic (random)
SPC's perspective	special cause	common cause
frequency content	low frequeny	high frequency
action taken	need to compensate for	dampened or filtered

base operating data. Improvements in process control are expected to reduce the variation by $\Delta\sigma^2$ $(= \sigma_B^2 - \sigma_I^2)$ in the controlled variables. Here σ_I^2 is the variance after improvements in control. As a result of this reduction the mean operating value can be shifted to $\bar{X}_I$, i.e. closer to the product specification or operating limit constraint without increasing the frequency of violation. For example if the deviations are assumed to be normally distributed, and if 2% one-sided constraint violations are tolerable, then the change in mean $(\Delta\bar{X})$ as a result of reducing the variance by $\Delta\sigma^2$ is given by:

$$\Delta\bar{X} = \sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{0.5 - 0.02}{0.5}\right)\Delta\sigma^2 = 2.0537\Delta\sigma^2 \tag{1.1}$$

The corresponding factor by which the mean can be shifted when 5% one-sided violations are tolerable is 1.6449 (Jones [22]). Operating at the new mean value provides the economic incentive. Quantification of the economic benefit is performed by using some form of a process model to determine the steady-state material and energy balance changes resulting from the improved control operation (Muske [36]). A brief overview of current process control strategies is discussed next while issues pertaining to process measurements are discussed in section 1.4.

Depending upon the way in which upsets (Table 1.1) are dealt with, there are two basic approaches to process control: automatic process control (APC) and statistical process control (SPC). It is possible for APC and SPC to provide complementary functions and they can be applied on the same process to improve the overall performance. This

study attempts to develop such an integrated approach. Before discussing the details of this new approach, a brief overview of APC and SPC is given next.

1.2 Automatic process control (APC)

Automatic process control is based upon the principles of feedback and feedforward. Here, deviations from the desired operating setpoint(s) are identified and corrective action is taken by adjusting the manipulated variable(s). However, APC does not attempt to eliminate the cause of poor operation. Manipulated variables are adjusted simply to compensate for the effects of the disturbance and thereby maintain the controlled variable at it's desired value. Since the sources of disturbances aren't affected, APC leaves the process susceptible to future disturbances from the same source. Plantwide control is a methodology to design APC systems for entire plants. Such control systems are usually implemented in an hierarchical (tiered) fashion.

1.2.1 Base control systems

At the lowest level of the control system hierarchy are the single-input single-output (SISO) proportional (P), proportional-integral (PI) or proportional-intergral-derivative (PID) loops providing fast control action (i.e. in the order of seconds). These controllers use the measurements transmitted from the field sensors as inputs and calculate the control moves. The controller outputs are sent to the actuators of the final control elements (which are usually control valves). Some loops may be implemented in a cascade fashion. In this scheme, the output of the upper level primary (i.e. master) controller is the set-point for the lower level secondary (i.e. slave) controller. The most basic arrangement consisting of mostly SISO loops and a few cascade loops is known as the *base control system*. This

arrangement stabilizes the plant and provides basic regulatory control for disturbance rejection but quite often this is sufficient to satisfy the economic and safety objectives and so is not necessarily optimal. If on-line analyzers are used for feedback then quality issues can also be addressed. Base control systems can be designed using a variety of methods. Luyben et al [29] have suggested a method based on empirical approaches (rules of thumb), engineering judgement and process knowledge. The method put forward by Chen and McAvoy [8] relies more on process models and optimization. The design of base control systems is however not within the scope of this study. Instead, this study assumes that a base control system has already been designed and is operating satisfactorily.

1.2.2 Model predictive control (MPC)

Very often the base control system is not enough to meet the control requirements. In earlier times, these shortcomings were overcome by augmenting the base control system with various selectors, overrides, decouplers, time-delay compensators, etc. Nowadays, optimizing, multivariable and model-based control systems provide control action on the order of minutes and are used as upper layers. Model predictive control (MPC) refers to a class of multivariable algorithms in which explicit models are used to compute a sequence of present and future manipulated variable adjustments. Amongst the several approaches to MPC, Dynamic matrix control (DMC) is one of the earliest and is still popular (Qin and Badgwell [41, 42]). It derives it's name owing to the use of models in the form of a

dynamic matrix, $\mathbf{A_D}$, given by:

$$\mathbf{A_D} \equiv \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1s} & \cdots & \mathbf{A}_{1S} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2s} & \cdots & \mathbf{A}_{2S} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rs} & \cdots & \mathbf{A}_{rn_u} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ \mathbf{A}_{n_y1} & \mathbf{A}_{n_y2} & \cdots & \mathbf{A}_{n_ys} & \cdots & \mathbf{A}_{n_yn_u} \end{bmatrix}_{P\cdot n_y \times M\cdot n_u} \tag{1.2}$$

where each $\mathbf{A}_{rs}$ has dimensions $P \times M$ and is formed from the step response coefficients ($\mathbf{s_{rs,i}}$) of the subprocess relating the rth process measurement to the sth manipulated variable. Hence for a single variable case (i.e. for $n_y = 1$ and $n_u = 1$),

$$\mathbf{A_D} = \begin{bmatrix} s_1 & 0 & 0 & \cdots & 0 \\ s_2 & s_1 & 0 & \cdots & 0 \\ s_3 & s_2 & s_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ s_P & s_{P-1} & s_{P-2} & \cdots & s_{P-M+1} \end{bmatrix}_{P\times M} \tag{1.3}$$

Here, M is the control horizon and P is the prediction horizon. At each sampling instant, the future control moves are calculated by solving:

$$\min_{\mathbf{\Delta u}} \; [\mathbf{E_p} - \mathbf{A_D}\mathbf{\Delta u}]^T\mathbf{\Gamma}^T\mathbf{\Gamma}[\mathbf{E_p} - \mathbf{A_D}\mathbf{\Delta u}] + \mathbf{\Delta u}^T\mathbf{\Lambda}^T\mathbf{\Lambda u} \tag{1.4}$$

This formulation is the traditional unconstrained dynamic matrix control (DMC) formulation where the objective function to be minimized penalizes squared input ($\mathbf{\Delta u}$) and controlled variable (i.e. state or measurement) deviations from the origin ($\mathbf{E_p} - \mathbf{A_D}\mathbf{\Delta u}$). It also includes separate diagonal controlled variable and input weight matrices, $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ respectively, to allow for tuning trade-offs. The matrix $\mathbf{\Gamma}^T\mathbf{\Gamma}$ has γ_r^2 ($r =$

$1, 2, \ldots, R$) as the leading diagonal elements of the rth diagonal matrix block, i.e.:

$$\mathbf{\Gamma}^T\mathbf{\Gamma} = \begin{bmatrix} \gamma_1^2\mathbf{I}_{P\times P} & 0 & \cdots & 0 \\ 0 & \gamma_2^2\mathbf{I}_{P\times P} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_R^2\mathbf{I}_{P\times P} \end{bmatrix}_{P\cdot R\times P\cdot R} \tag{1.5}$$

Moreover, the matrix of move suppression coefficients, $\mathbf{\Lambda}^T\mathbf{\Lambda}$, has λ_s^2 ($s = 1, 2, \ldots, S$) as the leading diagonal elements of the sth diagonal matrix block, i.e.:

$$\mathbf{\Lambda}^T\mathbf{\Lambda} = \begin{bmatrix} \lambda_1^2\mathbf{I}_{M\times M} & 0 & \cdots & 0 \\ 0 & \lambda_2^2\mathbf{I}_{M\times M} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_S^2\mathbf{I}_{M\times M} \end{bmatrix}_{M\cdot S\times M\cdot S} \tag{1.6}$$

It should be noted that $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ are usually held constant over the prediction and control horizons respectively. The closed form solution to Equation 1.4 is the unconstrained DMC control law:

$$\Delta\mathbf{u} = (\mathbf{A_D}^T\Gamma^T\Gamma\mathbf{A_D} + \Lambda^T\Lambda)^{-1}\mathbf{A_D}^T\Gamma^T\Gamma\mathbf{E_p} \tag{1.7}$$

Out of the control moves thus generated, only the ones corresponding to the current time step are implemented. The calculations are repeated at the next time step (i.e. at $k+1$) using a new value of $\mathbf{E_p}$. This approach, termed as *Receding horizon control* or *Moving horizon control* is depicted in Figure 1.3.

Inequality and equality constraints can also be included in the objective function. A closed form solution for the control law similar to Equation 1.7 can still be obtained when only equality constraints are present. This may be achieved by incorporating appropriate Lagrange multipliers. However when inequality constraints are also present, the control law is no longer linear and a quadratic program calculation becomes necessary.

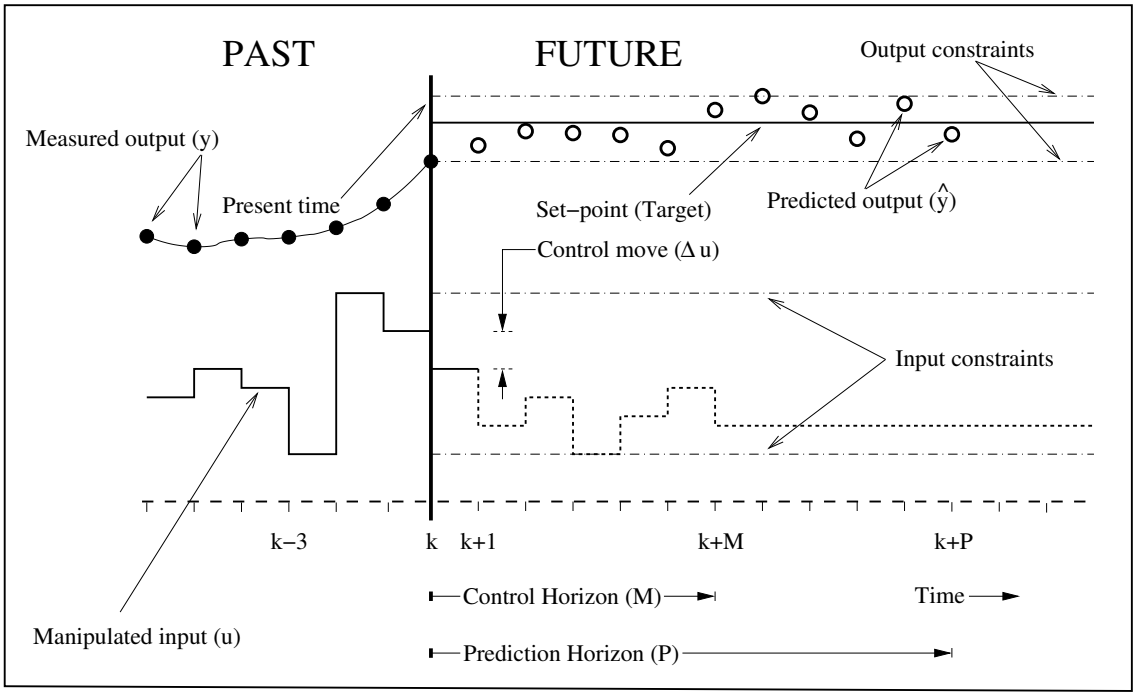

Figure 1.3: Schematic illustrating Receding horizon control.

Models are central to the MPC approach and so it is worthwhile discussing the various model forms used in MPC. Models used in linear MPC may be classified into parametric and non-parametric models. The *Finite Step Response* (FSR) model is the simplest form of a non-parametric model. The FSR model of a SISO process is the vector of step response coefficients:

$$\mathbf{s} = [s_1 \; s_2 \; s_3 \cdots s_i \cdots s_{n_s}]^T \tag{1.8}$$

where the settling time (i.e. model length denoted by n_s) is long enough so that the coeffcient values are relatively constant (i.e. the process is close to a new steady state). The *Finite Impulse Response* (FIR) model of a SISO process is the vector of impulse response coefficients:

$$\mathbf{h} = [h_1 \; h_2 \; h_3 \cdots h_i \cdots h_{n_s}]^T \tag{1.9}$$

As shown in Figure 1.4 the FIR model can be obtained from the FSR model and vice

versa:

$$\left.\begin{aligned} h_i &= s_i - s_{i-1} \\ s_i &= \textstyle\sum_{j=1}^{i} h_j \end{aligned}\right\} \tag{1.10}$$

It should be noted that the abscissas in Figure 1.4 are discrete time steps. The most commonly used parametric models are variants of the basic auto-regressive (ARX) models. State-space models are another form of representing linear difference relationships. The concept of state, the minimum amount of information required to describe the syetem, is the key feature. Linear, time-invariant, discrete, state-space models are given by:

$$\left.\begin{aligned} \mathbf{x}(k+1) &= \Phi\mathbf{x}(k) + \Gamma_u\mathbf{u}(k) + \Gamma_d d(k) + \Gamma_w w(k) \\ \mathbf{y}(k) &= \bar{y}(k) + z(k) \\ &= Cx(k) + D_u\mathbf{u}(k) + D_d d(k) + D_w w(k) + z(k) \end{aligned}\right\} \tag{1.11}$$

Here, $\mathbf{x}(k)$, $\mathbf{y}(k)$, $\mathbf{u}(k)$, $\mathbf{d}(k)$ and $\mathbf{w}(k)$ are the state, measurement, manipulated input, disturbance and noise vectors respectively at sampling time k (see Franklin et al [15] for further details). Non-linear models in the form of Artificial Neural Networks (ANNs), non-linear state space or Volterra Laguerre models can also be incorporated. But in such situations a non-linear program has to be solved. If the objective function involves logical constraints in the form of binary variables, the approach is referred to as Hybrid MPC and a mixed-integer program calculation becomes necessary (see Morari and Lee [34]; Morari et al [35]).

1.3 Statistical process control (SPC)

Statistical process control (SPC) involves monitoring the performance of a process over time to verify that it is remaining in a state of statistical control (Marlin [30] and Levinson [26]). Such a state of control is said to exist if certain process or product variables (usually small in number) remain close to their desired values and the only

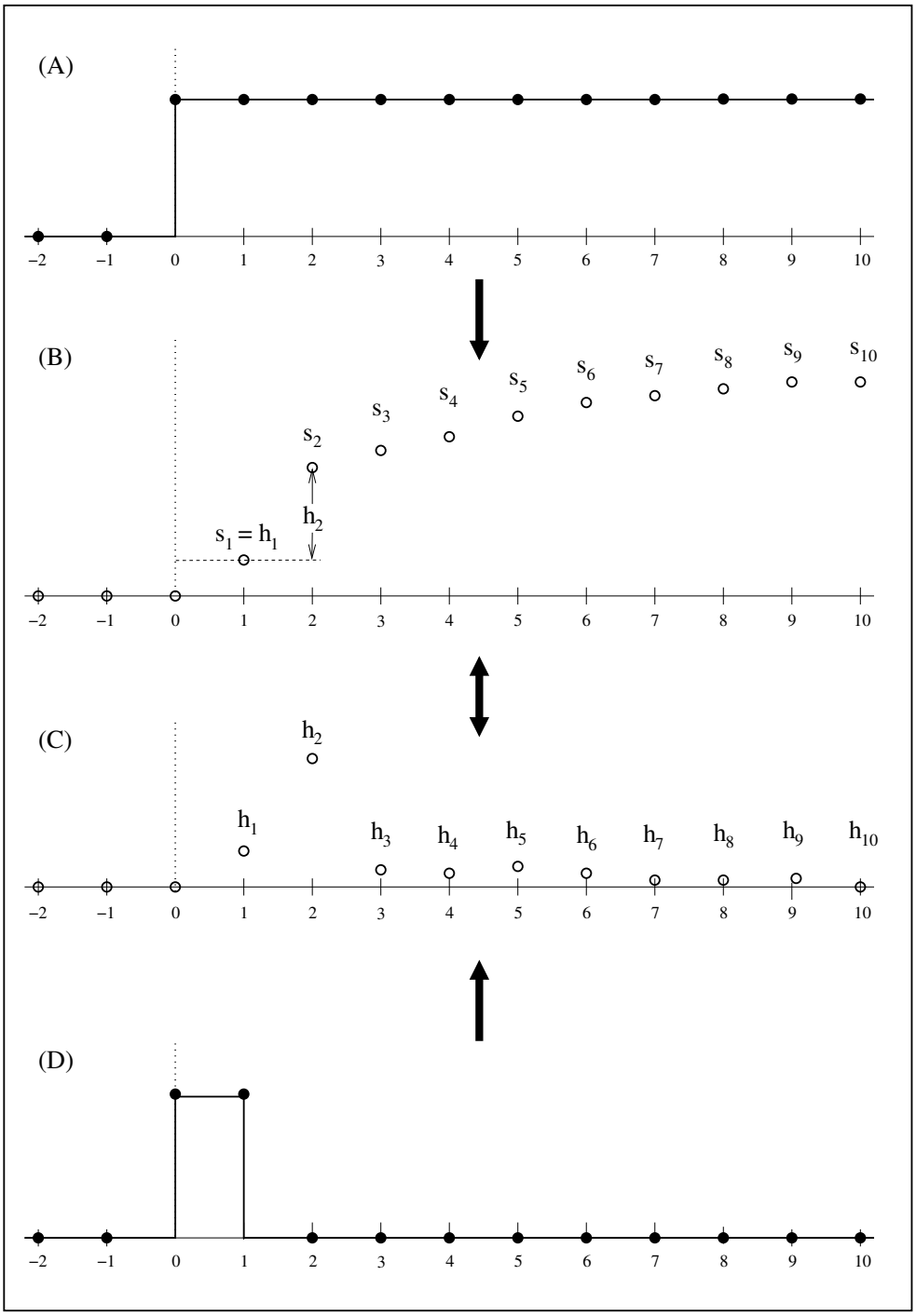

Figure 1.4: Non-parametric discrete time models: (A) Step input, (B) Finite step response (FSR) model, (C) Finite impulse response (FIR) model and (D) Impulse input.

source of variation is *common-cause* variation, that is, variation which affects the process all the time and is essentially unavoidable within the current process. Abnormal process conditions are considered as events having *special* or *assignable* causes. Their occurances are identified and the source (i.e. root cause) of such disturbances are eliminated.

The ultimate goal of both APC and SPC is to improve products and the processes used to make them. However, in contrast to APC, SPC achieves this goal by making the process less susceptible to future upsets. However, SPC alone cannot adequately control most process operations. The benefits are also long-term because although SPC uses experience and empirical models derived from real-time measurements, "control" is through infrequent manipulated variable movements and is usually not carried out in real-time.

The techniques used in SPC can be loosely classified into (1) Analysis and (2) Regression. Analysis involves drawing useful conclusions from a single block of data. Regression involves building quantitative relationships between more than one block of data. Before discussing further details about SPC, a few tools commonly used in statistics, time series analysis and signal processing are defined next.

1.3.1 Statistical hypothesis testing

Certain control problems can be posed as statistical hypotheses. Two issues important in hypothesis testing are sample classification and the level of sigificance parameter.

Matched and Independent samples

Samples may be classified into:

1. Matched samples: Matched samples (also called paired samples) can arise in the

following situations:

(a) Two samples in which the members are clearly paired, or are matched explicitly by the researcher. For example, IQ measurements on pairs of identical twins.

(b) Those samples in which the same attribute or variable, is measured twice on each subject, under different circumstances. This is commonly called repeated measurements. For example, the times of a group of athletes for 100 m. before and after special training or the milk yields of cows before and after being fed a particular diet.

2. Independent samples: Independent samples are those samples selected from the same population or different populations, which have no effect on one another. That is, no correlation exists between the samples.

The Wilcoxon hypothesis test is used in this study to determine the stationarity of the process under control. In the MATLAB environment, the `signrank` function in the MATLAB Statistics Toolbox can be used to carry out the Wilcoxon sign rank test for matched (or paired) samples while the tt ranksum function can be used to carry out the Wilcoxon rank sum test for independent samples.

Level of significance

Statistical hypothesis tests require the specification of the level of significance (α) for the study. Hypothesis tests deal with two types of errors:

1. Type I errors: These errors are the ones when TRUE null hypotheses are erroneously rejected.

2. Type II errors: These errors are the ones when FALSE null hypotheses are erroneously accepted.

The level of significance (α) is the probability of committing a Type I error. For example, if α is decreased, the chance of a Type I error is decreased. However, at the same time the chance of a Type II error is increased. Hence, there is a trade-off. $\alpha = 0.05$ is the most commonly used value of α (as reported in the literature) and also its default value in most statistical software. In other words, there is a 5% chance of an error in the conclusions drawn when the default value of α is used. It should also be noted that the quantity and/or quality of data should not influence the choice of α. If the data is allowed to dictate the choice of α, bias will enter the study.

1.3.2 Data pre-processing

Data pre-processing, or data transformation, prior to applying any statistical technique is often critical to success. There are four aspects to data pre-processiing:

- Scaling (also known as weighting or normalizing): This the most routinely used form of pre-processing. The aim of scaling is to ensure that every variable has an opportunity to participate in the statistical modeling and analysis irrespective of it's unit of measurement and range of variation. The various types of scaling are shown in Figure 1.5. The process of mean-centering involves calculating the average value for each variable and then substracting this average from every sample for that variable. Variance scaling involves calculating the variance for each variable and then dividing every sample for that variable by it's variance. Auto-scaling is mean-centering followed by variance scaling. Process measurement data is often auto-scaled.

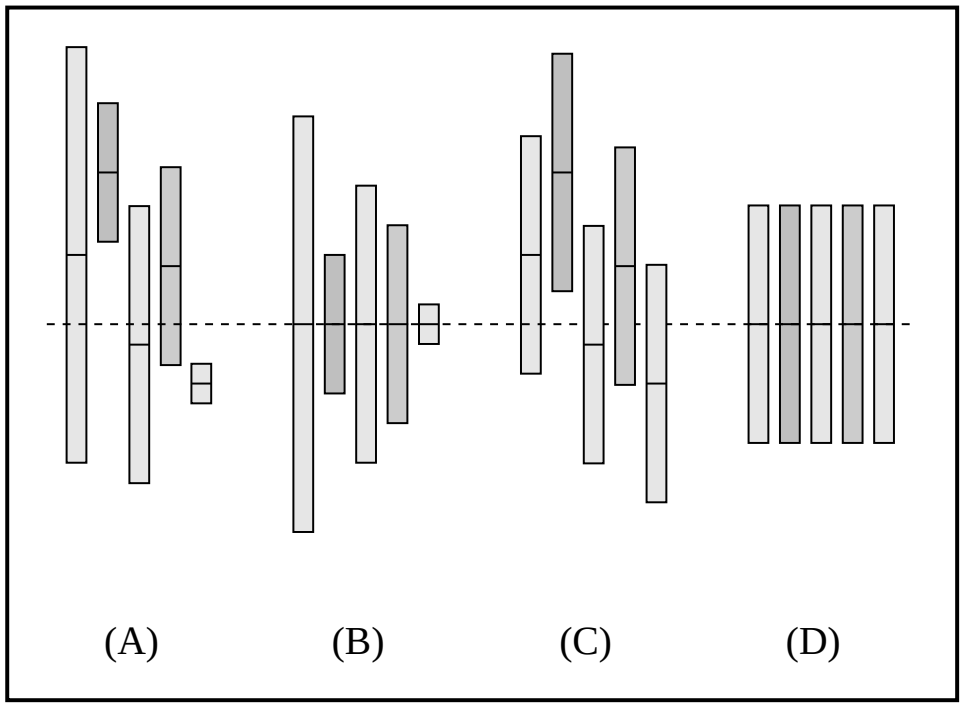

Figure 1.5: Scaling of a 5-variable data-set (Geladi and Kowalski [16]). The data for each variable are represented by a variance bar and it's center: (A) Raw data (without any scaling), (B) Mean-centered data, (C) Variance-scaled data and (D) Auto-scaled data.

- Linearization: The linear statistical methods discussed in this study assume a linear relationship between the variables. It is advantageous to appropriately deal with the known non-linearities during pre-processing. For example, reaction rates are exponential functions of temperature. In this case, the natural logarithm of the temperature should be used in the analysis. Similarly, for turbulent flow in a pipe, the flow rate isn't a linear function of the pressure drop ($F \propto \Delta P^{4/7}$).

- Differentiation: This study deals with scalar quality variables. However, quite often quality is described using a vector such as a molecular weight distribution, particle size distribution, etc. In such cases the data set consists of several chromatograms or spectra. Taking the first derivatives of spectra has the advantage of removing the effects of baseline offsets on the results. Taking the second derivatives of spectra has the advantage of resolving "peaks" and "shoulders" and thereby revealing the underlying addition of spectra. Filters such as the Savitzky-Golay [47] filter are also useful in improving the resolution.

- Outlier removal: Owing to an excessive amount of noise or errors present in them, some samples have a detrimental effect on the results. At times some variables have an excessive amount of leverage on the results. Such samples or variables should be removed during the pre-processing step.

1.3.3 Classical SPC and univariate statistical tools

The foundations of SPC have been laid by Shewart [48]. The Shewart (see Figure 1.6), the cumulative sum (CUMSUM) [19] and the exponentially weighted moving average (EWMA) [56] charts are widely used SPC tools in industry. Proceses improvements can also be attained by using univariate and multivariate statistical tools to carry out data

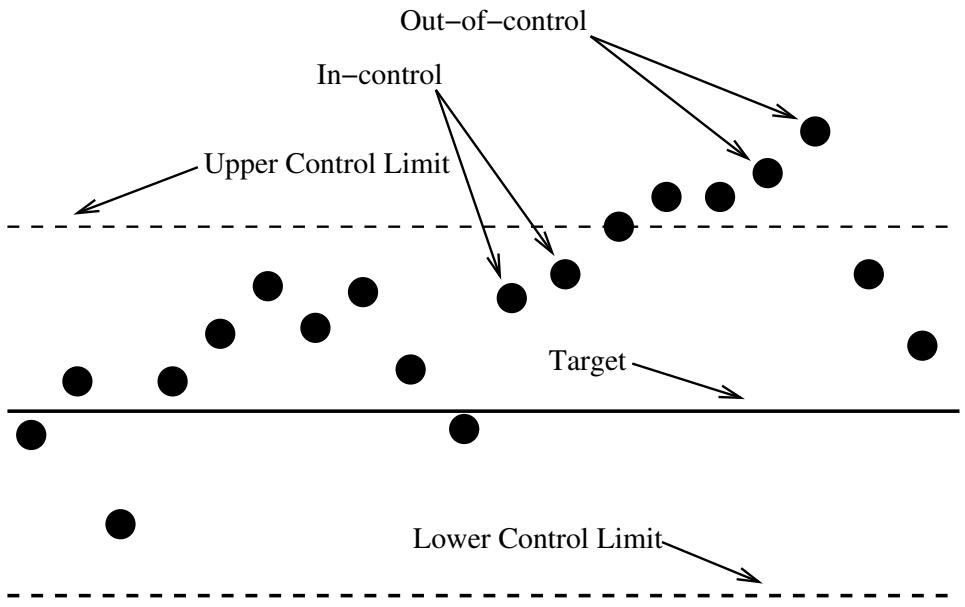

Figure 1.6: A typical Shewart chart.

analysis or for building regression models. These are discussed next.

Consider a data-set $\mathbf{X}$ of process measurements. $\mathbf{X}$ is of dimension $n_s \times n_y$ with $\mathbf{x}(\mathbf{t})(= x_1(t), x_2(t), ...x_n(t))$ as the vector of measurement variables (or simply measurements) collected at sample time t. Typically the measurements contain noise and many of them are linearly dependant, i.e. there is redundancy in the data. As a result, n_s and n_y are very large, i.e. this data set is of a very high dimensionality. In order to reduce the dimensionality and identify the underlying correlation, several statistical techniques may be employed.

Correlation Analysis

In order to determine the redundancy in two sets of time-series data, it is necessary to determine whether two variables are correlated (i.e. parallel or collinear). There are several measures of redundancy. Consider two process measurements $\mathbf{x}$ and $\mathbf{y}$, with means

$\bar{x}$ and $\bar{y}$ respectively, each a vector of length n_s. Covariance is defined as:

$$cov(x,y) = \frac{\sum_{i=1}^{n_s}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{1.12}$$

Correlation is usually expressed using the correlation coefficient ($r_{x,y}$) and is defined as:

$$r_{x,y} = \frac{\sum_{i=1}^{n_s}(x_i - \bar{x})(y_i - \bar{y})}{\left(\sqrt{\sum_{i=1}^{n_s}(x_i - \bar{x})^2}\right)\left(\sqrt{\sum_{i=1}^{n_s}(y_i - \bar{y})^2}\right)} \tag{1.13}$$

when two different measurements are being compared, this quantity is known as the *cross-correlation coefficient* while when two time periods of the same measurement are being compared, the term used is *auto-correlation coefficient*. $r_{\mathbf{x},\mathbf{y}}$ values range from -1 to +1. The extreme values can be interpreted as:

$$r_{x,y} = \begin{cases} 1 & \mathbf{x} \text{ and } \mathbf{y} \text{ are totally positively correlated} \\ 0 & \mathbf{x} \text{ and } \mathbf{y} \text{ are completely un-correlated i.e. independant} \\ -1 & \mathbf{x} \text{ and } \mathbf{y} \text{ are totally negatively correlated} \end{cases} \tag{1.14}$$

Internediate values of $r_{x,y}$ reflect lesser degrees of correlation. It should be noted that correlation does not imply a causal relationship.

Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR), also known as Ordinary Least Squares (OLS), is an Inverse Least Squares (ILS) method. It calculates the pseudoinverse of $\mathbf{X}$ as:

$$\mathbf{X}^+ = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{1.15}$$

Limitations of univariate techniques

Univariate techniques are not capable of dealing with collinear variables. Moreover, univariate techniques only perform some sort of noise averaging but do not remove noise in the truest sense.

1.3.4 Multivariate statistical tools

Multivariate statistical techniques are becoming increasing popular in many diverse fields where they are variously referred to as Econometrics (in Economics), Biometrics (in Biology) or Chemometrics (in chemistry, particularly analytical chemistry). These techniques are used to perform a myriad of different tasks such as exploratory data analysis, pattern recognition, sample classification, discriminant analysis, data mining, bioinformatics, fault detection, etc. Although there may be slight differences in the nomenclature used, the underlying fundamental principles are the same. Owing to this similarity, research in this area is often cross-disciplinary.

Extending classical SPC, which is traditionally univariate in nature, to multivariate cases, numerous researchers (Kresta et al [24], Wise and Gallagher [53]) have recently discussed applications in chemical process analysis and control. Several multivariate statistical techniques such as Principal Component Analysis (PCA), Principal Component Regression (PCR), Partial Least Squares (PLS) and others may be employed.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA), also known as the Karhunen-Loeve (KT) transform, was originally developed by Pearson [38]. It involves a matrix decomposition that gives rise to a new set of variables, known as the *Principal Components* (PCs), by transforming a given input matrix into two matrices. Another way of looking at this mathematical procedure is that PCA reveals the hidden "real" variables and so the PCs

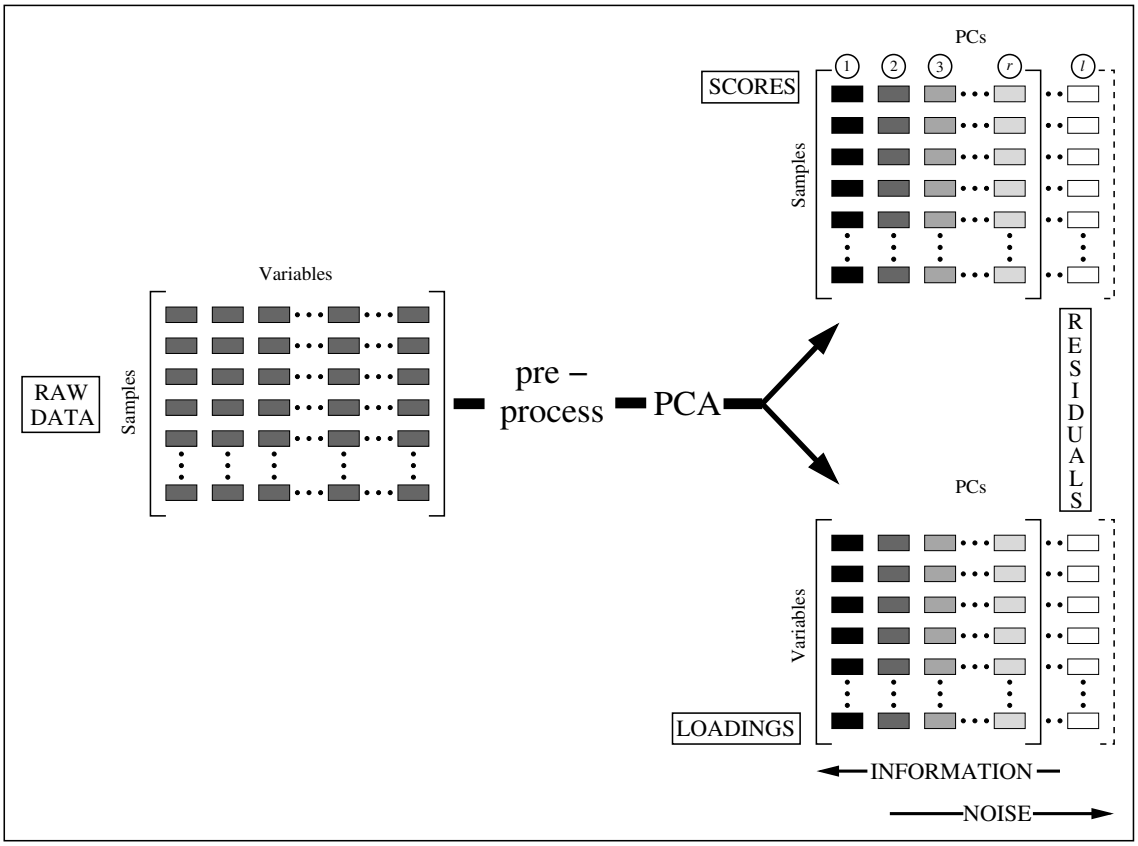

Figure 1.7: Schematic illustrating PCA. A darker intensity is used to depict a higher information content.

are often referred to as *Latent Variables* (LVs). Symbolically,

$$\left.\begin{aligned} \mathbf{X} &\equiv \theta_1\mathbf{p}_1^T + \theta_2\mathbf{p}_2^T + \ldots + \theta_r\mathbf{p}_r^T + \ldots + \theta_l\mathbf{p}_l^T \\ &= \textstyle\sum_{i=1}^{r}\theta_i\mathbf{p}_i^T + \sum_{i=r+1}^{l}\theta_i\mathbf{p}_i^T \\ &= \textstyle\sum_{i=1}^{r}(-\theta_i)(-\mathbf{p}_i^T) + \sum_{i=r+1}^{l}(-\theta_i)(-\mathbf{p}_i^T) \\ &= \boldsymbol{\Theta}\mathbf{P}^T + \mathbf{E} \\ &= \hat{\mathbf{X}} + \mathbf{E} \end{aligned}\right\} \tag{1.16}$$

The mathematical features of PCA, as depicted in Figure 1.7, are:

- The $\mathbf{p}_i$ vectors are known as the *PC factors* or *loadings*. They are orthonormal (i.e. $\mathbf{p}_i^T\mathbf{p}_j = 0$ for $i \neq j$, $\mathbf{p}_i^T\mathbf{p}_j = 1$ for $i = j$) and provide the direction of the PCs. For each PC, there are as many loadings as there are variables in the input matrix (i.e.

the $\mathbf{P}$ matrix is $n_y \times r$). The $\mathbf{p}_i$ vectors are the eigenvectors of the covariance matrix, i.e. for each $\mathbf{p}_i$,

$$\text{cov}(\mathbf{X})\mathbf{p}_i = \lambda_i \mathbf{p}_i \tag{1.17}$$

where the λ_is are the eigenvalues.

- The θ_i vectors are known as the *PC scores*. They are orthogonal (i.e. $\theta_i^T \theta_j = 0$ for $i \neq j$) and reflect the magnitude of the PCs. For each PC, there are as many scores as there are samples in the input matrix (i.e. the $\mathbf{\Theta}$ matrix is $n_s \times r$). The score vector $\theta_\mathbf{i}$ is the linear combination of the original $\mathbf{X}$ variables defined by $\mathbf{p_i}$. In other words, the $\theta_\mathbf{i}$ are the projections of $\mathbf{X}$ onto the $\mathbf{p_i}$.

$$\theta_\mathbf{i} = \mathbf{X}\mathbf{p_i} \tag{1.18}$$

- Each $\theta_i \mathbf{p}_i^T$ pair is referred to as the i^{th} PC. Their outer product forms a matrix of rank 1. They are arranged in order of decreasing eigenvalues (i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq \cdots \geq \lambda_l$). The variance captured by each PC is proportional to their eigenvalues. From a statistical standpoint, variance is equated with information. Hence, the first PC accounts for as much of the variability in the data as possible, and hence carries maximum information. Each succeeding PC accounts for as much of the remaining variability as possible.

- l is the maximum number of PCs. It is the smaller of the number of variables and the number of samples (i.e. $l = \min(n_s, n_y)$). The *PCA model*, denoted by $\hat{\mathbf{X}}$, is formed by retaining only a few PCs (i.e. $r < l$). The matrix formed by the minor components ($r + 1$ to l) is not included in the model truncated in this way and is referred to as the *Residual Matrix* (denoted by $\mathbf{E}$). This matrix contains the unimportant variance or noise and any non-linearities, if present.

- PCA is very closely related to Singular Value Decomposition (SVD). When all the components are retained (i.e. $r = l$) then the residual matrix vanishes. For this case, $\mathbf{\Theta} = \mu\mathbf{\Sigma}$ and $\mathbf{P} = \mathbf{V}$.

- Wise et al [55] have argued that a theoretical connection between PCA and state-space models (Equation 1.11) exists. It has been demonstrated that, for processes where there are more measurements than significant states, variations in the process states appear primarily as variations in the PCA scores, while noise mainly affects the residuals. Hence, when limits on the PCA residuals are being derived, only the noise properties of the system have to be taken into consideration while the dynamics of the process do not have to be considered explicitly.

From a process analysis and control perspective, PCA offers several advantages. In all cases PCA derives it's utility by determining the right value of r and discarding the trailing $r+1$ to l PCs. PCA can be used to build models for prediction/estimation. For example, new values of the scores can be estimated from new measurements as follows. First a PCA model is built using the calibration data, i.e. $\mathbf{X} = \mathbf{\Theta}\mathbf{P}^T$. Measurements from the new data are centered using the same mean and variance as the calibration data. For this scaled data, $\mathbf{X}^{new} = \mathbf{\Theta}^{new}\mathbf{P}^T$ and so $\mathbf{X}^{new}\mathbf{P} = \mathbf{\Theta}^{new}\mathbf{P}^T\mathbf{P}$. The loading vectors are orthonormal and so $\mathbf{P}^T = \mathbf{P}^{-1}$. Hence, a new score vector can be estimated as $\theta_{\mathbf{i}}^{\mathbf{new}} = \mathbf{X}^{\mathbf{new}}\mathbf{p_i}$. When PCA is used to solve linear regression problems the approach is known as *Principal Component Regression* (PCR). The pseudoinverse of $\mathbf{X}$ is:

$$\mathbf{X}^+ = \mathbf{P}(\mathbf{\Theta}^T\mathbf{\Theta})^{-1}\mathbf{\Theta} \tag{1.19}$$

Other applications of PCA are shown in Figure 1.8. For illustrative purposes, here PCA operates on an appropriately pre-processed database of five process measurements ($n_y = 5$)

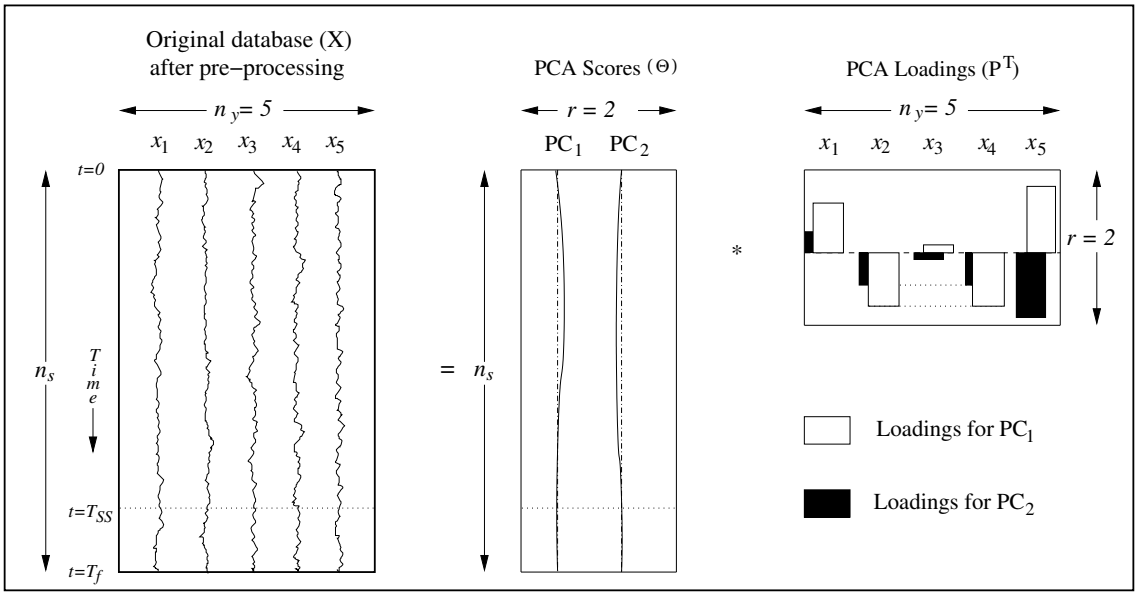

Figure 1.8: How PCA operates on time-series data. The original database is first pre-processed. For illustrative purposes, the number of PCs retained (i.e. r) is 2.

recorded over a sufficiently long time period ($n_s >> n_y$). Now if only the two leading PCs are retained ($r = 2$) in the model, the following features are observed:

1. Reduction in dimensionality of the variable space: PCA compresses a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables. In other words, it finds a lower dimensional space capturing the maximum amount of variance in the input data matrix without losing any significant useful information. This is very important because now there are fewer measurements (two instead of five) to monitor.

2. Filtering of noise: PCA separates the process noise from the data and removes it with the discarded PCs. As a result, the measurement profiles appear smoother.

3. Analysis of process characteristics: PCA results can be interpreted to better understand process behavior as follows:

- The PC scores show how the samples are related. Hence, when scores are plotted versus time, the score variable dynamics are reflective of the major process trends. If the score values for all the retained PCs remain the constant for a substantial period of time, the process is essentially at a steady state. For example, in Figure 1.8, the process is practically at a steady state from $t = T_{SS}$ onwards. θ_1 versus θ_2 plots do not reflect the dynamic aspects but do provide a convenient means of detecting deviations from normal behavior and are particularly suited for batch processes.

- The PC loadings show how the variables are related. Hence, two measurements which have similar loadings are practically identical (for example measurements x_2 and x_4 in Figure 1.8). If a measurement has very small loadings for all the retained PCs, it means that this measurement does not paricipate in the PCA model (for example measurement x_3). An alternative approach to analyze and interpret PC loadings information is using $\mathbf{p_1}$ versus $\mathbf{p_2}$ plots.

Orthogonal PCA

Rao [44] presented a variant to the standard PCA formulation.

$$\left.\begin{aligned} \max_{\mathbf{p}_i} \quad & (\mathbf{X}\mathbf{p}_i)^T(\mathbf{X}\mathbf{p}_i) \\ \text{subject to} : \quad & \mathbf{p}_i^T\mathbf{p}_i = 1 \\ & (\mathbf{X}\mathbf{p}_i)^T\mathbf{u} = 0 \end{aligned}\right\} \tag{1.20}$$

Without the constraint $(\mathbf{X}\mathbf{p}_i)^T\mathbf{u} = 0$, Equation 1.20 represents ordinary PCA which maximizes the covariance of $\mathbf{X}$. Zheng [59] and Zheng et al [60] suggested a different approach for orthogonal PCA calculations. In their approach, before PCA is applied to

the $\mathbf{X}$ data, a matrix $\mathbf{A}$ is defined as $\mathbf{A} \equiv (\mathbf{X})^T\mathbf{u}$. Next, SVD of $\mathbf{A}$ is used to transform the ordinary PCA $\mathbf{\Theta}$ scores into two sets (see Appendix). This transformation is represented as:

$$\mathbf{Z} = \mathbf{\Theta} * \mu = [\mathbf{R}, \mathbf{N}]$$

where subspace $\mathbf{R}$ denotes the matrix of transformed measurements that are correlated with the manipulated variable inputs $\mathbf{u}$ (i.e. $\mathbf{R}^T\mathbf{u} \neq 0$) and $\mathbf{N}$ is the matrix of transformed measurements that are not correlated with $\mathbf{u}$ (i.e. $\mathbf{N}^T\mathbf{u} = 0$).

Partial Least Squares (PLS)

Partial least squares (PLS) is also known as projection to latent structures. It is performed when one is interested in studying and monitoring the variations in $\mathbf{X}$ that are most influential on the quality and productivity variables, $\mathbf{Y}$. This technique tries to capture the maximum amount of covariance between $\mathbf{X}$ and $\mathbf{Y}$ while at the same time describing the variance within $\mathbf{X}$ and $\mathbf{Y}$.

$$\left.\begin{aligned} \mathbf{X} &\equiv \mathbf{\Theta}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &\equiv \mathbf{U}\mathbf{Q}^T + \mathbf{F} \\ \mathbf{Y} &\equiv \mathbf{X}\beta + \mathbf{\Gamma} \end{aligned}\right\} \tag{1.21}$$

PCR captures the maximum variance in $\mathbf{X}$. MLR achieves maximum correlation between $\mathbf{X}$ and $\mathbf{Y}$. PLS tries to do both by maximizing covariance between $\mathbf{X}$ and $\mathbf{Y}$. Hence PLS tries to provide the best of MLR and PCR. The pseudoinverse of $\mathbf{X}$ is calculated as:

$$\mathbf{X}^+ = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}(\mathbf{\Theta}^T\mathbf{\Theta})^{-1}\mathbf{\Theta} \tag{1.22}$$

Some of the other popular variants are Multi-way PCA and Multi-way PLS which are capable of handling three-dimensional data. Zheng et al [60] have demonstrated some

applications of these multi-way techniques in an industrial batch process.

Statistics associated with PCA models

Several statistics associated with PCA models can also be used as measures to detect abnormal behavior in processes. Two of the most commonly used ones are the *squared prediction error* (SPE) and the *Hotelling* T^2. SPE is often referred to as the lack of fit statistic or the Q-residual. For the kth sample from a set of measurements, these statistics are defined as:

$$Q_k = x_k(I - \mathbf{P}\mathbf{P}^T)x_k^T \tag{1.23}$$

$$T_k^2 = t_k\lambda^{-1}t_k^T \tag{1.24}$$

SPE indicates how well each sample conforms to the PCA model. As a result it also detects any new variations occuring in the process. The T^2 statistic is a measure of the variation in each sample within the model. In other words. it captures larger than normal variations.

Measures of model performance

Model validation is a means of measuring model performance, i.e. it's fit and predictive power. Model performance is determined by evaluating "average" deviations of the model estimates from the measured data used to build it (callibration set) and new data (test set). Obviously, the best test of a model is how it works with samples it has not seen before because the real objective is to apply models to new data. In either case, the Root Mean Square (RMS) error is a means of quantifying the performance:

Determination of the number of PCs to retain

One of the critical decisions to be made during PCA or PLS model development is the determination of the number of LVs to retain. Since this study is primarily concerned with PCA, the following discussion deals with the number of PCs retained, r. As more PCs are retained, noise is filtered out to a lesser degree and begins to get incorporated into the model. As a result, the fit to the calibration set improves. However, as shown in Figure 1.9 (A), the validity of the model when applied to new data (i.e. test set) eventually declines. For process measurements, a knowledge of the number of sources of variation during database generation can be used to determine r. For example, if two manipulated variables are perturbed and there is one active disturbance during the identification experiment, then $r = 2 + 1 = 3$. If however, the number of sources of variation is not known, there isn't a single definite method that can be used as the cut-off criteria. In such a case, one of the following, often conflicting, rules of thumb may be used for determining r:

1. Identifying a "knee" or "elbow" (a break or sharp change) in the eigenvalue versus number of PCs plot (Cattell [4]). This approach, also known as the *Scree test*, is depicted in Figure 1.9 (B).

2. Retaining PCs with an eigenvalue greater than a threshold value (for example, for auto-scaled data, this threshold could be 1.0 as per the Kaiser criterion). See Jolliffe [21] for details. This is shown in Figure 1.9 (B).

3. Determining the minima in the PRESS versus number of PCs plot. The PRESS (Prediction Error Sum of Squares) is the square residual error summed for each PC. The residuals are calculated for the test-sets in cross-validation. This is shown in

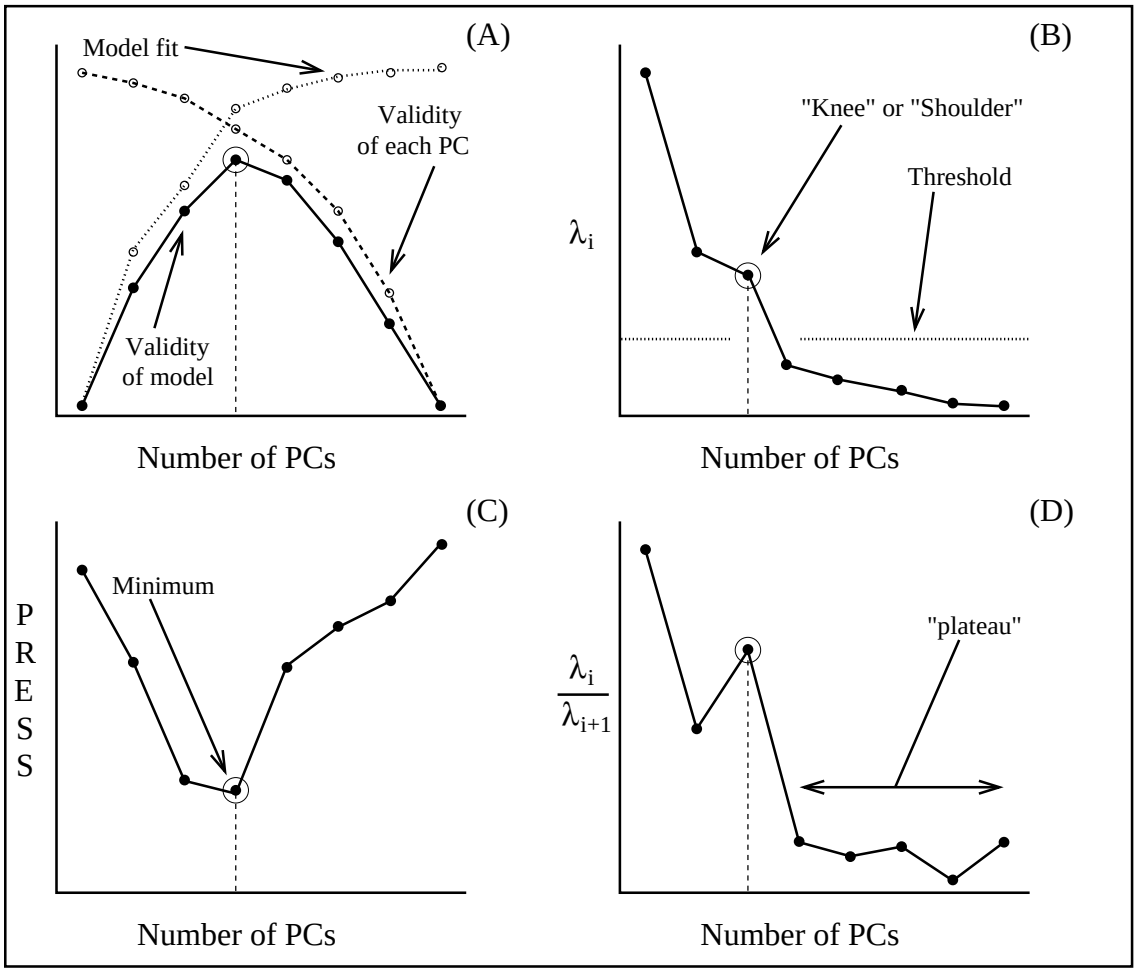

Figure 1.9: Determining the number of PCs to retain.

Figure (1.9 (C)).

4. Identifying a "plateau" in the ratios of eigenvalues or a small change in successive eigenvalues in the ratio of eigenvalue versus number of PCs plot. This is shown in Figure 1.9 (D).

5. If the noise level is known, retaining PCs with %variances greater than the noise level. For example, if the measurements are obtained using thermocouples with a known signal to noise ratio of 9, then r should be chosen such that only about (100-10=) 90% variance is captured by the PCs.

Several other methods with a higher degree of mathematical rigor have also been proposed. For example, Qin and Dunia [43] proposed a method based on the variance of reconstruction error criterion. Valle et al [51] compared this approach to other methods.

1.4 Process measurements

No matter what the approach to process control, process measurements are an essential part of the control system. Field sensors for pressure, temperature, level and flow (referred to as PTLF) provide process measurements for the base control system. As can be seen in Table 1.2, the time constants, in seconds, for typical elements in base control systems and PTLF measurements are quite small, i.e. such control loop elements and measurements are fast [1]. Moreover, PTLF measurements are usually cheap, robust, quite reliable, readily available and easy to install. However, measurements for product properties of interest from a commercial and quality control perspective, such as compositions, viscosity, particle size, etc; are usually obtained through analyzers. Such analyzers are usually a part of the upper layer. The analysis may be carried out on-line or

Table 1.2: Typical control loop element time constants (adapted from Riggs [46]).

Element/measurement	Time constant (s)
Control valve	3 to 15
Control valve with valve positioner	0.5 to 2
Thermocouple with thermowell	6 to 20
Resistance temperature detector (RTD) with thermowell	6 to 20
Pressure sensor	< 0.2
Differential pressure level indicator	< 1

off-line. Off-line analysis is usually carried out in quality control laboratories. In general, both on-line and off-line analyzers have the following disadvantages:

1. They are speciallized sensors and so some of them are not readily available.

2. They are expensive.

3. The associated sampling and analysis step can lead to significant time delays.

4. The output is usually not continuous.

5. They are not very reliable or robust and require frequent re-callibration.

1.4.1 Inferential control using soft-sensors

One approach to tackle the problems associated with analyzers is to use inferential control systems. Here, the *primary variables* which are difficult to measure are "inferred" from the *secondary variables* (i.e. PTLF) that are easier to measure. *Soft-sensors* are algorithms implemented as computer models that perform the task of inference. There are several approaches for developing soft-sensors. Although there isn't a clear demarcation, inferential control techniques using soft-sensors may be broadly classified into *direct* and *indirect methods* (Amirthalingam et al [1]).

Direct methods

In this approach, an *estimator* for the primary variable is coupled with a regular feedback controller. Static, linear estimators are typically used. These are built using either steady state gain information or are identified directly from plant test data. Regression techniques such as MLR, PCR or PLS are used for this purpose. Artificial neural networks (ANNs) find use when the estimation requirements are nonlinear. Time series regression is a means of incorporating dynamics in such estimators.

Indirect methods

The indirect method, also referred to as the model-based approach, is based on optimal prediction and control theory. In this approach primary variable measurement and it's control are integrated. In other words, the soft-sensor is embedded within the controller. A state estimator, also referred to as an *observer*, is coupled with a state feedback controller.

The *data-based* or *data-driven approach* is very attractive because in recent years, process PTLF measurements are being collected and recorded with greater ease. This development is the result of widespread use of sophisticated computer-based data acquisition and logging tools available in distributed control systems (DCS). Davis et al [9] review the current status and applications of process data analysis and interpretation technologies. They provide a detailed discussion on the practical importance of these methods and the diversity of their applicability through an integrated perspective. Of the several approaches presented in their review [9], this thesis is restricted to multivariate linear projection and model based approaches.

1.4.2 General guidelines for selecting secondary measurements

The criteria used for selecting secondary measurements (Brosilow and Joseph [3]) are:

1. The cost of installing and maintaining the sensor.
2. The ability of the control system to cope with sensor failure.
3. The ability to get "good" estimates of the output.

Both the first and second criteria suggest using a small number of measurements. The second criterion favors simpler control structures that do not depend on many measurements. As the complexity of the control system increases, it becomes increasingly difficult to cope with even temporary sensor failures. With the kind of formulations discussed in this thesis, the third criterion leads to specifying the following four key charecteristics for the selected measurement to possess:

1. It should be correlated with product quality.
2. It should be affected by the stochastic upsets.
3. It should be affected by the manipulated setpoints.
4. It should lead the product quality variable's dynamic response.

1.4.3 Measurement selection using process models

Kookos and Perkins [23] have proposed a mixed integer linear programming (MILP) based approach. They took the dynamic aspects of the process into consideration but the method is only applicable to systems where the number of measurements employed is equal to the numbers of manipulated variables.

Pannocchia and Brambilla [37] have approached measurement selection problem using consistency tests for the models. The accuracy of PLS models in the feedback mode is examined. The approach emphasizes steady-state errors (rather than dynamics) and is oriented towards a linear, PI type controller, i.e. direct methods for inferential control.

In conclusion, the measurement selection problem for inferential control when process models are not available needs to be addressed.

1.5 Optimal control using data-based methods

SPC is essentially an off-line quality control technique. However the tools and techniques discussed in section 1.3 can be modified and adapted to build on-line applications too. Since a multivariate statistical model is essential for this purpose, a process data-base with a significant degree of variability in it's measurements is necessary.

1.5.1 Data-base generation

The sources for the data used to build inferential control applications could be either historical plant operation data or from designed experiments for model identification. It should be noted that the data-set used should involve the manipulated variables moves be independent otherwise correlations are built into the data. These correlations are detrimental to performance if control is applied through the model being developed. Hence, quite often, historical data alone isn't sufficient to build soft-sensors for quality control. In either case, the process is expected to be at a steady state and the measurements need to be somehow varied about this steady state. Historical operation data are records of process measurements collected over a long period of time (weeks, months or maybe years). In such a case, common cause variations such as process noise and disturbances

are the sources of variability. Plant tests for model identification are designed experiments where the variability in the measurements is achieved by systematically changing a set of manipulated variables which are known to affect the measurements under consideration.

Historical operation data

Amirthalingam et al [1] have suggested the following guidelines for choosing slices of data from the large amounts of data usually available in historical databases. As per them, one should select a period

1. in which the system's production rate did not deviate significantly from the nominal value.
2. in which the reliability of the critical measurements were high (such as after a shut-down maintenance).
3. that includes no serious plant upset.
4. where the disturbances were likely to be representative of most other times.
5. where input manipulations were not unusually high.

Plant test data

Although historical operation data is useful in analyzing the steady state relationships between the measurements and in building soft-sensors, plant tests are essential when information regarding dynamic behavior is being sought after. Whether the implemention is direct or indirect, a dynamic relationship between the manipulated variables and the controlled variables is essential to build soft-sensors based controllers. Well designed plant tests are conducted to identify dynamic process models by perturbing the manipulated

variables in a systematic manner. Pseudo Random Binary Sequences (PRBS) are the most commonly used form of excitation. A PRBS is a truncated form of a Random Binary Sequence (RBS) with a period m (i.e. the RBS repeats itself after m data points). PRBS inputs are persistently exciting of order m. A higher order for the persistency of excitaion is helpful in generating more accurate models. Impulse and step inputs are of order zero while sinusoidal inputs are of order 2. As a result, for impulse, step and sinusoidal inputs, bias-free estimation of the model parameters is not possible even with an infinite data set, particularly when noise is present.

1.5.2 Control in the score space

Since the indirect, model-based approaches to soft-sensor development have significant advantages over the direct approaches, the indirect approaches are discussed in further detail. Piovoso and Kosanovich [39] presented an approach in which PCA models were used for on-line control. They termed this approach as score control. Chen et al [6] developed a lagged PCA model which was used within the MPC framework. Only stationary upsets were considered and so quality control was achieved by driving the scores to zero. McAvoy [31] extended this approach by using orthogonal PCA and demonstrated how quality control can be achieved even in the absence of quality measurements. He used a dynamic score model in the form of an auto-regressive exgenous input (ARX) model identified in the presence of a disturbance. A 44% reduction in variation in the product quality was achieved when the disturbance encountered was the same as the one present during identification. Controller performance deteriorated when a non-stationary disturbance was encountered, and McAvoy [31] showed how a steady state proces model coupled with intermittent lab results can be effective in tackling this shortcoming.

Although the results obtained by McAvoy [31] are very promising, the methodology needs further investigation.

1.6 Research philosophy

The chance of applicability of any new technology improves and the possibility of failure decreases with it's simplicity. So the basic philosophy guiding this study is to achieve a control strategy that is "simple, yet effective". The control strategy developed in this study is easy to troubleshoot and straightforward enough for plant operators to understand. Process control literature abounds with publications presenting complicated control algorithms and their behavior simulated on linearized or overly simplistic process models dealing with fictitious components. Instead the process models used in this study are fairly detailed and realistic. Their dynamics have been benchmarked against industrial data and thereby they are more convincing.

Critiquing the process control literature, Bequette [2] pointed out that it is desirable to compare control techniques and to include manipulated variable responses in addition to controlled variable responses. This aspect is taken into consideration. Also, the aim of research is to benefit a larger community and so the results are presented in a reproducible manner.

1.7 Research goals

The basic premise of this research study is to develop new strategies for inferential model predictive control using statistical techniques. The idea is to utillize historical databases and plant testing to generate multivariate statistical models as soft-sensors which provide an alternative to on-line quality measurements. Specifically, the principal

objectives are:

1. To develop multivariable extensions to McAvoy's formulation [31].

2. To develop alternatives to the basic formulation so as to take advantage of an on-line analyzer, if available.

3. To develop guidelines for tuning the controllers developed.

4. To analyze and understand the effect of the nature of the disturbance (i.e. frequency, size, direction and stationarity) on controller performance.

5. To develop schemes whereby the wealth of information contained in historical plant operation databases is utillized to build soft-sensors.

6. To develop strategies for selecting the best set of measurements to be used in developing the score model.

7. To better understand orthogonal PCA and to develop a strategy to determine the number of principal components to retain during score model development.

8. To develop a dynamic score model identification methodology specific to the type of controllers being discussed.

9. To use well known process models to demonstrate the effectivenes of these schemes.

In short, all aspects related to the design and analysis of an approach integrating statistical process control and model-based, feedback process control will be thoroughly investigated and practical guidelines for it's implementation will be provided.

1.8 Thesis outline

This Ph.D. thesis is organized as follows. In the next chapter, Chapter 2, the test-bed problems used to evaluate the performance of the advanced process control strategies presented in this thesis are described. This includes the Tennessee Eastman challenge process (TE) and the Azeotropic distillation column (AZ) of the Vinyl Acetate process. Chapter 3 presents the basic inferential model-predictive controller. This controller is essentially a multivariable extension to McAvoy's [31] formulation. It's behavior is simulated on the two test-cases described in Chapter 2. Chapter 4 discusses implementation issues pertaining to the steps prior to on-line implementation. This order of presentation is chosen so as to first make a case for the new approach and then discuss it's implementation in detail. Chapter 4 provides details about a unique scheme for utilizing historical plant operation databases, combined with plant test data, for developing soft-sensors. It also discusses a new methodology for measurement selection and several approaches for identification specifically developed for the controllers under study. A better understanding of the orthogonal PCA calculation and it's use is also provided. Chapter 5 discusses an alternative controller formulation. The controller's ability to deal with different kinds of disturbances and possible extensions that might be necessary under certain disturbance scenarios are also discussed in this chapter. Finally Chapter 6 concludes with a summary of the results and recommendations for future work. Appendix A briefly reviews some linear algebra preliminaries relevant to this study. Appendix B discusses the details of the Propane-propylene splitter case study. The source code in the C programming language and a MEX interface to MATLAB is also provided here. The notation used throughout the thesis is described in the next section.

1.9 Notation

Matrices are shown in capital and bold, for example $\mathbf{X}$. Vectors are denoted by a lower case and bold, for example $\mathbf{x}$. Scalars are denoted in smaller case, for example x.

Abbreviations

APC	Automatic process control
AZ	Azeotropic
DMC	Dynamic matrix control
MP-SPC	Model predictive statistical process control
MPC	Model predictive control
MV	Manipulated variable
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares
SP	setpoint
SPC	Statistical process control
TE	Tennessee Eastman

Symbols

$\mathbf{A_D}$	Dynamic matrix
$\mathbf{E}$	Residual matrix for PCA
$\mathbf{E_p}$	Error matrix in MPC
$\mathbf{h}$	vector of impulse response coefficients
M	Control horizon (scalar)
P	Prediction horizon (scalar)
$\mathbf{s}$	vector of step response coefficients
$\mathbf{X}$	Data-set
$\mathbf{X}$	Process measurements

Greek symbols

σ	variance

Superscripts

$\square^T$	transpose of matrix or vector
$\square^{-1}$	inverse of matrix
$\square^\circ$	degrees
$\hat{\square}$	model or predicted

Subscripts

$\square_B$	Base operation
$\square_I$	Improved operation
$\square_k$	time step.
$\square_r$	process measurement
$\square_s$	manipulated variable

Other

$\overline{\square}$ average or mean $\square$.

$\Delta\square$ difference or change in $\square$.

Chapter 2

Test-beds for advanced process control studies

An essential part of advanced process control research involves testing the technology with realistic problems. Before on-line applications in commercial scale plants, new control systems or strategies are evaluated at a pilot-plant scale. Often, extensive computer simulations are conducted to test new ideas before these can be considered for on-line application at any scale. Some of the popular test-bed problems are:

1. The Tennessee Eastman (TE) challenge problem suggested by Downs and Vogel [12].
2. The Vinyl Acetate (VAc) process suggested by Luyben and Tyreus [28].
3. The Shell procees control problem (see Prett and Morari [40]).
4. The recyle process proposed by Reyes-DeLeon and Luyben [45].
5. The hydrodealkylation (HDA) of toluene process (see Douglas [13]).

Based on the availability of nonlinear dynamic simulations, process complexity and popularity in the process control literature, out of the above list, two processes have been chosen as test-beds for this study. The first is the Tennessee Eastman challenge problem, henceforth referred to as TE. The second case study focuses on the Azeotropic distillation column of the Vinyl Acetate process, henceforth referred to as AZ. In both cases, this study assumes that a base control system for the process is already in place and the plant is running at a steady state. The controller tuning parameters used are the same as those reported in the relevant publications and have not been optimized further.

2.1 The Tennessee Eastman (TE) challenge process

Downs and Vogel [12] suggested the Tennessee Eastman challenge process (TE) problem. It consists of five major pieces of equipment: reactor, condenser, separator, compressor, and stripper with a gas recycle stream. There are seven components involved: reactants A, C, D and E; inert B; products G and H and by-product F. All the components, except for the inert B, are condensible. All the reactions are irreversible and exothermic. The stoichiometry is:

$$\left.\begin{array}{c} A(g) + C(g) + D(g) \rightarrow G(g) \\ A(g) + C(g) + E(g) \rightarrow H(g) \\ A(g) + E(g) \rightarrow F(g) \\ 3D(g) \rightarrow 2F(g) \end{array}\right\} \tag{2.1}$$

Hence, 1 mole of G is produced by 1 mole of D and 1 mole of H is produced by 1 mole of E. Further details about the process, typical disturbances, base operating conditions and control objectives are provided in the original paper and so are not reproduced here.

This test case has been extensively studied and several researchers have published base control structures for stabillizing the plant. Here, the one suggested by McAvoy and Ye [32] is adopted. Figure (2.1) shows the process flowsheet along with the controller pairings for the base control system.

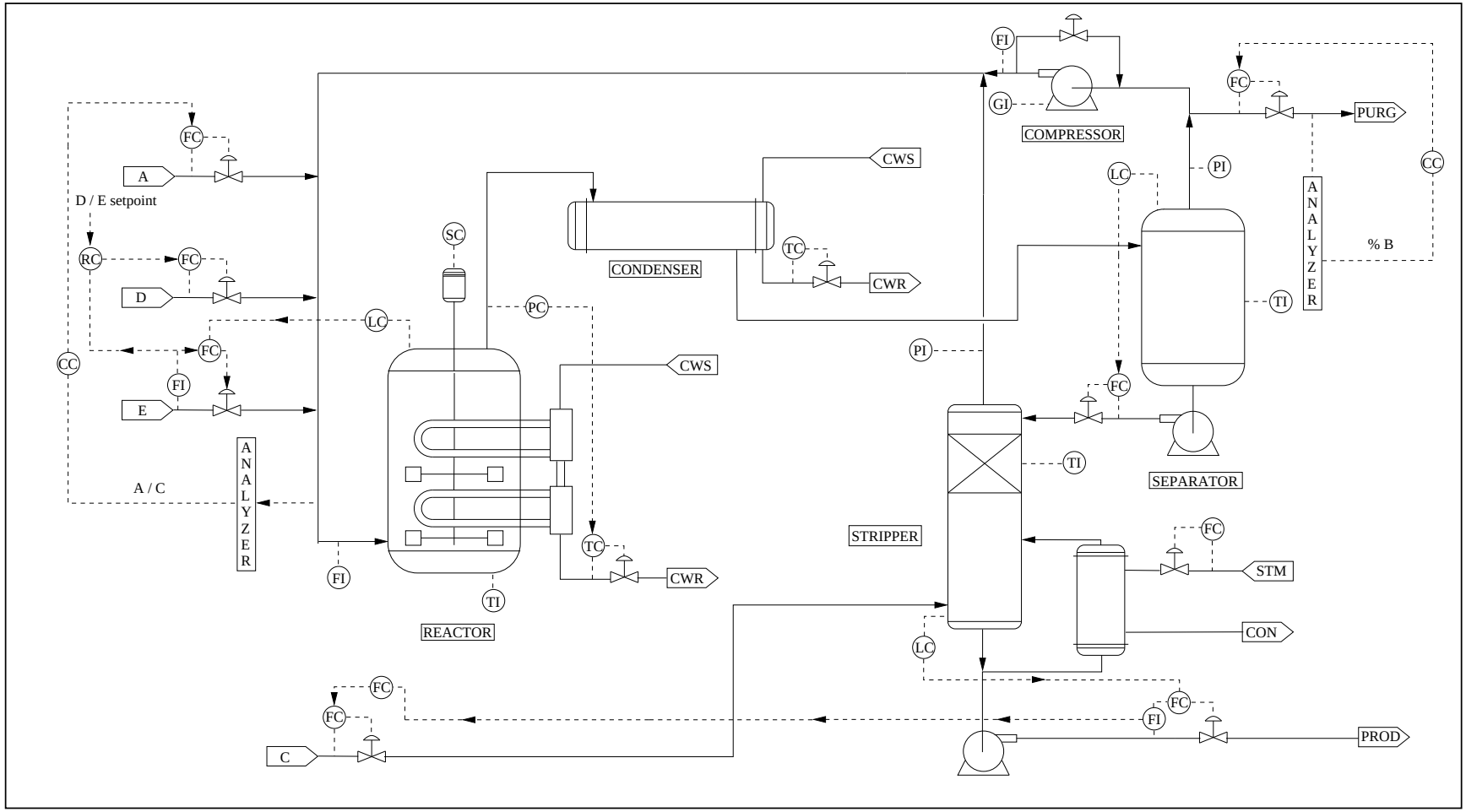

Figure 2.1: The Tennesse Eastman process with it's base control system.

The ratio of the composition of G to that of H in the product stream (referred to henceforth as G/H ratio for brevity) is the quality variable. McAvoy and Ye [32] point out that the D/E ratio in the feed can be used to control the G/H ratio. The E-feed flow is already being used to control the reactor level. Owing to this, the reactor level and the D-feed flow may be used to change the D/E ratio. Hence, the control objective is to maintain the G/H ratio in the product ratio constant using the D-feed flow setpoint and/or the reactor level setpoint as the two manipulated variables.

Table (2.1) is a list of candidate secondary measurements that may be used to build a soft-sensor. It should be noted that when the D-feed SP is being used as a manipulated variable, Y(2) should not be used to build the soft-sensor. The same applies to Y(8) when reactor level is being manipulated.

Table 2.1: List of candidate secondary measurements for TE.

Symbol	Name	Symbol	Name	Symbol	Name
Y(1)	A-feed flow	Y(13)	Separator pressure	Y(25)	React Feed %C
Y(2)	D-feed flow	Y(14)	Separator exit flow	Y(26)	React Feed %D
Y(3)	E-feed flow	Y(15)	Stripper level	Y(27)	React Feed %E
Y(4)	C-feed flow	Y(16)	Stripper pressure	Y(28)	React Feed %F
Y(5)	Recycle flow	Y(17)	Product flow	Y(29)	Purge Gas %A
Y(6)	Reactor feed	Y(18)	Stripper temperature	Y(30)	Purge Gas %B
Y(7)	Reactor pressure	Y(19)	Steam flow	Y(31)	Purge Gas %C
Y(8)	Reactor level	Y(20)	Compressor work	Y(32)	Purge Gas %D
Y(9)	Reactor temperature	Y(21)	Reactor cooling water temperature	Y(33)	Purge Gas %E
Y(10)	Purge flow	Y(22)	Condensor cooling water temperature	Y(34)	Purge Gas %F
Y(11)	Separator temperature	Y(23)	React Feed %A	Y(35)	Purge Gas %G
Y(12)	Separator level	Y(24)	React Feed %B	Y(36)	Purge Gas %H

The base control system has been implemented as a cascade arrangement of single-loop PI controllers. Table (2.2) lists the inner loop (primary) controller pairings for the base control system and their tuning parameters while Table (2.3) does the same for the outer loop (secondary) controllers.

Table 2.2: Inner cascade PI controller constants for TEC base control system.

Controlled	Manipulated	K_c	T_R (min)
A-feed flow	A Valve	150 $(\%/kscmh)$	0.075
C-feed flow	C Valve	14 $(\%/kscmh)$	0.13
D-feed flow	D Valve	0.026 $(\%/kg.h^{-1})$	0.10
E-feed flow	E Valve	0.017 $(\%/kg.h^{-1})$	0.10
Purge flow	Purge Valve	200 $(\%/kscmh)$	0.06
Separator Exit flow	Separator Exit Valve	2.0 $(\%m^3h{-}1)$	0.12
Product flow	Product Valve	3.0 $(\%m^3h^{-1})$	0.12
Steam flow	Steam Valve	2.5 $(\%/kg.h^{-1})$	1.5
Condenser Cool. Water Temp.	Condenser Cool. Water Valve	-8.0 $(\%/^0C)$	1.8
Reactor Cool. Water Temp.	Reactor Cool. Water Valve	-10.0 $(\%/^0C)$	0.70

Table 2.3: Outer cascade PI controller constants for TEC base control system.

Controlled	Manipulated	K_c	T_R (min)
Reactor level	E-feed setpoint	500 ($kg.h^{-1}/\%$)	200
Separator level	Separator exit flow setpoint	-0.12 ($m^3h^{-1}/\%$)	200
Stripper level	Product flow setpoint	-0.07 ($m^3h^{-1}/\%$)	200
Reactor pressure	Reactor cooling water temp. setpoint	-0.02 ($^0C/kPa$)	30
Product flow	C-feed flow setpoint	0.6 ($kscmh/m^3h^{-1}$)	60
A/C in Reactor feed	A-feed setpoint	1.0 ($kscmh$)	164
B in purge	Purge setpoint	-0.0366 ($kscmh/\%$)	250

The authors of the original paper, Downs and Vogel [12], have provided several disturbance scenarios. They have referred to these disturbances as IDVs. For illustrative purposes, this study considers two out of the 20 different IDVs. IDV(1) is a step in the A to C feed ratio with the B composition kept constant for stream 4 (i.e. C-feed). IDV(8) is a random variation in A, B and C feed compositions for stream 4. It is a well known fact that multivariable controller performance is dependent on the type and direction of the disturbance inputs. As a result, it is desirable to monitor controller performance over an appropriately wide frequency range for the IDVs. As per the authors [14], the disturbance models for the TEC process were designed to model the types of random variation experienced in actual processes while at the same time having a continuous first derivative so that numerical integrators work satisfactorily. The models are not based on any published techniques (i.e. transfer function first order plus dead time (FOPTD), ARX, etc). There are four parameter vectors: *sspan*, *szero*, *hspan* and *hzero* which determine the nature of the disturbances. However, as a matter of policy, the authors [14] have decided not to disclose modeling details. In order to better understand the nature and effects of IDV(8), the original FORTRAN code has been modified and the mole fraction of component A in stream 4, which in the original version was inaccessible, is extracted and analyzed. Figure (2.2) shows the effect of IDV(8) on product quality (i.e. G/H ratio) with just the Base control system in place. It can be seen that the fluctuations in IDV(8) have a detrimental effect on product quality. In order to reduce the variability in the G/H ratio, a control system over and above the Base control system is necessary. Before designing such a control system however, it would be worthwhile to analyze the mole fraction of component A in stream 4 for it's frequency content. It can be seen from Figure (2.3) that not only is there a significant range of frequencies being covered but also

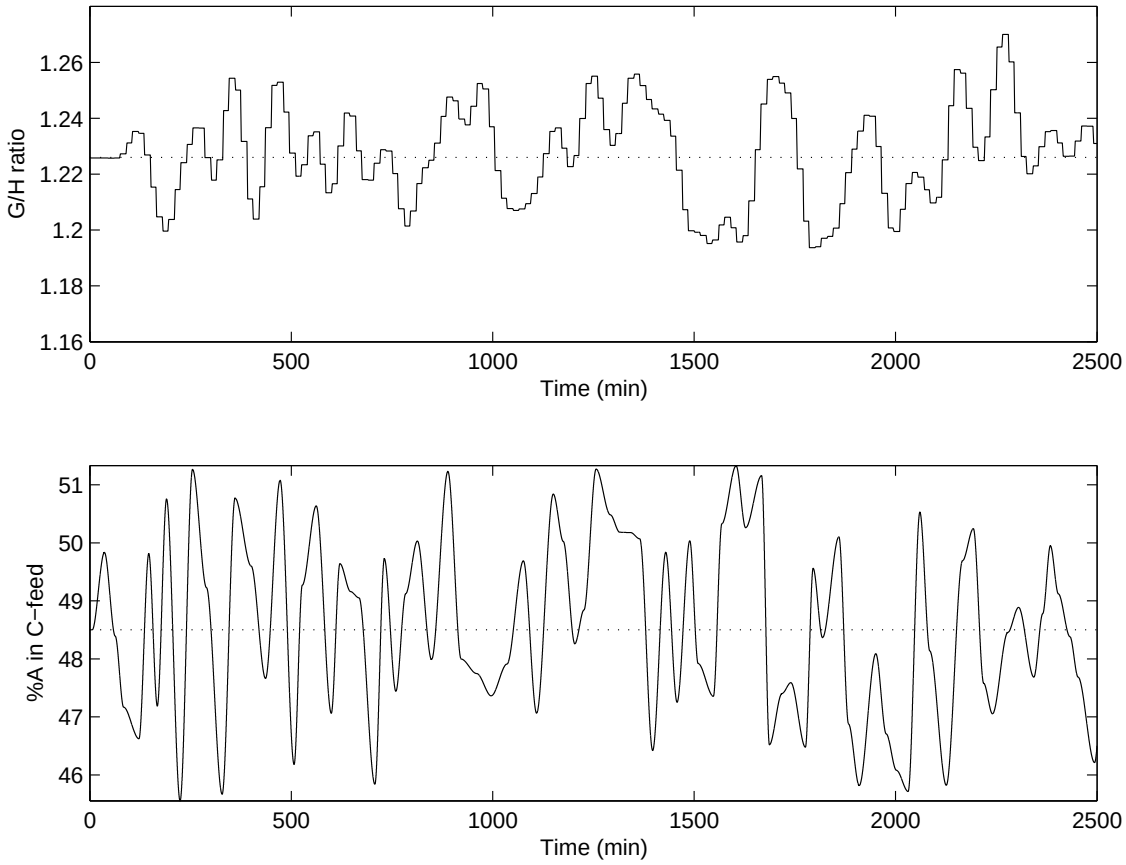

Figure 2.2: Effect of IDV(8) on product quality.

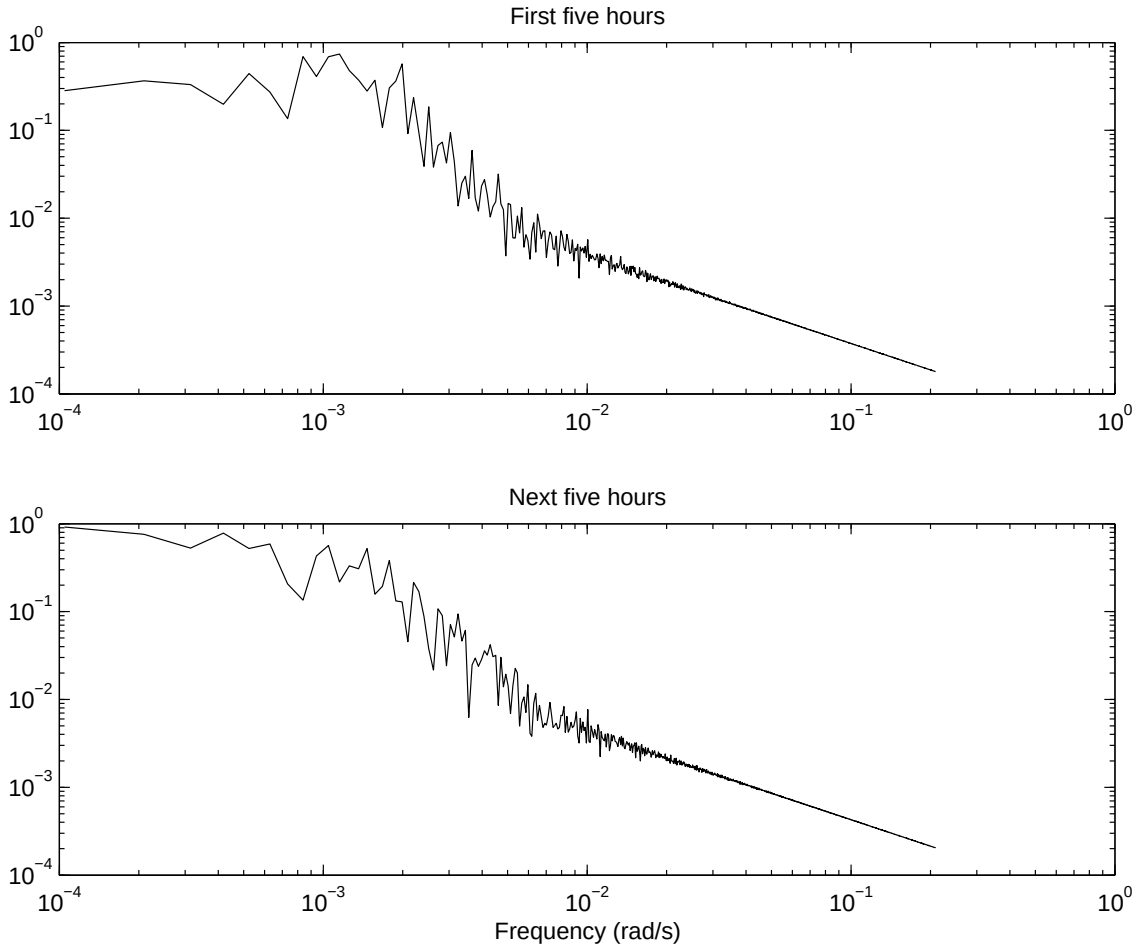

Figure 2.3: Spectral decomposition of IDV(8) for TEC problem.

that the nature of the disturbance changes with time. Another key system property to take note is the open-loop time constant. The D/E ratio in the reactor feed is the primary handle on the G/H ratio in the product. The two manipulated variables available affect the product quality through the D/E ratio. Figure (2.4) is the response of the G/H ratio in the product to the bump-testing of the D/E ratio in the reactor feed. It can be seen that for small step sizes ($\pm 4\%$), the response is essentially linear. Moreover, the open-loop time constant is about 112 minutes and so the ultimate frequency is about 1×10^{-3} rad/s.

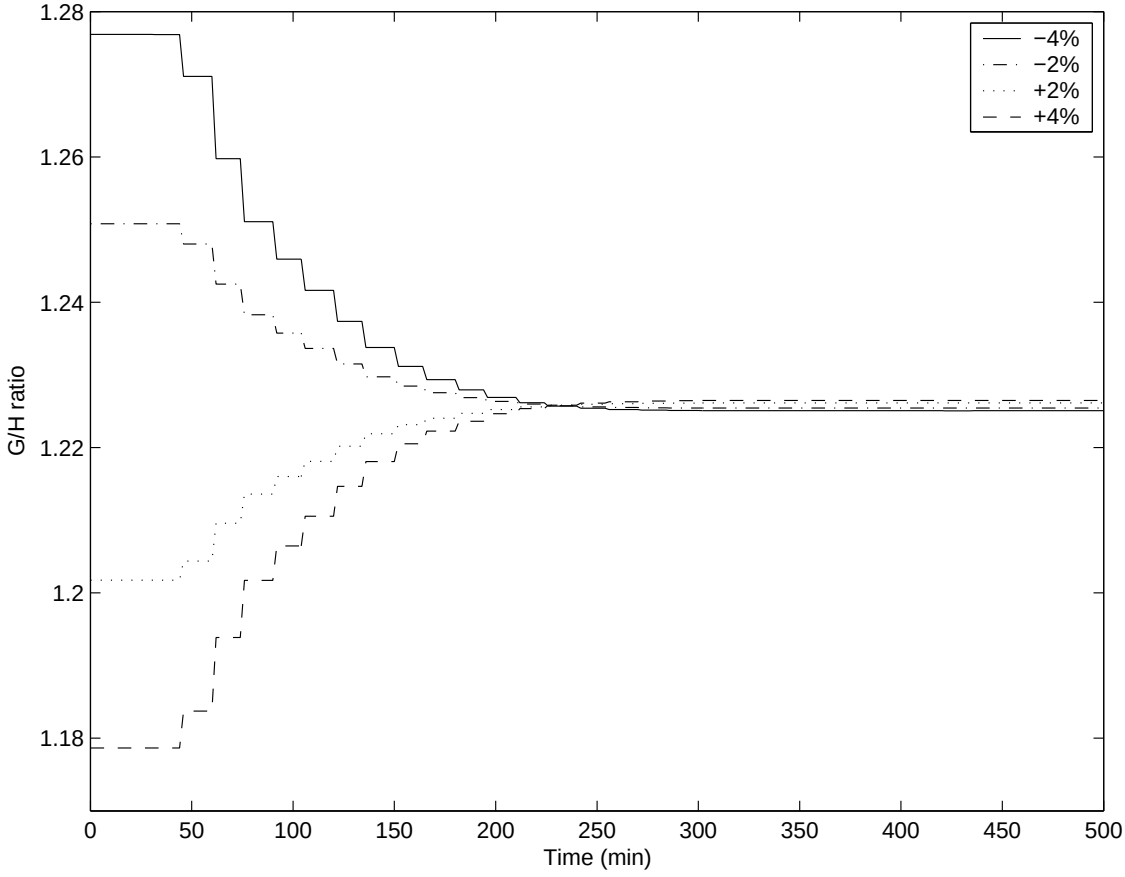

Figure 2.4: Response of G/H ratio in product to the bump testing of D/E setpoint with different step sizes.

2.2 The Azeotropic (AZ) tower

Although the Tennessee Eastman challenge process (TE) is by far the most popular industrial test-bed process amongst researchers, the reaction section of a vinyl acetate manufacturing process, referred to as the Vinyl Acetate (VAc) process [28], is considered a better case study for process control research. The reason for such a view is that the VAc process has common real components in a realistically large process flowsheet with standard chemical unit operations, both gas and liquid recycle streams, and energy integration. It consists of eleven basic unit operations: reactor, vaporizer, separator, compressor, absorber, CO_2 removal system and azeotropic distillation column with decanter. Besides these, there are several coolers, heaters, a steam drum and a process-to-process heat exchanger. There are seven components involved: ethylene (C_2H_4), oxygen (O_2) and acetic acid (HAc i.e. CH_3COOH) are the raw materials; ethane (C_2H_6) is an inert; vinyl acetate (VAc i.e. $CH_2 = CHOCOCH_3$) is the product while water (H_2O) and carbon dioxide (CO_2) are the by-products. Further details about the process, typical disturbances, base operating conditions and control objectives are not reproduced here because they are provided in the original paper [28] and in Chen et al [7]. The latter publication also provides access to a dynamic model for the process and the details of the base control system.

This study focuses only on the azeotropic distillation tower which is in the final stage of the process. The tower is highly non-linear in it's behavior. The column bottoms level (LB), decanter organic level (LO) and decanter aqueous levels (LA) are integrating in nature and so need to be controlled. The decanter temperature is controlled using condenser cooling water flow. However, these four controllers are not enough to stabilize the tower because the tower exhibits parametric sensitivity to organic reflux flow and feed

compositions. An additional control loop is necessary to prevent the tower from settling to the undesirable, water-rich, steady state. One way to deal with this is to control the fifth tray temperature (T(5)) using the reboiler heat duty (QR).

The mole fraction of water in the bottoms stream (xBH_2O) is the quality variable. Luyben et al [29] point out that the organic reflux flow rate (FR) can be used to control xBH_2O. The reboiler heat duty (QR) can also be used a manipulated variable for this purpose but is already being used to control the 5^{th} tray temperature. This latter pairing is absolutely essential to ensure stability so instead the 5^{th} tray temperature can be used as a manipulated variable. Hence, the control objective is to maintain the mole fraction of water in the bottoms stream constant at 0.09 using the organic reflux flow setpoint and the 5^{th} tray temperature setpoint as the two manipulated variables. The soft-sensor to be built has the temperatures of trays 1 through 20 as the candidate secondary measurements.

Figure (2.5) shows the process flowsheet along with the controller pairings for the base control and composition control system. Table (2.4) provides the controller tuning parameters used. It should be noted that these values are those reported in Chen et al [7] and have not been optimized further. The chromatographic analyzer has a 10 minute sampling frequency and a 10 minute dead time. For an open-loop time constant of about 20 minutes, this delay is very dominant. Owing to this delay, the controller gain is kept very small.

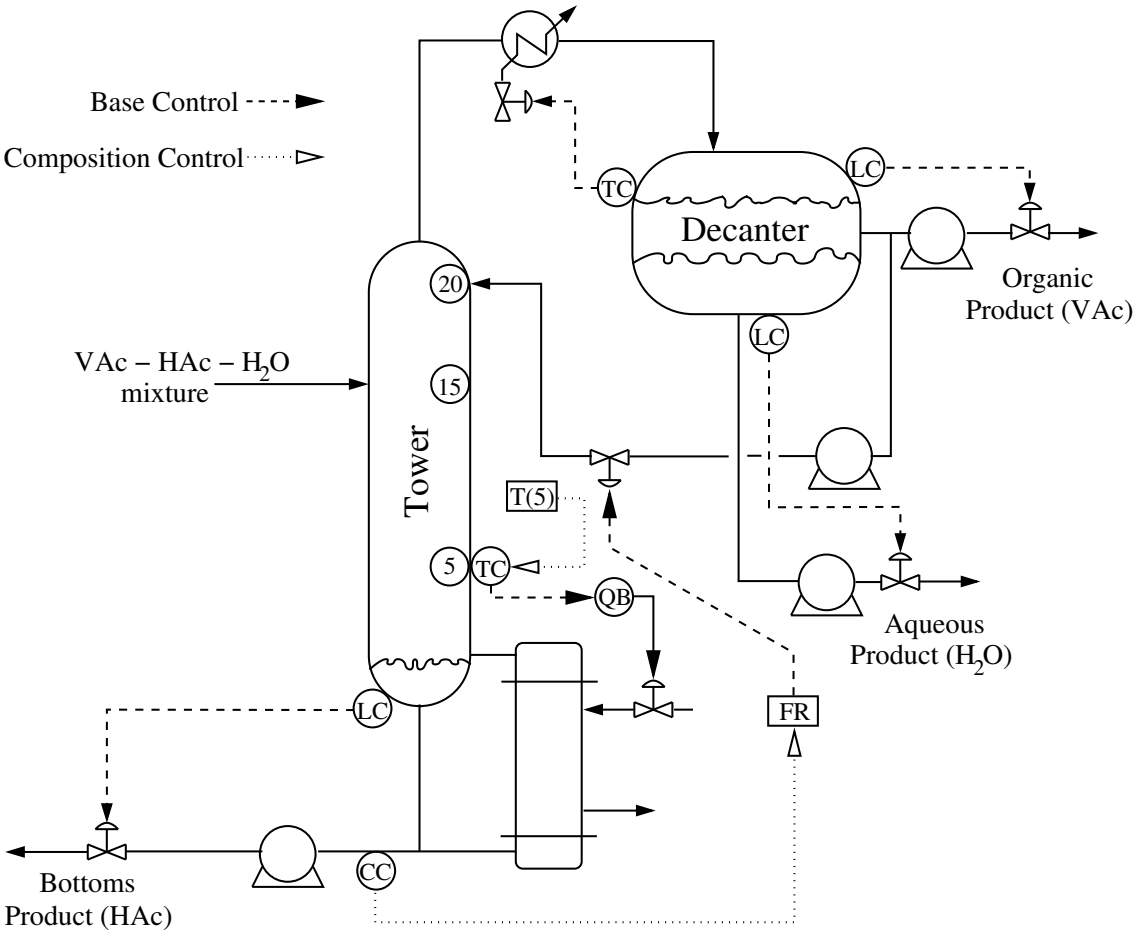

Figure 2.5: The Azeotropic distillation tower of the Vinyl Acetate process with it's base control system.

Table 2.4: Control structure, controller parameters, setpoints and steady state values for Azeotropic tower (AZ).

Loop No.	Controlled variable	setpoint	Manipulated variable	S.S. value	Type	K_c	T_R (min)
1	H_2O in the column bottom ($xB(2)$)	0.0937	column reflux flow rate (FR)	4.985	PI	0.5	60
2	fifth tray temp ($T(5)$)	110^{o}C	reboiler steam valve (QR)	6.726e+004	PI	20	30
3	decanter temp (TD)	45.845^o C	column condenser duty (QC)	6.042e+004	PI	1	5
4	decanter organic level (LO)	50%	organic product flow rate (FO)	0.829	P	1	
5	decanter aqueous level (LA)	50%	aqueous product flow rate (FA)	0.835	P	1	
6	column bottom level (LB)	50%	column bottom flow rate (FB)	2.159	P	1	

Figure (2.6) shows the spectral decompostion of the feed disturbance. It can be seen that the disturbance encompasses a significant range of frequencies. For the two time periods considered, the nature of the disturbance also changes. Figure (2.7) shows the effect of the feed disturbance (a fluctuation in x_F) on the product quality (x_B).

Figure (2.8) shows the response of the quality variable (i.e. mole fraction of water in the column bottoms) to a step change in the two manipulated variables. It can be seen that in both cases the overshoot is much greater than the steday state gain. Moreover, the steday state gain when T5-SP is bumped is almost zero. Hence, owing to the nonlinear thermodynamics (i.e. vapor-liquid equilibria) relating the three components, a one-to-one correlation between tray temperatures and composition does not exist. This behavior is in contrast with most research publication on sift-sensors which deal with binary systems.

2.3 Summary

The test-bed processes used in this study exhibit a wide spectrum of interesting dynamics. The Tennessee Eastman challenge problem (TE), an extensively studied process, is open-loop unstable and could shut-down due to an imbalance in stoichiometry leading to component build up. The Azeotropic distillation tower (AZ) of the Vinyl Acetate process is highly non-linear and exhibits parametric sensitivity to organic reflux flow and feed compositions. As a result controlling the column bottoms and decanter organic and aqueous levels is not enough. The (fifth tray temperature) - (reboiler heat duty) loop is also necessary to prevent the tower from settling to the undesirable steady state.

These peculiarities have to be accounted for by the base control system and this study assumes that such a system, as a lower layer, is already in place. The set-points for

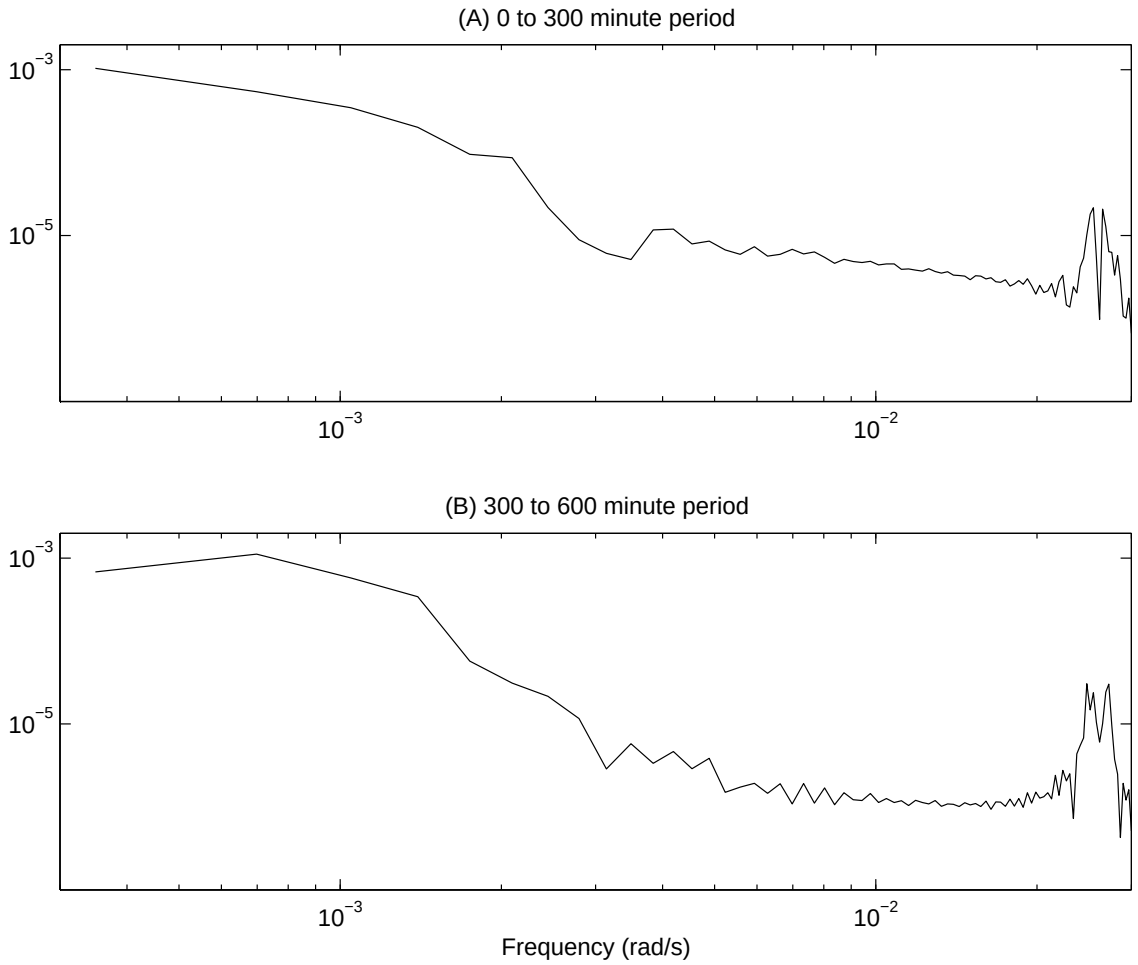

Figure 2.6: Spectral decomposition of the random fluctuation in the composition of the feed to the Azeotropic (AZ) tower.

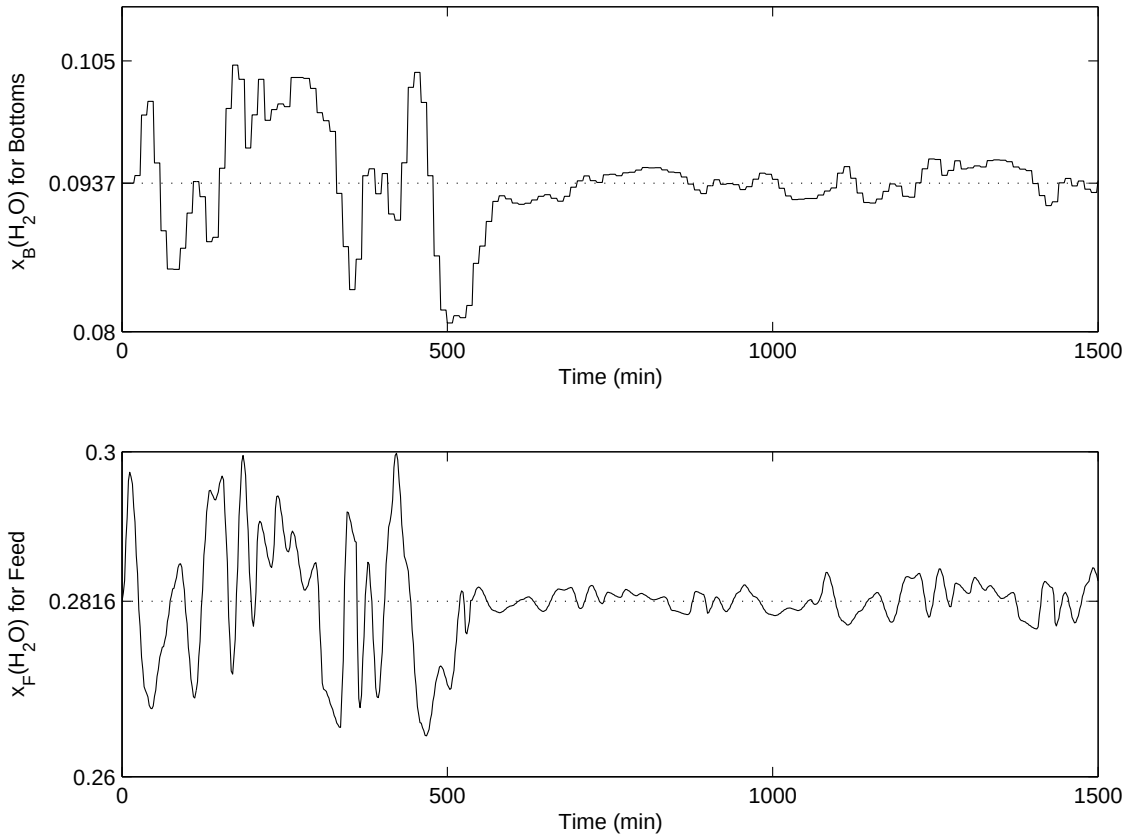

Figure 2.7: Effect of feed disturbance on product quality.

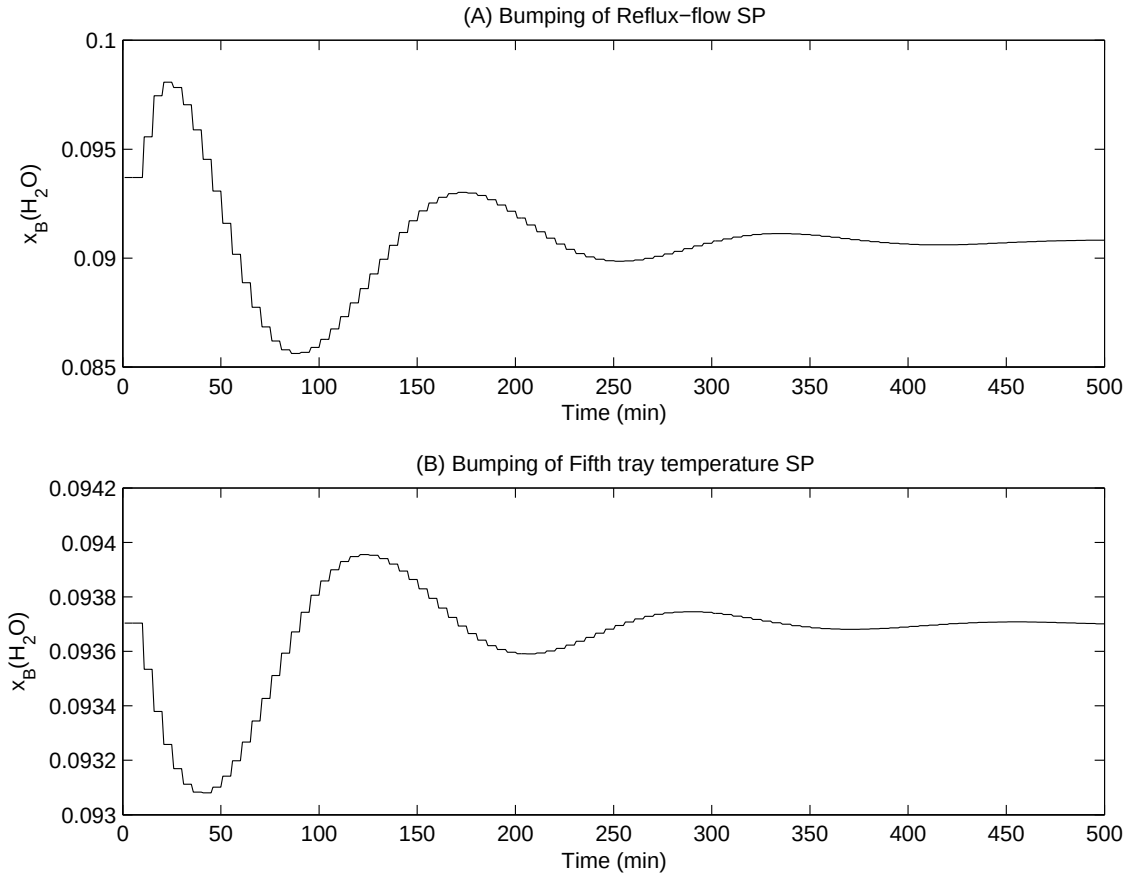

Figure 2.8: Response of $x_B(H_2O)$ (i.e. bottoms product composition) to the bump testing of (A) Reflux-flow SP and (B) Fifth tray temperature SP.

Table 2.5: Overview of test cases used in this study.

		Test case	
		Tennessee Eastman (TE)	Azeotropic (AZ) tower
Quality variable (QV)	label	G/H ratio	$x_B(H_2O)$
	setpoint	1.226	0.0937
	sampling freq. (min)	15	10
	dead time (min)	15	10
Manipulated variables	MV1	D-feed SP	Reflux-flow SP (FR-SP)
	MV2	Reactor-level SP	Fifth tray temperature SP (T5-SP)

the lower level controllers are the manipulated variables for the controllers, being developed in this study, in a cascade arrangement. The lower level controllers ensure basic regulatory performance and stability while the upper layer controllers provide quality control and improved dynamics. If, however, the upper layer controllers are not properly designed, there is a potenial for undesirable effects such as instability.

Both cases provide an opportunity for improved quality control and/or huge savings in energy costs and hence a justification for improved process control. Simulations for both test-bed processes have been benchmarked against dynamics observed in their industrial counterparts.

Chapter 3

Basic controller formulation and performance evaluations

In this chapter, the basic formulation for inferential control, termed as "Model predictive statistical process control" (MP-SPC), is described first. Controller performance is analyzed in detail and results for stationary and non-stationary disturbances using simulations for the two test case processes (the ones described in Chapter 2) are presented next. Finally, controller tuning strategies for optimal and robust performance are suggested.

3.1 Basic controller formulation (MP-SPC)

This formulation is essentially an extension of McAvoy's formulation [31] to multivariable systems, i.e. using multiple manipulated variables to achieve score control. This section primarily discusses the on-line implementation step. The idea is to first make a case for this approach and once enough promise for it's success is put forward, other aspects dealing with the implementation can be taken up.

Consider a plant with it's base control system in place running at steady state. In order to achieve the control objectives, the controllers being developed here, the second-tier in the control system, adjust the set-points of the base controllers. The number of manipulated set-points, i.e. manipulated variables (MVs) available is n_u. For the sake of developing the basic formulation, disturbances are restricted to being stationary and hence the net steady state adjusments on the manipulated variables are zero.

Assume that a finite step response (FSR) model relating the collinear PC scores (denoted by $\mathbf{z}$) to the MV moves is available. It should be noted that the $\mathbf{z}$ variables are

the scores obtained using ordinary PCA and then suitably transformed such that they become collinear with the MVs. Owing to this transformation, the number of collinear score variables is equal to the number of MVs, i.e. n_u. As defined in Equation 1.2, the dynamic matrix ($\mathbf{A_D}$), a matrix of size $P \cdot n_u \times M \cdot n_u$, is constructed using the step response coefficients of the score model. Here, P denotes the prediction horizon and M refers to the control horizon. Future values of $\mathbf{z}$ in response to the future control moves $\mathbf{\Delta u}$ can be predicted using:

$$\mathbf{z} = \mathbf{A_D}\mathbf{\Delta u} \tag{3.1}$$

The $\mathbf{z}$ variables are essentially deviation variables since they have been scaled to zero mean and unit variance. At time $t = k$, the future values of $\mathbf{z}$ can be predicted using:

$$\mathbf{z^f} = \mathbf{\Delta z} + \mathbf{A_D}\mathbf{\Delta u} + \mathbf{z^P} \tag{3.2}$$

Here, $\mathbf{\Delta z}$ is the difference between the measured value and the predicted value of $\mathbf{z}$ at the present time instant, k. $\mathbf{\Delta z}$ is responsible for introducing feedback action into the controller. $\mathbf{z^P}$ reflects the contributions of the past moves of $\mathbf{u}$. The future error if no control action is taken (i.e. for $\mathbf{\Delta u} = 0$) is defined as the difference between the $\mathbf{z}$ set-point and $\mathbf{z^f}$:

$$\mathbf{E_P} = \mathbf{z^{sp}} - \mathbf{z^f} = \mathbf{z^{sp}} - \mathbf{\Delta z} - \mathbf{z^P} \tag{3.3}$$

where $\mathbf{z^{sp}}$, the set-point for $\mathbf{z}$, is 0 for stationary disturbances. McAvoy [31] defined a MP-SPC problem as:

$$\left.\begin{array}{c} \min \\ \Delta u(t), \ldots \Delta u(t+M-1) \end{array} \begin{array}{l} \sum_{l=1}^{P} \|\Gamma[E_l - A_{D,l}\Delta u]\|^2 + \sum_{l=1}^{M} \|\Lambda[\Delta u(t+l-1)]\|^2 \\ \\ \text{subject to: } \sum_{l=1}^{M}[\Delta u(t+l-1)] + \mathcal{U} = 0 \end{array}\right\} \tag{3.4}$$

The minimization of the objective (cost) function is performed at each time step k. The output weights $\mathbf{\Gamma}$ and input weights $\mathbf{\Lambda}$ are usually held constant. The sum of the past moves $\mathcal{U}$ is a column vector of size n_u, i.e.

$$\mathcal{U} = \sum_{\text{past}} \mathbf{\Delta u} = \left[\sum_{i=1}^{k-1} \Delta u_1(i), \sum_{i=1}^{k-1} \Delta u_2(i), \cdots, \sum_{i=1}^{k-1} \Delta u_{n_u}(i)\right]^T \tag{3.5}$$

It can be seen that this formulation is similar to the traditional DMC formulation given by Equation 1.4 with two additional features. First, the score variables are being treated as the controlled variables and their deviations from the origin are being minimized. Second, the equality constraints on the optimizer, i.e. the changes in $\mathbf{u}$, force the net sum of the control moves to be zero. The aim of the control action is to counteract the effect of disturbances and thus maintain the process at steady state. Since the upsets are assumed to be stationary around a steady state, the net change in the control moves should also be zero. This is a key feature of MP-SPC.

Lagrange multipliers (contained in a column vector of size n_u) are used to incorporate the contraints into the objective function and thereby making the optimization problem unconstrained. The augmented Lagrange function, also known as the Lagrangian, is then given by:

$$\left.\begin{array}{rl} \min & \sum_{l=1}^{P} \|\Gamma[(E_l - A_{D,l}\Delta u)]\|^2 + \sum_{l=1}^{M} \|\Lambda[\Delta u(t+l-1)]\|^2 \\ \Delta u(t), \ldots \Delta u(t+M-1), \lambda & \\ & +\lambda(\sum_{l=1}^{M}[\Delta u(t+l-1)] + \mathcal{U}) \end{array}\right\} \tag{3.6}$$

Define the $(M \cdot n_u) \times (M \cdot n_u)$ Hessian matrix ($\mathcal{H}$) as:

$$\mathcal{H} = \mathbf{A_D}^T \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{A_D} + \mathbf{\Lambda}^T \mathbf{\Lambda} \tag{3.7}$$

and the $(M \cdot n_u) \times 1$ gradient vector ($\mathcal{G}$) as:

$$\mathcal{G} = -2\mathbf{A_D}^T \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{E_p} \tag{3.8}$$

The augmented Lagrange function now becomes:

$$\mathcal{L}(\mathbf{\Delta u}, \lambda) = \mathbf{\Delta u}^T \mathcal{H} \mathbf{\Delta u} + \mathcal{G}^T \mathbf{\Delta u} - \lambda^T (\sum_{l=1}^{M} \mathbf{\Delta u} - \mathcal{U}) \tag{3.9}$$

In order to determine the minimum value of the cost function, the partial derivatives with respect to $\mathbf{\Delta u}$ and λ need to be zero. For any vector $\mathbf{x}$ and matrix $\mathbf{A}$:

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \tag{3.10}$$

if $\mathbf{A}$ is symmetric and

$$\frac{\partial \mathbf{A}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A} \tag{3.11}$$

Next an $n_u \times (M \cdot n_u)$ block diagonal matrix is defined:

$$\alpha = \text{diag}(\underbrace{[\overbrace{11\cdots1}^{M}], \cdots, [\overbrace{11\cdots1}^{M}]}_{M \cdot n_u})$$

The minimum of the cost function given in Equation 3.6 can now be obtained through the solution of the following $M \cdot n_u$ equations:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{\Delta u}} = 2\mathcal{H}\mathbf{\Delta u} - \alpha^T \mathbf{\Delta u} + \mathcal{G} = 0$$

and the following n_u eqations:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \alpha \mathbf{\Delta u} - \mathcal{U} = 0$$

These equations have to be solved simultaneously at each time step k. This problem may be represented compactly as an exactly determined problem of solving a set of $((M+1) \cdot n_u) \times 1$ linear equations with $((M+1) \cdot n_u) \times 1$ unknowns (variables):

$$\begin{bmatrix} \mathbf{\Delta u} \\ \lambda \end{bmatrix} = \begin{bmatrix} 2\mathcal{H} & \alpha^T \\ \alpha & \emptyset \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{G} \\ \mathcal{U} \end{bmatrix}$$

Out of the control moves thus calculated, only the ones corresponding to the present time step (k), i.e. the $\mathbf{\Delta u}(k)$ are executed. At the next time step ($k+1$), the same calculation sequence is repeated.

3.2 Analysis of controller performance

3.2.1 Handling of stationary disturbances

In order to gain insight into the performance of the model predictive statistical process controller (MP-SPC) in dealing with stationary disturbances in the feed, consider a sinusoidal disturbance in the C-feed, i.e. stream 4 for the Tennessee Eastman (TE) process and a sinusoidal disturbance in the feed to the column for the Azeotropic (AZ) tower. The frequency and size of these input sinusoids is constant with respect to time. For the case of the Tennessee Eastman (TE) process, the input disturbance was constructed by varying the mole fraction of A in a sinusoidal fashion about it's steady state value of 0.485 and the mole fraction of B about it's steady state value of 0.005. The size of the inputs was 2% of their steady state values and so the maximum value for the mole fraction of A was 0.4947 and that for the mole fraction of B was 0.0051. The mole fraction of C was adjusted in order to ensure that the sum of the mole fractions of the three components equals unity. In the case of the Azeotropic (AZ) tower, the input disturbance was generated by varying the mole fraction of H_2O in the feed to the column in a sinusoidal fashion about it's steady state value of 0.2816. The maximum value of this input was 0.1% of it's steady state value and so the maximum value of $x_F(H_2O)$ was 0.2819. It was necessary to have such a small perturbation in order to minimize the chances of destabilizing the tower. The mole fraction of vinyl acetate was kept constant at 0.2064 and that of acetic acid was adjusted in such a way that the sum of the mole fractions of the three components is 1.0 at all time.

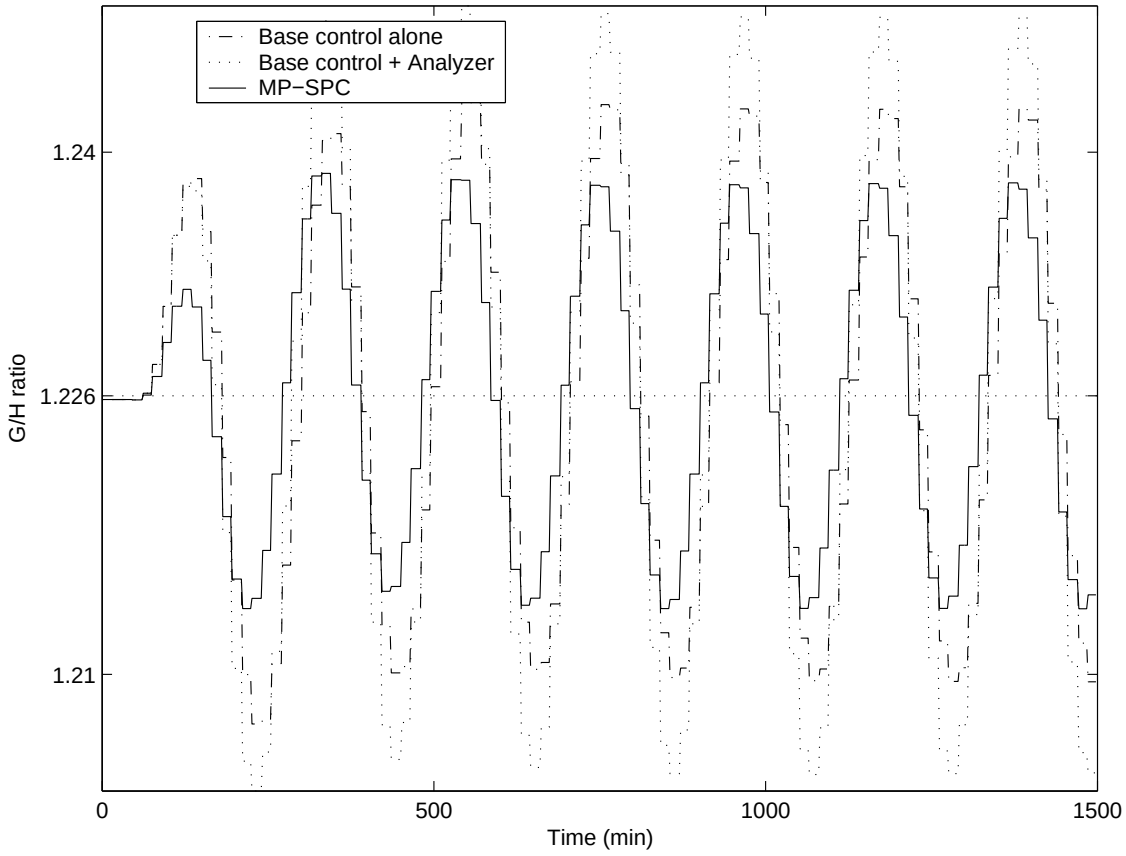

Figure 3.1: Comparative response to a hypothetical sinusoidal disturbance in stream 4 for base control system (with and without an analyzer for the quality variable) and MP-SPC for the Tennessee Eastman (TE) process.

It should be noted that the case of D-feed SP being used as the manipulated variable for the TE process and the case of fifth tray temperature SP being used as the manipulated variable for the AZ tower were chosen as representative configurations for MP-SPC.

For these input disturbances and MP-SPC configurations, it can be seen in Figure 3.1 and Figure 3.2 that, for both cases, there is an initial transient period and then the product quality variable value settles down to a steady sinusoidal oscillation. The stationary value about which the G/H ratio for the TE case oscillates is 1.226. The corresponding value for $x_B(H_2O)$ of the AZ case is 0.0937. Both of these are the target values for the respective quality variables. As shown in Figure 3.1, the maximum value

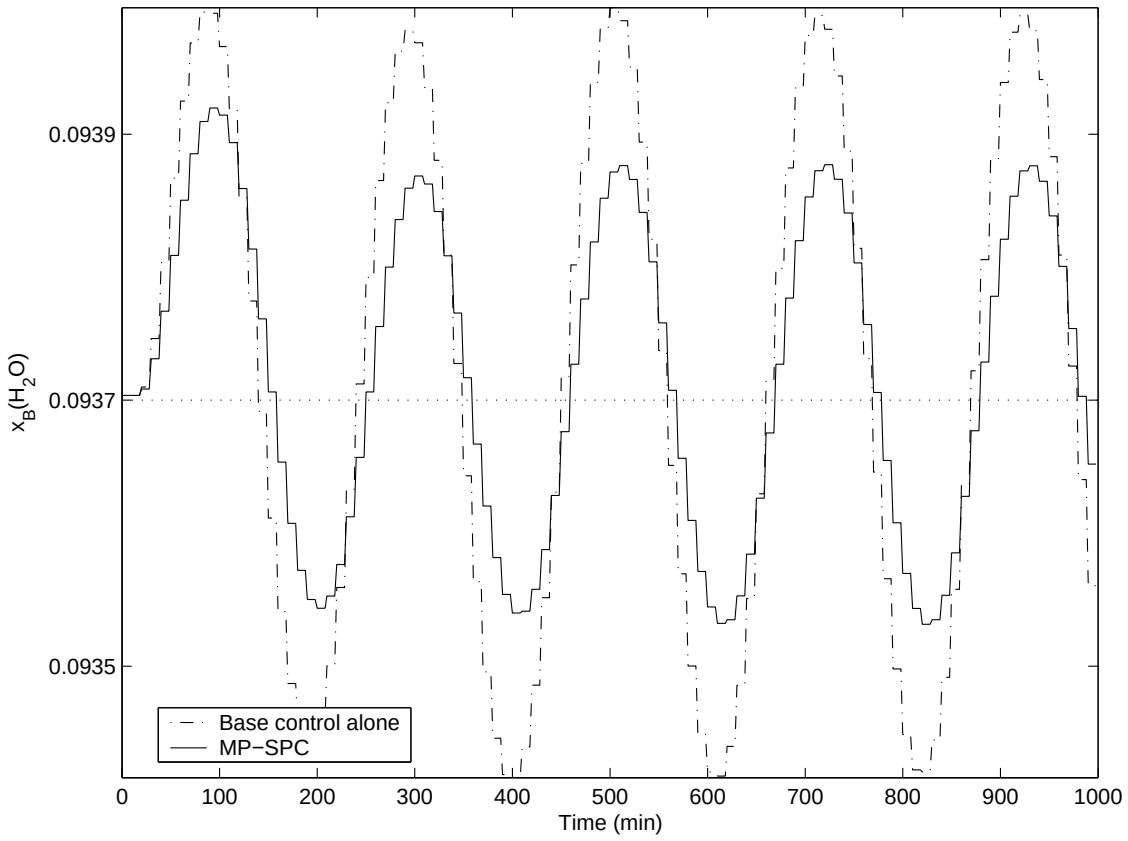

Figure 3.2: Comparative response to a hypothetical sinusoidal disturbance in the feed for base control system alone (i.e. without an analyzer for the quality variable) and MP-SPC for the Azeotropic (AZ) tower.

of the G/H ratio when the Base control system alone is in place is 1.2429 and that when an on-line analyzer for the product quality variable is available alongwith the Base control system is 1.2484. Hence, for this case, not only does the addition of the on-line analyzer not provide any benefit but it worsens the performance of the control system instead. The reason for this detrimental performance is discussed later. For the case when MP-SPC is put on-line, the maximum value of G/H ratio is lowered to 1.2388. It should also be noted that all three responses are in phase and have the same frequency. This means that the reduction in variance has not been achieved as a result of sampling or filtering but because of an attenuation in the amplitude. Figure 3.2 shows how the Base control

system and MP-SPC handle the input disturbanace. The response of the case when the Base control system is augmented with an on-line analyzer for the bottoms composition is not shown because this behavior is identical to that for the Base control system alone. In this case too, when compared with the Base control system alone, MP-SPC reduces the maximum deviation of the quality variable from it's steday state value as evident from the fact that he maximum value of $x_B(H_2O)$ for the Base control system is 0.0940 while that for MP-SPC is 0.0939. Again, both the responses have the same frequency and are in phase and so the reduction in variance is brought about by a smaller amplitude.

These disturbances are obviously hypothetical. However, they are very useful in demonstrating that MP-SPC "shrinks" the size of the deviations from the target values for the quality variable. The underlying mechanism at work whereby a reduction in the variance in the score variables gets translated into a reduction in the variance in the product quality, as observed above, will be discussed later. Instead, the reason why the on-line analyzer for the product quality does not provide any benefit is investigated next.

For the Tennessee Eastman (TE) process, the product (stream 11) is analyzed at a sampling frequency of 0.25 hours and a dead time of 0.25 hours. This means that a new measurement of the G/H ratio is available every 15 minutes and this measurement is 15 minutes old. This delay is quite significant when it is compared with the open-loop time constant (from the D/E set-point to the G/H ratio) of about 110 minutes. As a result the frequency responses (which will be discussed in detail later) of the Base control system alone and the Base control system with the on-line analyzer are quite different. For the case when the on-line analyzer is available, the disturbance hits at or near the ultimate frequency of the loop. The amplitude ratio for frequencies in this range is greater than unity and so the control system magnifies the disturbance instead of shrinking it.

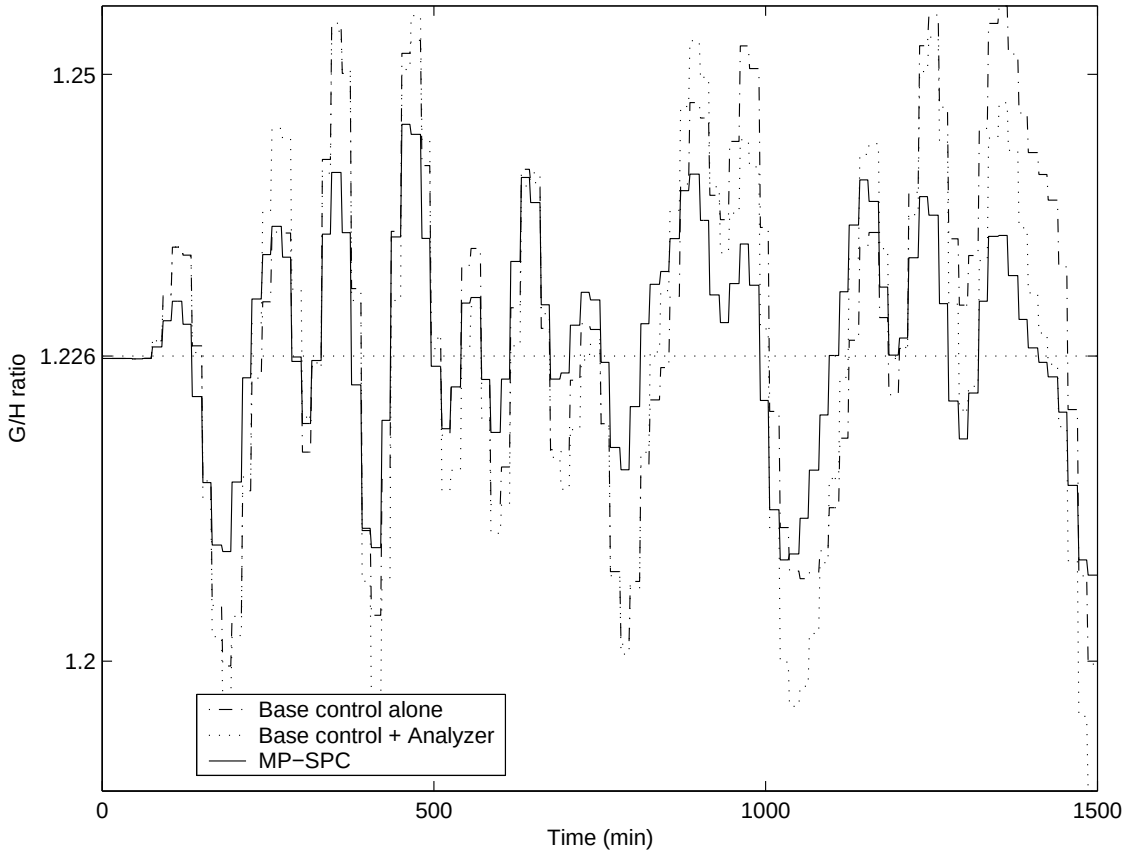

Figure 3.3: Comparative performance of base control system (with or without an on-line analyzer) and MP-SPC for the Tennessee Eastman (TE) process.

For the case of the Azeotropic (AZ) tower, as discussed in section 2.2, the controller gain suggested by Chen et al [7] for the Reflux-flow SP paired with bottoms composition loop is kept very small. As a result, the on-line analyzer has a very slow response and does not impart any advantage in dealing with disturbances with a significant frequency content. Moreover, the magnitude in the frequency response for both the Base control system and the Base control system + on-line analyzer are similar.

In order to observe the performance of MP-SPC in dealing with more realistic disturbances, the effect of a random fluctuation in the C-feed (stream 4) for the Tennessee Eastman (TE) process, termed IDV(8), was simulated for 2500 minutes. A time period of 0 to 1500 minutes is given in Figure (3.3). For the period simulated, it can be seen

that the G/H ratio oscillates around a steady state value of 1.226. The maximum value of the G/H ratio is 1.2457, the minimum is 1.2073 and the variance is 8.0169×10^{-5}. Compared to the base control system (i.e. constant D/E policy), this represents a 61% reduction in the product quality variability. Hence, significant improvements have been achieved without actually measuring either the G/H ratio in the product stream or the disturbance i.e. composition of the C-feed. The quantity that affects the G/H ratio in the product is the D/E ratio in the feed to the reactor. IDV(8) does not upset the D/E ratio in the feed or the G/H ratio in the product directly. However, this disturbance enters the process through stream 4 and upsets other process variables which in turn upset the D/E and G/H ratio, i.e. there is an indirect effect. Most of the process variables which are directly affected by IDV(8) can be controlled using the Base control system alone but the variability in the G/H ratio cannot be efficiently handled because the Base control system does have a direct effect on the G/H ratio. MP-SPC has a direct affect on the G/H ratio because it manipulates the D-feed flow which is essentially changing the D/E ratio. Owing to this direct effect, MP-SPC performs better than the Base control system alone.

Besides the D-feed SP, the performance of another manipulated variable (MV), the Reactor-level SP was also evaluated. These two MVs were the two univariate (i.e. single MV) configurations examined while evaluating MP-SPC for the Tennessee Eastman (TE) process. In order to study a multivariate configuration, the above two MVs were used in combination. The performace results for the three configurations are summarized in Table 3.1 where the variance of the G/H ratio in the product stream for the time periods 0 to 1250 min and 1250 to 2500 min are given. It can be seen that when compared to the Base control system alone, the % improvement provided by R-level SP is lower (about

13%) than that provided by D-feed SP (about 60%). The reason why R-level SP is not as effective an MV as D-feed SP is that the score variable exhibits an inverse response to the MV moves when R-level SP is being used. Inverse response significantly retards the performance of MP-SPC and so the improvements provided by using both the MVs together is also not significantly greater (about 75%) than that provided by the univariate configuration involving D-feed SP.

As regards the Azeotropic (AZ) tower, the response to the random fluctuation in the feed to the column (the disturbance described in section 2.2) was simulated for 2500 minutes. A time period of 0 to 800 minutes is shown in Figure (3.4). For the period simulated, it can be seen that the mole fraction of water ($x_B(H_2O)$) fluctuates around a steady state value of 0.0937. With the fifth tray temperature SP (T5-SP) used as the manipulated variable for MP-SPC, the maximum value of $x_B(H_2O)$ is 0.1032, the minimum value is 0.0841 and the variance is 1.5447×10^{-5}. This reduction in variablity represents an improvement of about 49% over the case when only the Base control system is being used. Comparing the minimum $x_B(H_2O)$ value of 0.0797 for the Base control system to 0.0841 for MP-SPC, it might seem that MP-SPC does not offer any significant advantages. However, it should be noted that the difference in the minimum values (0.0841-0.0797 = 0.0044) as a consequence of improved control can be taken advantage of by moving the set-point for $x_B(H_2O)$ closer to the constraint, As a result of this new set-point, the material and energy balance are recalculated and the new operatng conditions would result in substantial savings in energy and other operating costs.

Table 3.1: Performance of Base control, Base control + analyzer and MP-SPC for the Tennessee Eastman (TE) process in dealing with IDV(8) as quantifed through the variance in G/H ratio. % improvements with respect to the Base control system is expressed within brackets.

Formulation	Time period (min)	D-feed SP as MV	R-level SP as MV	Both
Base control	1:1250	1.8438×10^{-4}		
	1251:2500	4.0298×10^{-4}		
Base control + analyzer	1:1250	2.3349×10^{-4}		
	1251:2500	4.5638×10^{-4}		
MP-SPC	1:1250	7.9556×10^{-5} (56.8%)	1.6087×10^{-4} (12.7%)	3.8935×10^{-5} (78.9%)
	1251:2500	1.4816×10^{-4} (63.2%)	3.4638×10^{-4} (14.0%)	1.0572×10^{-4} (73.8%)

Table 3.2: Performance of Base control, Base control + analyzer and MP-SPC for the Azeotropic (AZ) tower in dealing with feed disturbances as quantifed through the variance in $x_B(H_2O)$. % improvements with respect to the Base control system is expressed within brackets.

Formulation	Time period (min)	FR-SP as MV	T5-SP as MV	Both
Base control	1:1250	2.3554×10^{-5}		
	1251:2500	3.6731×10^{-5}		
Base control + analyzer	1:1250	2.3554×10^{-5}		
	1251:2500	3.6732×10^{-5}		
MP-SPC	1:1250	2.2810×10^{-5} (3.2%)	1.1173×10^{-5} (52.5%)	1.2059×10^{-5} (48.8%)
	1251:2500	3.5978×10^{-5} (2.0%)	1.9514×10^{-5} (46.9%)	2.1133×10^{-5} (42.5%)

Hence, significant improvements have been achieved without actually measuring either the product quality (i.e. $x_B(H_2O)$) or the disturbance (i.e. the composition of the feed, x_F). It should be noted that the disturbance, i.e. the fluctuation in the feed composition has a direct effect on the product composition. This direct influence is owing to the material balance between the input and output in the absence of any reaction or net accumulation in the column.

The performance of a second manipulated variable (MV) for MP-SPC of the Azeotropic (AZ) tower, the Reflux-flow SP (FR-SP), has also been examined. The two MVs in combination, i.e. the fifth tray temperature SP (T5-SP) and the the Reflux-flow SP (FR-SP), have been used to form the multivariate configuration. The performance of these three configurations in dealing with the random disturbance in the feed is summarized in Table 3.2. In this table, the variance of $x_B(H_2O)$ for the time periods 0 to 1250 min. and 1250 to 2500 min. are given. Compared to the Base control system alone, the % improvement provided by Reflux-flow SP (FR-SP) is significantly lower (about 3%) than that provided by the fifth tray temperature SP (about 50%). The reason why FR-SP is not as effective an MV as T5-SP is because of the much larger time lag between the Reflux flow valve and point at which the bottoms composition is measured. The fifth tray tempaerature (T5-SP) operates via the Reboiler heat duty valve which is very close to the bottoms flow whereas there are 20 trays separating the Reflux flow valve and the bottoms flow. In both cases, the lag between the point at which the disturbance enters the system (i.e. the feed) and the approximate location at which the scores are "measured" is the same. The lag between the point at which the scores are evaluated and the bottoms flow is also about the same. The principal difference between the two configurations is the physical location of the two MVs and the resulting lags associated with these locations. Another feature

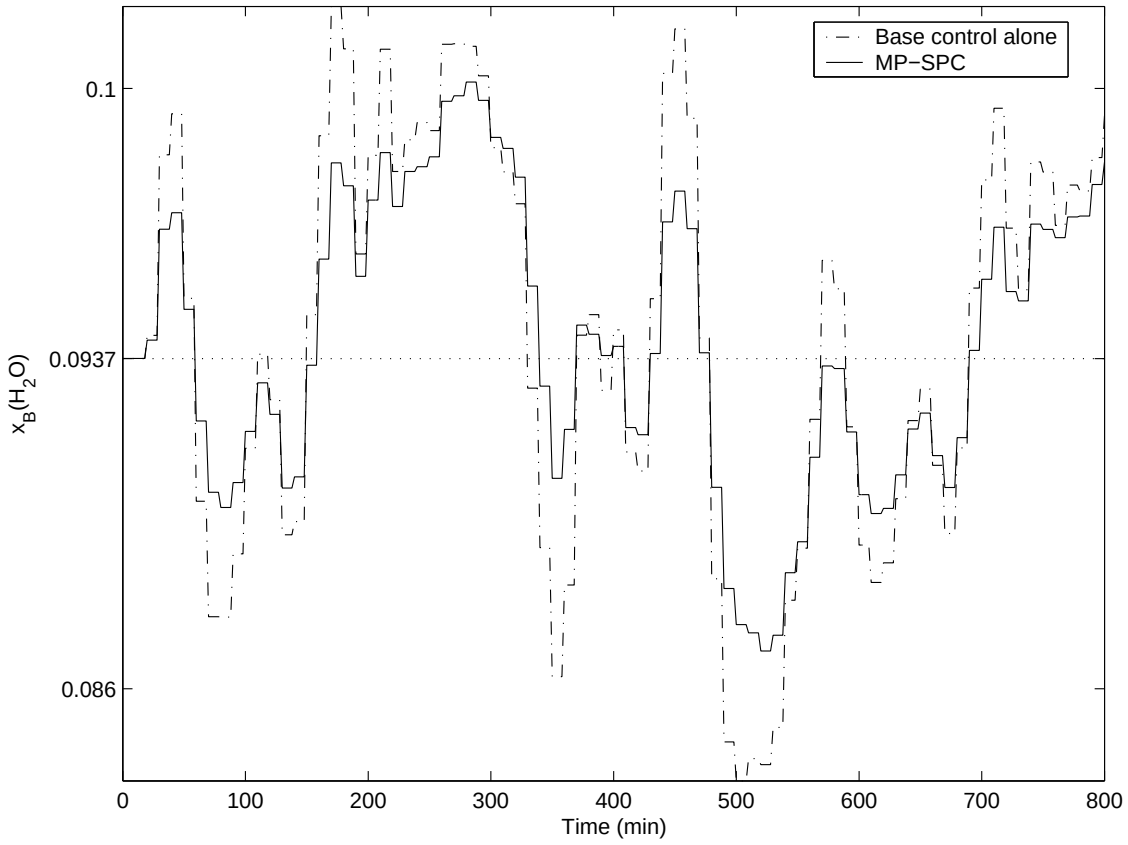

Figure 3.4: Comparative performance of base control system alone (i.e. without an on-line analyzer) and MP-SPC for the Azeotropic (AZ) tower.

worth noting for this case study is that the performance of the multivariate configuration is slightly worse than that of the univariate case. It can be seen in Table 3.2 that the % improvement provided when both FR-SP and T5-SP are used as MVs is about 46% but that provided when only T5-SP is used is about 50%. The reason for this is that the multivariate model is not very accurate. The benefit provided by the extra MV is overshadowed by a larger model-plant mismatch. Hence, it is more advantageous to use a univariate formulation rather than a multivariate one when the extra MV does not provide any significant benefit on it's own.

The mechanism whereby MP-SPC achieves improved quality control is explained using Figure 3.5 and Figure 3.6. It can be seen that the shape of the response of the score variable (z) is similar to that of the quality variable, (i.e. G/H or $x_B(H_2O)$ as the case may be). The peaks (maxima) and valleys (minima) in the top curve lead the corresponding ones in the bottom curve by about 50 minutes in the case of the Tennessee Eastman (TE) process and by about 18 minutes in the case of Azeotropic (AZ) tower. Hence, not only does the z variable act as an inferential measurement of the quality variable but it also provides a significant lead time. This lead time is the reason why MP-SPC does better than the case when an on-line analyzer is available. Another way of interpreting the relationship between the z variable and the quality variable is that the disturbance is first picked up by the z variable and then it travels though the process and shows up at the analyzer after a time lag that can be attributed to the process and analyzer dead time. Without this time lag, the performance of the Base control system + Analyser would have been equivalent to that of MP-SPC. It should be noted that the foregoing analysis using Figure (3.5) and Figure (3.6) was based upon simulations and not on real plant operation data. Time series data from a real plant would be a little more difficult to interpret since

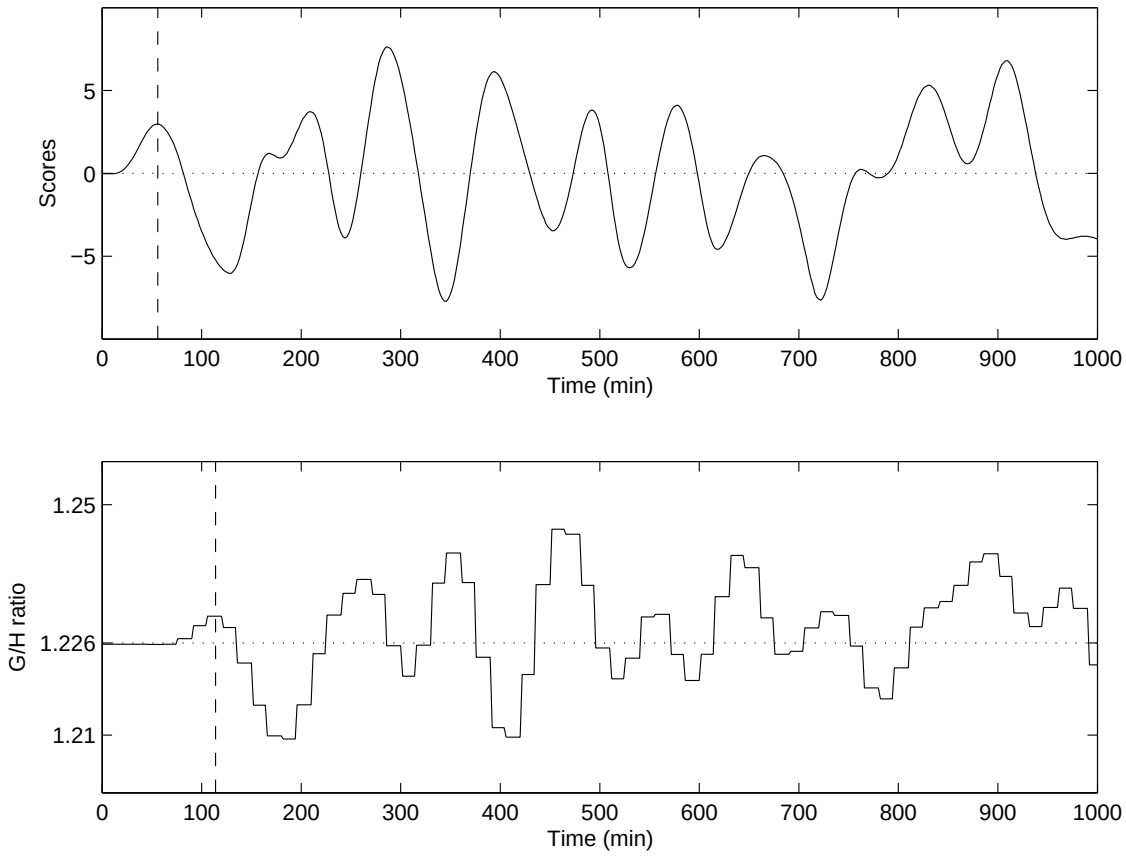

Figure 3.5: Response of quality variable and score variable for the Tennessee Eastman (TE) process.

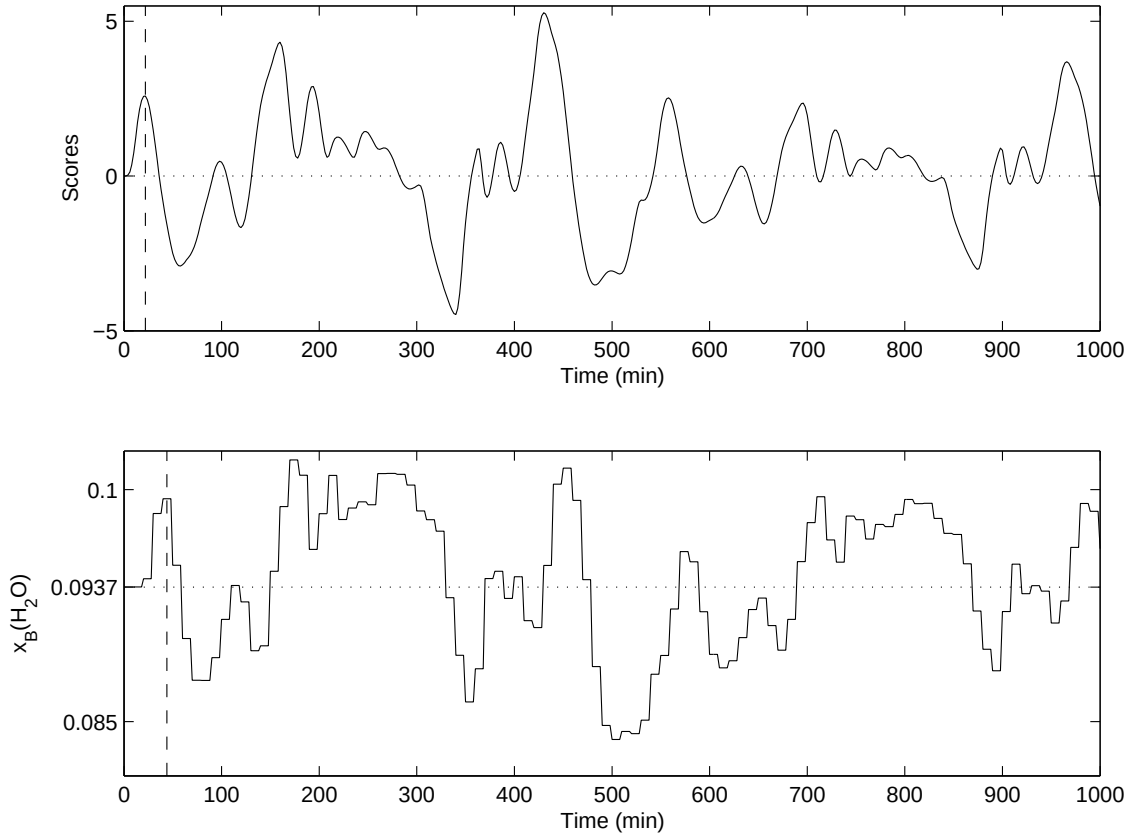

Figure 3.6: Response of quality variable and score variable for the Azeotropic (AZ) tower.

the start time and the steady state are not well defined. In such a situation, correlation analysis should be carried out and this topic is discussed next.

3.2.2 Effect of feedback on cross-correlation

A key assumption of the MP-SPC methodology is that a reduction in the variability of the scores translates into a reduction in the variability of the quality variable. For this assumption to be valid, the cross-correlation between the scores and the quality variable needs to be consistent across the different types of forcing the process and the model encounters, i.e. during model identification, in the presence of disturbances and in the presence of feedback control action. In this study, this consistency is examined as follows. The cross-correlation coefficient between the score variable and the quality variable is calculated at different time shifts (τ). This function, referred to as the cross-correlation function ($r_{z,q,\tau}$), is evaluated before and after MP-SPC is put on line. A significant change in the $r_{z,q,\tau}$ versus τ plot would mean that the correlation structure has been changed due to the feedback control action of MP-SPC. This change in correlation structure could hamper the performance of MP-SPC. An extreme case of alteration in the correlation structure would involve a change in the sign of the cross-correlation coefficient. In such a situation, a key assumption of the MP-SPC methodology: a reduction in the variability of the scores translates into a reduction in the variability of the quality variable, would no longer be valid. As a result, MP-SPC would reduce the variability of the score variable but the variability of the quality variable would increase. It is also possible that the lead time betwwen the score variable and the quality variable gets affected by the feedback control action. If the maximum value of $r_{z,q,\tau}$ occurs at a negative time shift, it would mean that the score variable does not lead the quality variable. Hence, it is important

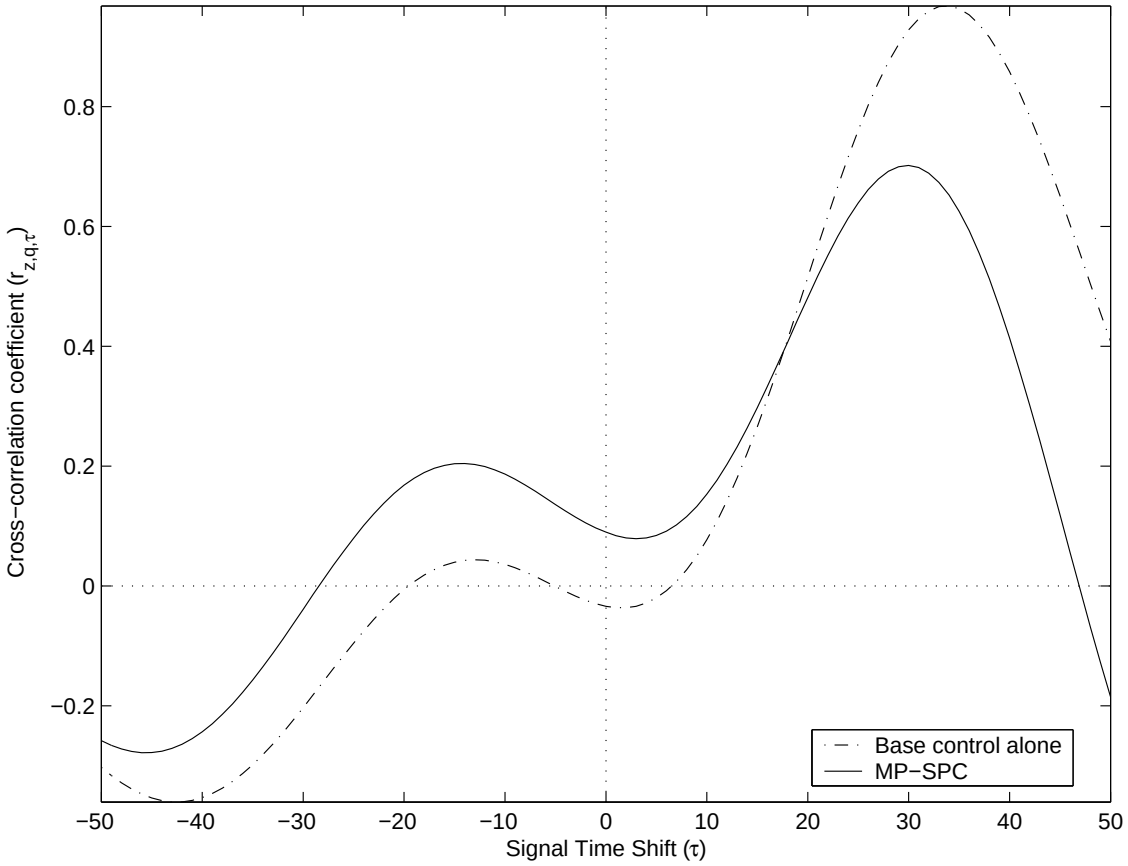

Figure 3.7: The effect of feedback on the cross-correlation between the score variable and the quality variable for the Tennessee Eastman (TE) process.

to evaluate the cross-correlation coefficient at negative time shifts in order to ensure that $r_{z,q,\tau}$ is small for $\tau < 0$.

Figure (3.7) shows the effect of putting MP-SPC on line on the cross-correlation between the score variable (z) and the quality variable (G/H) for the Tennessee Eastman (TE) process. This particular plot pertains to the case of the D-feed SP being used as the manipulated variable to tackle IDV(8). However, similar trends are observed when the Reactor-level SP is used as the MV or when both MVs are used together.

It can be seen that the cross-correlation functions are almost identical and so it can be concluded that the cross-correlation is not influenced by feedback effects i.e. the

score variable and the quality variable are consistently correlated. It can be seen that for negative time shift values, the score and the quality variable are negatively correlated but the cross-correlation is very small (the maximum value being about -0.3). This means that for all practical purposes, at any time instant, the future values of the score variable are independent of the present value of the quality variable. However, for positive time shifts, the cross-correlation function for the case involving the Base control system alone achieves a peak value of about 0.9 at a time shift of about 32 time units. Since the sampling time is 2 minutes, 32 time units correspond to 64 minutes. Hence, it can be concluded that the score variable leads the quality variable by about 64 minutes and that the two have a strong positive correlation between them. Feedback reduces the strength of the cross-correlation but not significantly.

In a similar fashion, Figure (3.8) shows how the score and the quality variables are related in the case of the Azeotropic (AZ) tower. For illustrative purposes, the case where the fifth tray temperature SP (T5-SP) is manipulated to deal with feed disturbances is shown. It should be noted that, identical trends are observed when the Reflux flow rate SP (FR-SP) is used as the MV or both MVs are used together.

Again, for negative time shifts the maximum value of the cross-correlation is only about -0.4 and so it can be concluded that future values of the score variable cannot be predicted from the present values of the quality variable. Moreover, the maximum value of the cross-correlation is about 0.8 when only the Base control system is in place. The peak occurs at a time shift of about 10 time units and for a sampling time of 2 minutes, 10 time units correspond to 20 minutes. Hence quality variable values upto 20 minutes into the future can be inferred from the present values of the score variable. Putting the MP-SPC on line alters the correlation structure to a very small extent.

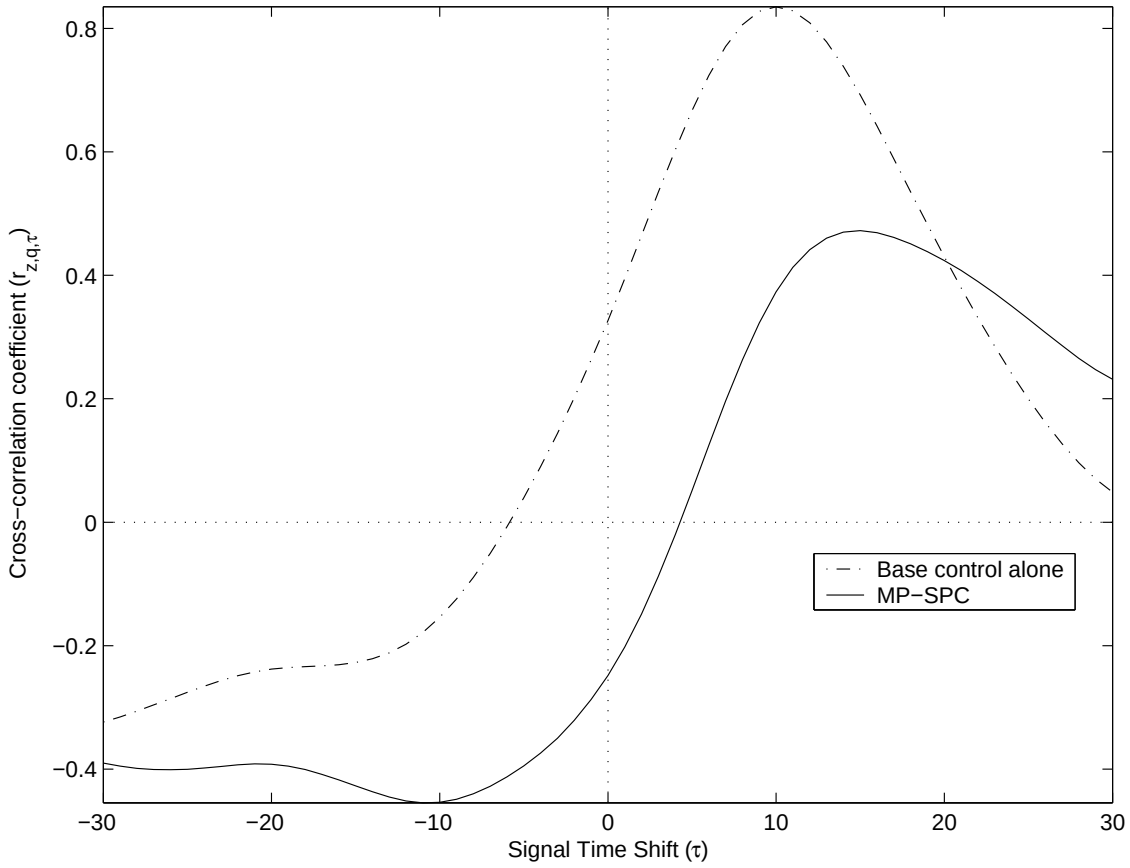

Figure 3.8: The effect of feedback on the cross-correlation between the score variable and the quality variable for the Azeotropic (AZ) tower.

Hence, a comparison of the cross-correlation functions before and after putting MP-SPC on-line confirms the consistency of the correlation between the score and the quality variable. Moreover, this analysis also confirms the previous conclusions that the score variables can be used to infer the future values of the quality variables and that the scores provide a sufficiently large lead time (about 50 minutes for the Tennessee Eastman (TE) process and about 18 minutes for the Azeotropic tower) which should be very advantageous during feedback control.

3.2.3 Handling of non-stationary disturbances

In the foregoing discussion, it was assumed that the random upsets encountered are stationary. Next, the performance of MP-SPC in the case of non-stationary disturbances is evaluated. For the Tennessee Eastman (TE) process, consider a step disturbance in the composition of the C-feed stream over and above the random disturbance. In other words, this is the simultaneous occurance of IDV(1) (i.e. step change in stream 4: A/C feed ratio changes while composition of B is constant) and IDV(8) (i.e. random variation in composition of A, B, C in C-feed i.e. stream 4) described in Downs and Vogel [12]. This step disturbance occurs after 10 hours (600 mintes) of operation. Figure 3.9 shows the comparative performance of the MP-SPC and that of the base control system alone (i.e. using a fixed D/E ratio) or with an on-line analyzer. It can be seen that MP-SPC continues to perform better than the base control system with or without the on-line analyzer. Moreover, in contrast to McAvoy's [31] results, there is no offset. The reason why McAvoy [31] had offsets in the response of MP-SPC when dealing with IDV(1) was because of the way in which the model was identified. The model he used was identified in a PRBS experiment in the presence of IDV(8). So the signal to noise ratio was extremely

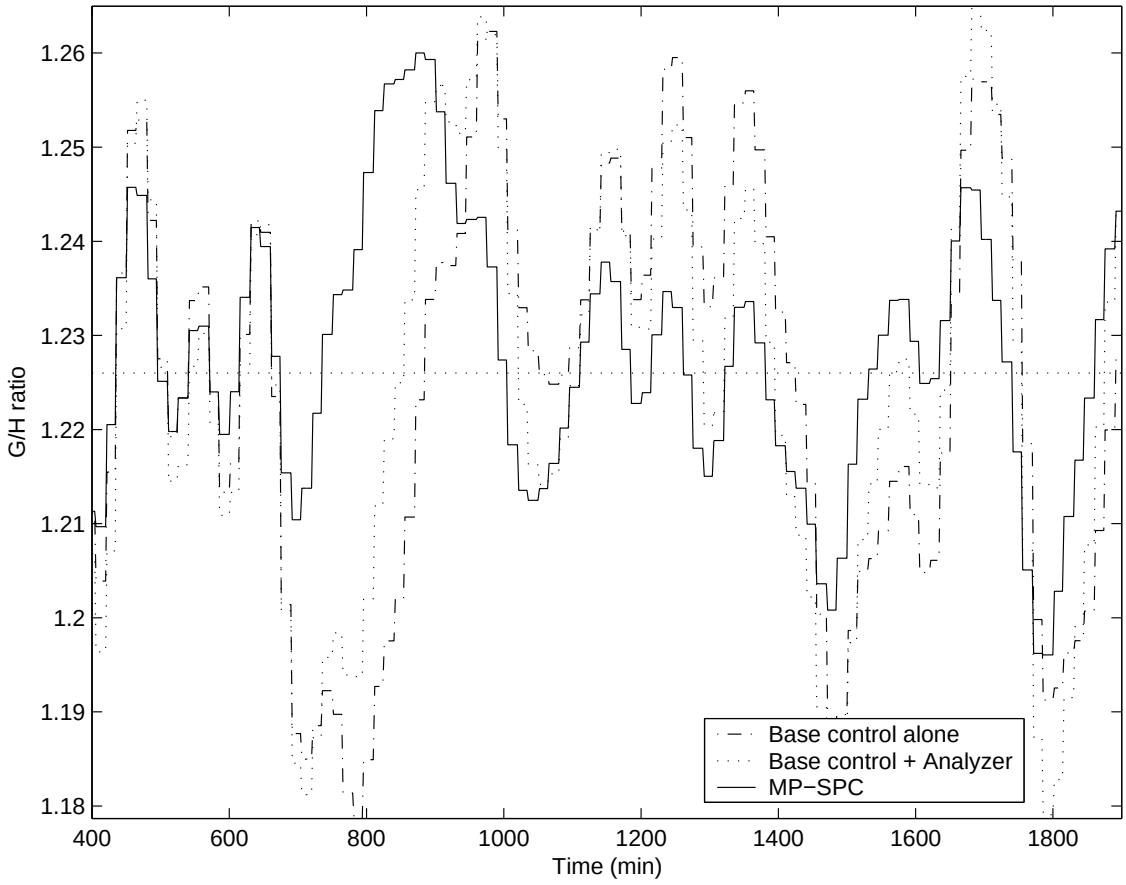

Figure 3.9: Comparative performance of base control system (with and without an on-line analyzer) and MP-SPC in dealing with a non-stationary disturbance for TE.

poor and the resulting model primarily incorporated the effect of the disturbance on the score variable rather than the effect of the MV. When MP-SPC encountered a new disturbance, i.e. IDV(1) + IDV(8) instead of IDV(8) alone, the model was no longer valid. In other words, the cross-correlation between the score variable and the quality variable was inconsistent. Hence, the model-plant mismatch was greater than 100% and this sign change in gain was responsible for the offset.

As mentioned before, neither IDV(1) nor IDV(8) upset the D/E ratio directly. The effect of IDV(1) and IDV(8) is on other process variables and this indirectly affects the G/H ratio in the product stream. This is the reason why a non-stationary disturbance

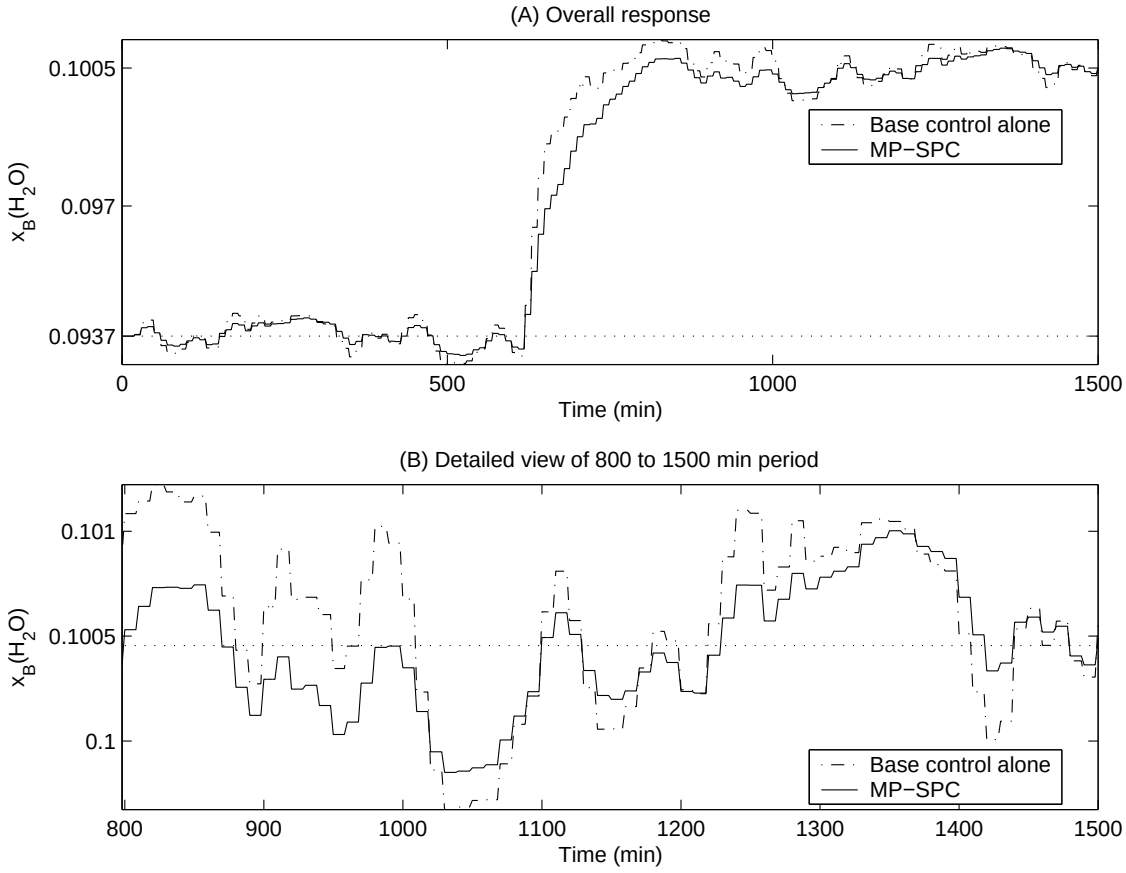

Figure 3.10: Comparative performance of base control system alone (i.e. without an on-line analyzer) and MP-SPC in dealing with a non-stationary disturbance for Azeotropic (AZ) tower.

in the feed composition does not result in a step response in the product composition but only in a larger than normal deviation from the steady state values. The process variables which are directly affected by IDV(1) and IDV(8) can be controlled using the Base control system alone but the variability in the G/H ratio cannot be efficiently handled.

Next consider a step disturbance over and above the stationary disturbance in the feed to the AZ tower. Figure (3.10) shows that for all three controller formulations, the column reaches a new steady state of 0.1005 and oscillates about this new steady state. Even at this new steady state, MP-SPC performs the best. However, MP-SPC is unable to drive the column back to it's original steady state of 0.0937. From the above discussion it can be concluded that MP-SPC is able to provide a substantial reduction in variability in the quality variable in the presence of both stationary and non-stationary disturbances.

3.2.4 Sensitivity interpretion of the Lagrange multipliers

The optimization problem being solved in MP-SPC is a linear program with a constraint on the manipulated variable moves such that the net sum of the moves is zero. This constraint was incorporated into the objective function using Lagrange multipliers (denoted as λ_i). These Lagrange multipliers can be interpreted to reveal certain characteristics of the process. Consider the top portion of Figure (3.11). It shows how the Lagrange multipliers associated with the two manipulated, D-feed SP and Reactor-level SP for the Tennessee Eastman (TE) process evolve with time in the presence of IDV(8). It is known that the constraints with the largest absolute values of the Lagrange multiplier are the ones whose right hand sides affect the optimal objective function value the most. Since both the manipulated variables and the score variables have been scaled to zero mean and unit variance, one does not have to account for any difference in units. It can be seen

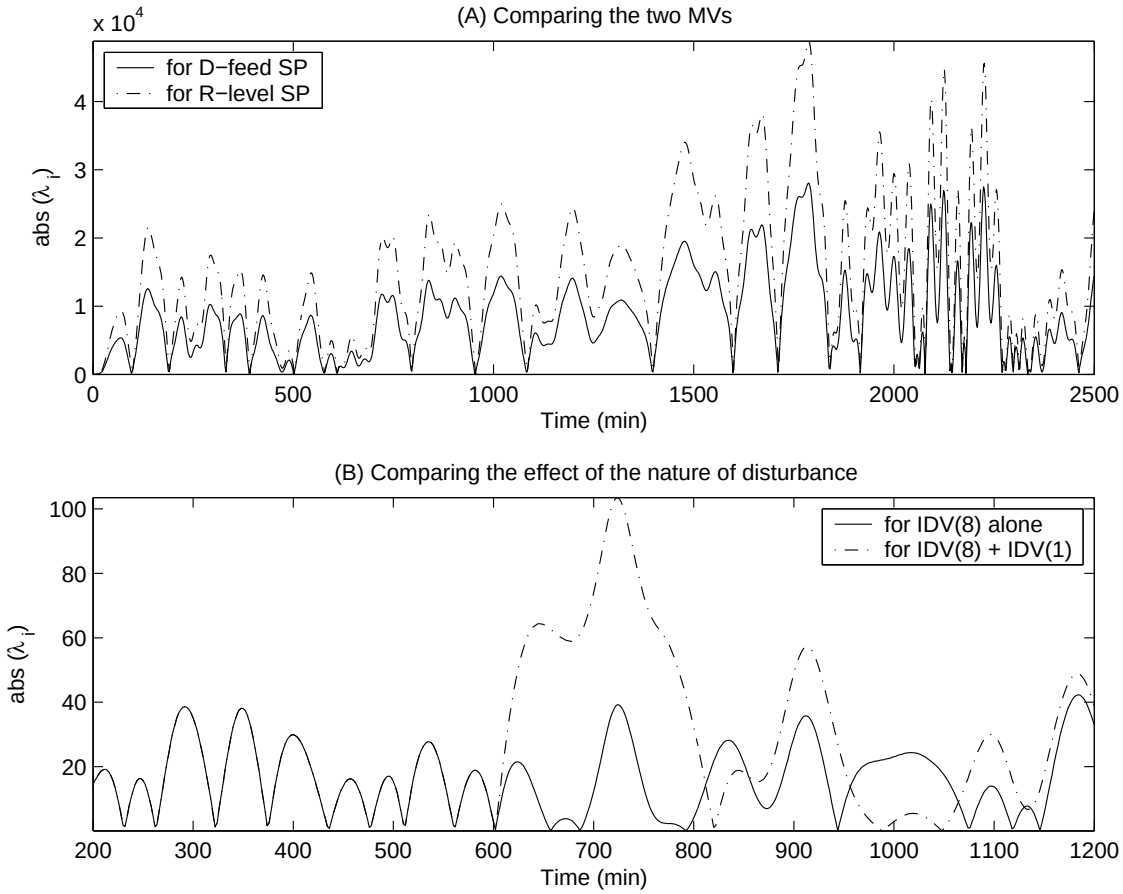

Figure 3.11: Sensitivity interpretion of the Lagrange multipliers for the Tennessee Eastman (TE) case: (A Comparing the two manipulated variables and (B) Comparing the effect of the nature of disturbance.

that the Lagrange multiplier value associated with D-feed SP is consistently smaller than that associated with the Reactor-level SP. Hence, compared to the equality constraint associated with the D-feed SP, the equality constraint associated with the Reactor-level SP has the more dominant effect on the objective function value.

Figure 3.11 (B) compares the time evolution of the Lagrange multiplier for the Tennessee Eastman (TE) process when a single manipulated variable is used to deal with disturbances with or without a non-stationary component. In both cases, the D-feed SP is used as the MV and an identical stationary disturbance, IDV(8), is present. However, in one case, a step disturbance enters the process after 10 hours (600 minutes) of operation. It can be seen that for the case in which there is a non-stationary component in the disturbance, the Lagrange multiplier values assume much larger values, at times 2 or 3 times the values, as compared to the case where there is no significant non-stationary component in the disturbance. Hence, larger than normal values of the Lagrange multiplier(s) can be used as an indicator of the presence of a non-stationary distrubance.

3.3 Controller tuning

Controller tuning involves deciding the most appropriate parameters that lead to robust and optimal performance. As a result, this aspect is very critical for a successful controller implementation. For MP-SPC, the available tuning parameters are:

- the Control horizon i.e. number of control moves (M)
- the Prediction horizon (P)
- the weight on the input i.e. manipulated variable moves ($\mathbf{\Lambda}$), and
- the weight on the output errors i.e. the difference between the actual and target

score values ($\mathbf{\Gamma}$).

Zafiriou [58] has elucidated the following tuning guidelines for linear, unconstrained model predictive controllers (MPCs):

1. The control action becomes more aggressive as

   - P is decreased
   - M is increased
   - $\mathbf{\Lambda}$ is decreased

   This is the general trend with respect to P and M. Although this trend is helpful on most ocassions, sometimes P and M do not have a significant effect on the controller performance at all. $\mathbf{\Lambda}$ is usually the main tuning parameter.

2. As the control action becomes more aggressive (i.e. the system response becomes faster), there is a greater risk of instability, irrespective of the extent of model-plant mismatch.

3. $P = M$ values should be avoided since these will lead to "perfect control".

It should be noted that for univariate cases, i.e. when only one manipulated variable is being used, the effect of increasing the input weight ($\mathbf{\Lambda}$) has exactly the same effect as that of decreasing the output weight ($\mathbf{\Gamma}$) by a proportional amount. In other words, there is a loss of degree of freedom in choosing the tuning parameters. Keeping this in mind when tuning univariate cases, the output weights should be kept fixed (for example at unity). Another point that is worth mentioning is that since the control horizon (M) and the prediction horizon (P) are discrete variables, if an optimzation routine is used to tune the controllers, a mixed-integer program would be necessary. The function relating

controller performance to it's tuning parameters is usually non-linear and quite often non-smooth with multiple local optima. Since under these circumstances the optimzation calculations would be very cumbersome, a trial-and-error type, line search procedure is usually adopted.

Although the above guidelines are meant primarily for linear, unconstrained MPC, they are applicable to MP-SPC too. However, it should be investigated whether the following two unique features of MP-SPC pose any extra challenges. First, there is an equality constraint on the sum of the manipulated variable moves. Second, the primary goal of MP-SPC is to reduce the variability caused by random stationary disturbances with a significant frequency content (as opposed to set-point changes or load distrubances that are steps).

The strategy adopted to tune the MP-SPC controllers is as follows. First the weight on the output errors, $\mathbf{\Gamma}$ is kept fixed (at unity for univariate cases). Next, "reasonable" initial guesses for the control horizon (M) and prediction horizon (P) are used. A good starting value for M is 4 and that for P is about a third of the truncation number of the model being used (i.e. $n_t/3$). Next, starting with a very large value as an initial guess for the weight on the control moves ($\mathbf{\Lambda}$), it's value is gradually decreased. The controller becomes more aggressive and it's performance progressively improves as $\mathbf{\Lambda}$ is decreased until a certain optimum is reached. Beyond this optimum, controller performance begins to deteriorate. For this optimum value of $\mathbf{\Lambda}$, M and P values are adjusted and a new set of optimal M and P values is determined. The procedure is repeated until an optimal set of $\mathbf{\Lambda}$, M and P values is obtained iteratively. For univariate cases, $\mathbf{\Gamma}$ is kept fixed at it's initial value. However, for multivariate cases, once optimal $\mathbf{\Lambda}$, M and P values are determied through an inner iteration, $\mathbf{\Gamma}$ is changed in an outer iteration. It should be noted that when

Table 3.3: MP-SPC tuning parameters for the Tennessee Eastman (TE) process and the Azeotropic (AZ) tower cases.

Tuning parameters ⇒		Control horizon (M)	Prediction horizon (P)	Output weights ($\mathbf{\Gamma}$)	Input weights ($\mathbf{\Lambda}$)
Case	MV				
Tennessee Eastman (TE) process	D-feed SP	4	15	1	0.8
	R-level SP	2	14	1	0
	Both	5	9	[1 112]	[16 6]
Azeotropic (AZ) tower	FR-SP	12	175	1	1.5
	T5-SP	3	190	1	0.12
	Both	3	35	[87 1]	[160 98]

two manipulated variables are being used, the size of the $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ vectors is 2×1. There is a loss of one degree of freedom and so only one element of $\mathbf{\Gamma}$ has to be determined. Even under these circumstances, the total number of continuous tuning parameters that has to be determined is 3. Alongwith M and P, the total number of tuning parameters that has to be determined iteratively is 5 and so the tuning procedure is extremely time consuming. Tuning parameters for the Tennessee Eastman (TE) process and the Azeotropic (AZ) tower for different controller configurations obtained using the above described approach are given in Table 3.3. In order to obtain the tuning parameters, the controller performance was observed for IDV(8) for the Tennessee Eastman (TE) problem. A lower variance in the ratio of G to H in the product stream is considered desirable. For the Azeotropic (AZ) tower, a random fluctuation in the feed composition (the one described in section 2.2) was considered and the aim was to minimize the variation in the product composition, i.e. mole fraction of water in the bottoms (denoted by $x_B(H_2O)$). The period of observation was 2500 minutes in both cases. The tuning parameters listed in Table 3.3 have been optimized for these conditions.

Next a study of the effect of various tuning parametsrs on controller performance is conducted by varying one parameter about it's optimal value and keeping the others

constant. It can be seen from Figure 3.12 (A) and Figure 3.13 (A) that M has a substantial effect on controller performance. An optimal value of M for IDV(8) for the TE problem is 4 when D-feed SP is being used as the MV while for the AZ tower, the optimal value is 3 when the fifth tray temparature SP is the MV. The values of P and $\mathbf{\Lambda}$ were held constant during this exercise. Similar trends were observed when other configurations were used. It should be noted that the value of M for the case when Reactor-level is used as the manipulated variable is 2. This value of M is a result of the fact that the score variable exhibits inverse response to the MV moves. For situations when there is inverse response, controllers need to be severely detuned. As a result, $M = 2$ values are typical. It should be kept in mind that if one attempts to get rid of the inverse response in the closed-loop behavior when the open-loop exhibits inverse response, then the system would become unstable.

From Figure 3.12 (B) and Figure 3.13 (B) it can be seen that P does not affect the controller performance significantly. The values of M and $\mathbf{\Lambda}$ were held constant during this exercise. It should be noted that for a continuous process, the chance that the linearized model is inaccurate is higher after a long time. So a very large P is not advisable. Moreover, in this specific case, the short term prediction provided by the model is reasonable. However, for the long term, it becomes increasingly difficult to invert the model. For example, it is impossible to invert the model when it is of an integrating type.

Figure (3.14) shows that the weight on the manipulated variable moves $\mathbf{\Lambda}$ has a significant effect on controller performance. As mentioned before, for the case when Reactor-level was used as the manipulated variable for the TE process, the score variable exhibits inverse response to the MV moves. For this case, the controller was detuned using $M = 2$. As a result of this detuning, a $\mathbf{\Lambda}$ of zero would give an optimal performance.

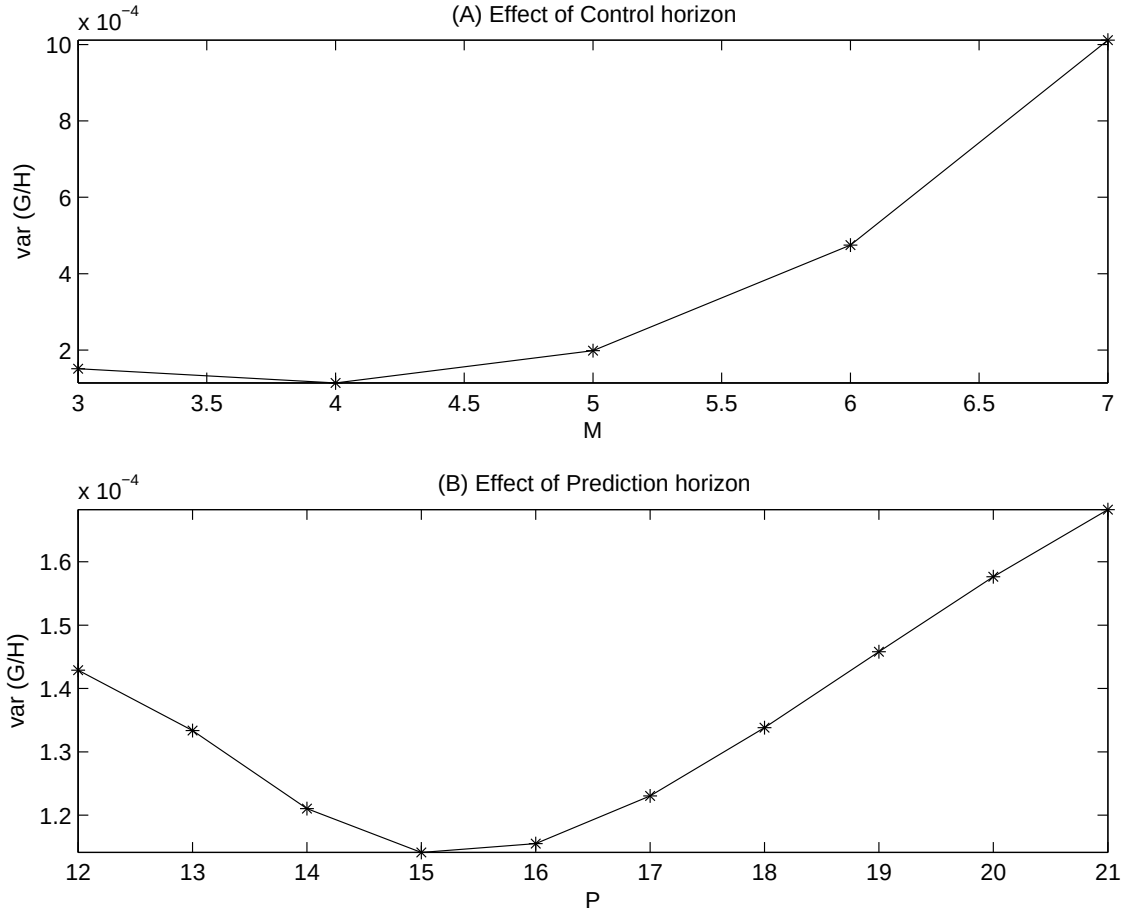

Figure 3.12: Effect of control horizon (M) and prediction horizon (P) on controller performance for the Tennessee Eastman (TE) process.

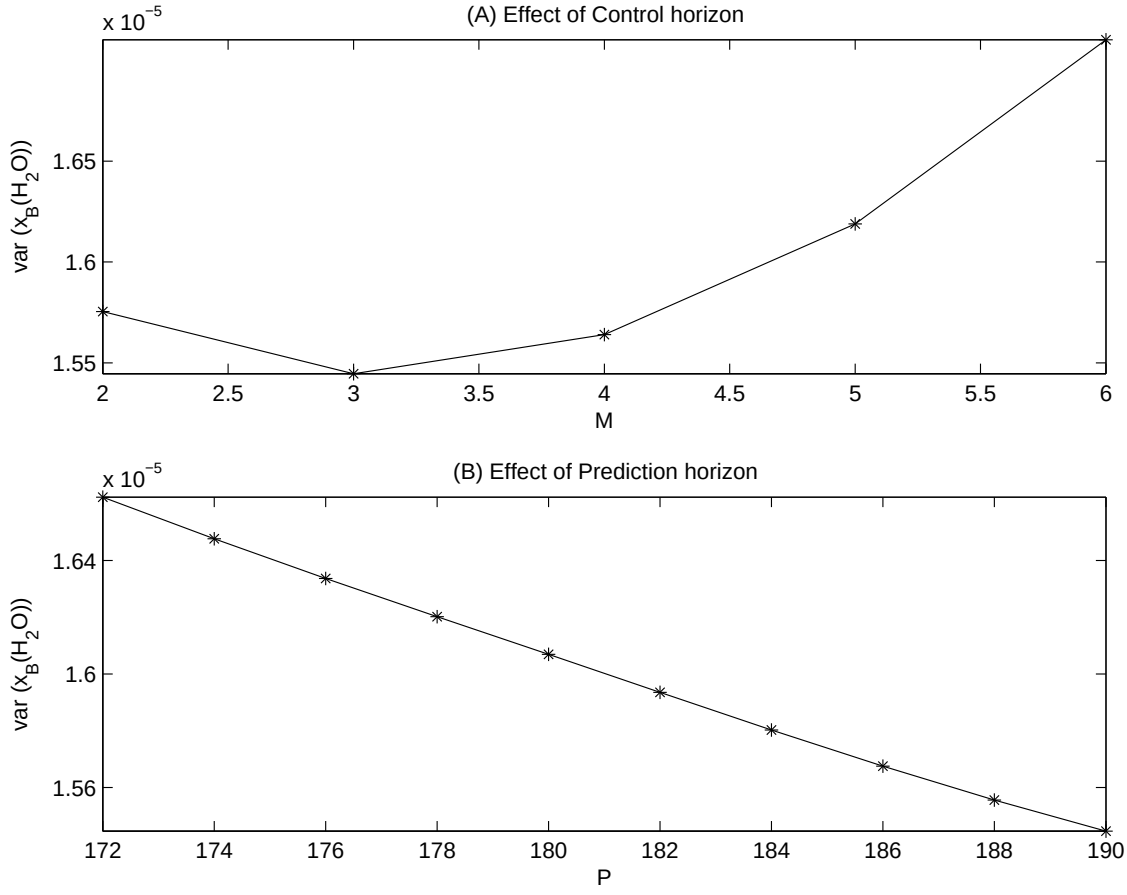

Figure 3.13: Effect of control horizon (M) and prediction horizon (P) on controller performance for the Azeotropic (AZ) tower.

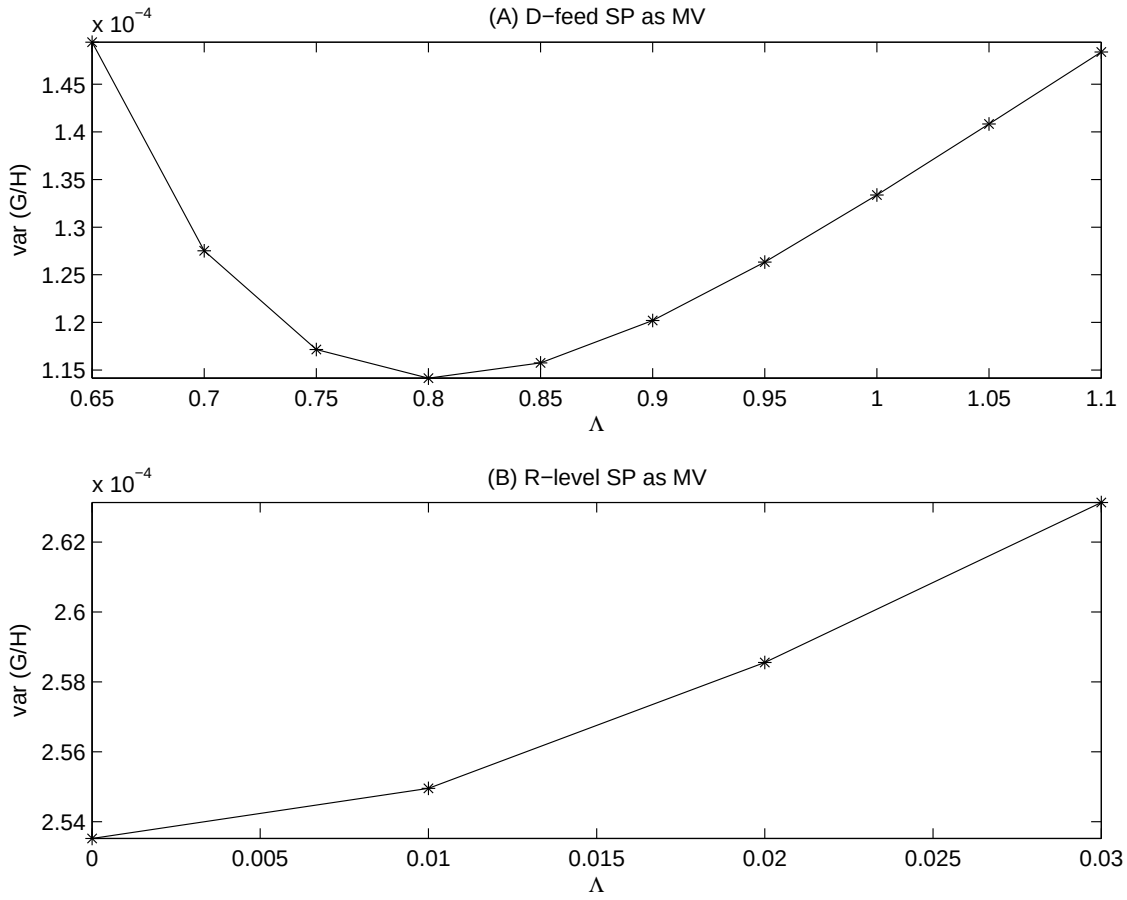

Figure 3.14: Effect of weight on Δu ($\mathbf{\Lambda}$) for univariate MP-SPCs controlling the Tennessee Eastman (TE) process: (A) D-feed SP as MV and (B) R-level SP as MV.

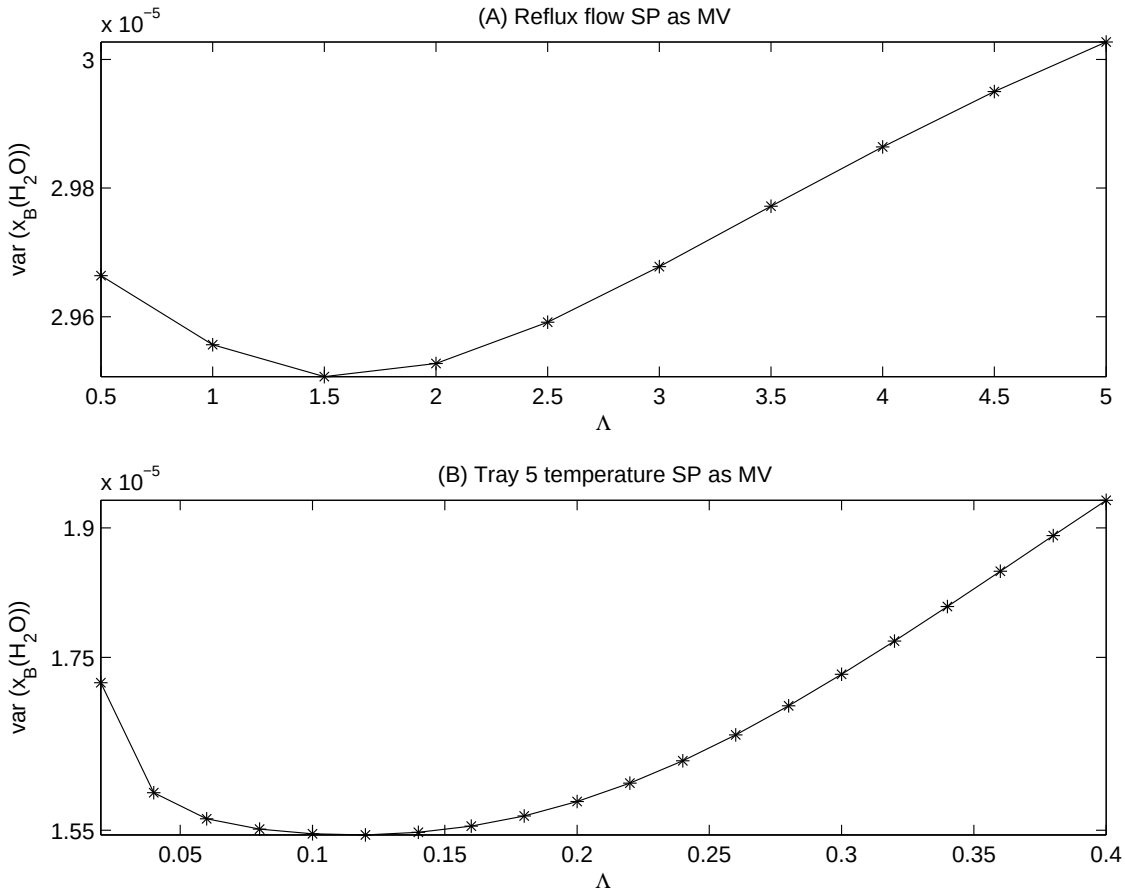

Figure 3.15: Effect of weight on Δu ($\mathbf{\Lambda}$) for univariate MP-SPCs controlling Azeotropic (AZ) tower: (A) Reflux flow SP as MV and (B) Tray 5 temperature SP as MV.

In general, for situations when there is inverse response, controllers need to be severely detuned and $M = 2$ and $\mathbf{\Lambda} = 0$ values are typical. A similar trend is also observed for the Azeotropic (AZ) tower. It can be seen in Figure (3.15) that $\mathbf{\Lambda}$ has a substantial effect on controller performance. Based on these trends, it can be concluded that the weight on the input moves, i.e. $\mathbf{\Lambda}$ should be used as the primary tuning parameter for MP-SPC.

3.4 Conclusions

This study deals with quality control in continuous processes using multivariate statistical tools. In this chapter, a feedback based score control methodology, called model predictive statistical process control (MP-SPC), is discussed in detail. Results based on the Tennessee Eastman (TE) process and the Azeotropic (AZ) tower case studies showed a significant reduction in the variabilty of the product quality in the presence of both stationary and non-stationary disturbances. The most important feature of this approach is that quality control is achieved without using real-time quality or disturbance measurements. MP-SPC owes it's strength to the fact that not only is the controlled variable (i.e. the score variable) correlated with the quality variable, but also that the scores have a significant lead time over the quality variable. It is also ascertained that feedback effects of putting the MP-SPC on line do not alter the cross-correlation structure between the score variable(s) and the quality variable significantly, provided that the correct manipulated variables and measurements are used in the MP-SPC formulation.

The next chapter will involve implementation issues: generation of historical databases, selection of process measurements, PCA model development and identification issues.

Chapter 4

Database generation, measurement selection, score model development and identification issues

In the previous chapter, a novel control methodology that integrated model predictive feedback control with statistical process control, termed MP-SPC, was presented. On-line implementation of MP-SPC was discussed in detail. Simulations of industrially benchmarked case studies were used to demonstrate the significant reductions in product variability achieved by MP-SPC. In this chapter, the steps that are necessary prior to on-line implementation are discussed.

The MP-SPC methodology involves the following steps:

1. Database generation and pre-processing: A process measurement database that combines historical operation data with plant testing data is generated as described in section 4.1. All the possible different disturbance scenarios MP-SPC is expected to encounter (i.e. during control in the future) are incorporated into the historical operation data. Next, the data is pre-processed.

2. Measurement and manipulated variable selection: A set of continuous, real-time secondary measurements (Y) and a set of manipulated variables (u) are selected. The selection of manipulated variables (MVs) is not within the scope of this study. For the case studies being evaluated, it is assumed that a knowledge of the most appropriate MVs for quality control is available. Measurement selection is discussed in section 4.2.

3. Plant testing using PRBS forcing: The manipulated variables (i.e. MVs) are varied

sequentially in a Pseudo-Random Binary Sequence (PRBS) as a part of a designed identification experiment. Deciding the sampling time (Δt), the length of the PRBS experiment (N_e) and other aspects of plant testing is described in section 4.4.1.

4. Reduction of measurement dimensionality using orthogonal PCA: The process measurement database is scaled to zero mean and unit variance. Principal component analysis (PCA) is performed on the scaled database. The scores obtained using ordinary PCA are transformed to obtain scores that are collinear with the MVs. Various issues dealing with orthogonal PCA calculations that generate these collinear scores (denoted as $\mathbf{z}$) are explained in section 4.3.

5. Dynamic PC score model identification: A discrete time linear dynamic model is developed from the data matrix consisting of the transforned scores (i.e. $\mathbf{z}$) and MVs (i.e. $\mathbf{u}$). The various aspects of score model identification are discussed in section 4.4.2 and section 4.4.3.

6. On-line implementation using an MPC algorithm: Once a dynamic PC score model relating the MVs to the collinear score variables is available, this model is incorporated into a linear model predictive control framework.

It should be noted that the order of steps 1 through 5 (i.e. the database generation, measurement selection, PCA model development and dynamic PC score model identification steps) is not fixed and is subject to adjustments. The choices made in one step might influence those made in another step and so actual implementation might not follow the above mentioned sequence strictly.

4.1 Database generation and pre-processing

Model-based and inferential control methodologies traditionaly use plant test data alone to build the models that are eventually incorporated into the controllers. It would be worthwhile to explore the possibility of exploiting the information contained in historical databases to build such models. If the historical operation data incorporates many different disturbance scenarios, then the model built using such data would be generated using many different sources of variation. As a result, when compared to the information contained in data from plant tests alone, the information contained in a combined database (i.e. historical operation data and data from plant tests) would be richer and the reduced model constructed using such a combined database would be able to capture the dynamics of the process more effectively. MP-SPC employs an approach based on this philosophy of combining historical operation data with data from plant tests. The overall scheme for this approach is explained in Figure 4.1.

The pre-processing step primarily involves scaling and outlier removal. For the case studies being evaluated in this study, it is assumed that linearization or differentiation of the measurements is not required. Measurement data from the historical operation data ($\mathbf{Y_H}$) and data from plant testing ($\mathbf{Y_P}$) are combined and scaled using the mean of the plant test data and the variance of the combined data. Historical operation data is usually not strictly stationary but contains small amounts of drift in the measurements. So, in such cases, the mean of the historical operation data would be quite different from the mean of the plant test data. One of the purposes of centering the data to zero mean is to ensure that the steady state value and hence the target setpoint for the score variables in MP-SPC is zero. However, if the dynamic model identified from the plant test data is scaled using a combined mean that is different from the mean for the plant test data, then

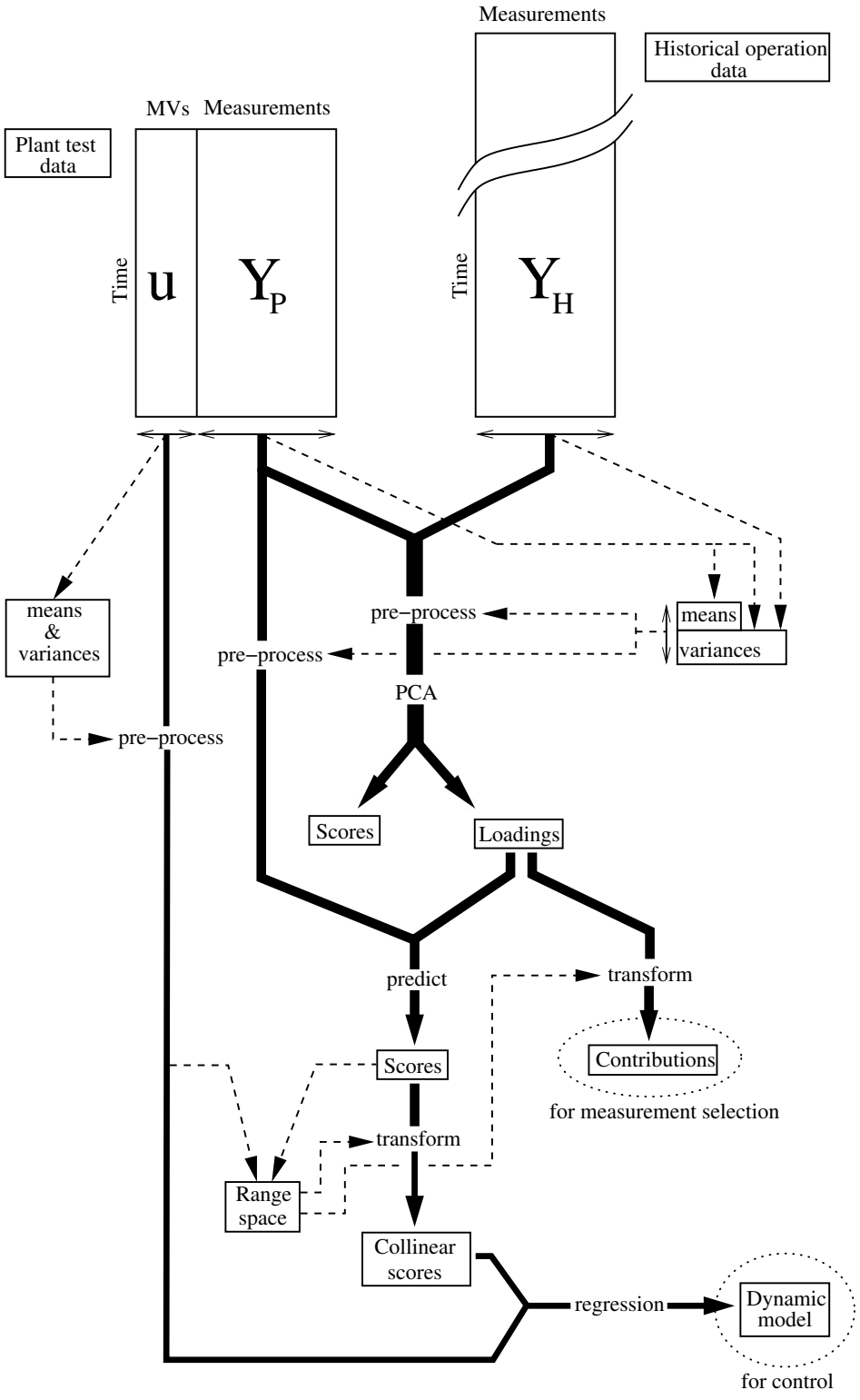

Figure 4.1: Schematic illustrating the use of both historical operation data and plant test data for measurement selection, score model development and identification.

the incorrect mean value would give rise to an erroneous step response model. It has been reported (see for example Brosilow and Joseph [3]) that model errors caused by a non-steady state initial condition can jeopardize controller performance. For this reason, the mean of the plant test data alone is used to scale the combined database. The combined variance is used to convert the data into unit variance because for some measurements, the historical operation data might have more variability while for some the plant test data might have more variability. Using a combined variance to scale the data would ensure that the model would encompass the entire range of variability for all the measurements. Thereby, the score variables would also always vary between a minimum value of -3 and a maximum value of 3.

Once the process measurement database that combines both historical operation data ($\mathbf{Y_H}$) and data from plant testing ($\mathbf{Y_P}$) is appropriately pre-processed, principal component analysis is performed to obtain the static, ordinary PCA model as shown in Figure 4.1. Samples with a very high Q-residual value are considered as outliers and are removed from the database. The pre-processing and PCA calculations are then repeated for the database. It should be noted that PCA treats samples as unrelated and unordered and so removing samples with high Q-residual values does not impair the PCA calculations.

The manipulated variable changes during normal plant operation (i.e. changes in setpoints for the Base control system) aren't independent of each other. As a result, correlations are built into such data. Moreover, the level of excitation is also not enough and the resulting signal to noise ratio is poor. Owing to these shortcomings, historical operation data do not contain any useful dynamic information and so are not used beyond this point (i.e. once the PCA model has been developed). The loadings from the PCA model of the combined database are used to obtain the scores for the plant test data.

It should be noted that any sample removed from the plant test data during the outlier removal step should be reinserted into the plant test data when the loadings from the PCA model are used to obtain the scores for the identification experiment. This is done to ensure that the dynamic information contained in the plant test data is preserved. The ordinary PCA scores for the plant test data are transformed using the corresponding manipulated variable ($\mathbf{u}$) information in such a way that the resulting scores are collinear with the MVs. A dynamic model describing the response of the collinear scores to step or impulse changes in the MVs is obtained using a suitable regression technique. This dynamic model will be used for control. The loadings from the PCA model obtained earlier (i.e. for the combined database) are transformed in an identical fashion into "contributions". These contributions reflect how much each measurement participates in the PCA model being used for control. This information is useful in measurement selection.

In order to generate the historical database for the Tennessee Eastman (TE) process, process operation was simulated for different disturbance scenarios, with the Base control system in place, for a total of 50,000 minutes. Table (4.1) lists the various disturbances that are active during each time period. The disturbance IDs, stream numbers and descriptions are the ones reported in the origianl paper by Downs and Vogel [12]. It can be seen that the historical database encompasses a wide range of disturbance types and directions which is typical of continuously running chemical process plants. Figure (4.2) shows some of the key measurements recorded in the database: E-feed flow, Reactor pressure, Reactor temperature and D-feed flow. It should be noted that except for IDV(13) i.e. a slow drift in the reaction kinetics, all the disturbances are random fluctuations. In other words, the disturbances are stationary about the same steady state (i.e. the one corresponding to the operating mode for 50/50 G/H ratio in the product). Similarly, Figure (4.3)

Table 4.1: Disturbances incorporated in the historical database for the Tennessee Eastman (TE) process.

Time period (hr)	Disturbance		
	ID	Random fluctuation	Stream
0 to 70	IDV(8)	A, B and C compositions of the C-feed	4
70 to 200	IDV(9)	Temperature of the D-feed	2
200 to 400	IDV(10)	Temperature of the C-feed	4
400 to 700	IDV(11)	Temperature of the cooling water in the inlet to the Reactor cooling bundle	12
700 to 800	IDV(12)	Temperature of the cooling water to the Condenser	13
800 to 833.33	IDV(13)	Reaction kinetics (slow drift)	-

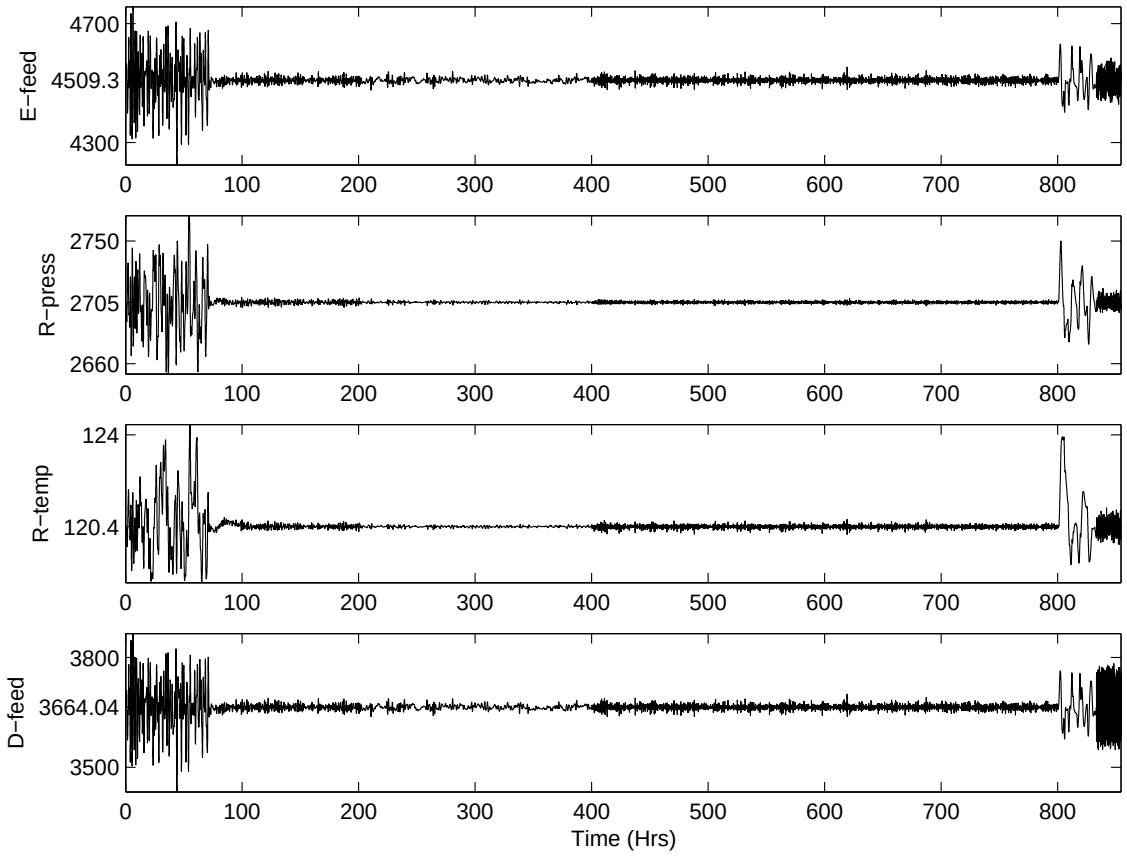

Figure 4.2: The historical database for the Tennessee Eastman (TE) process.

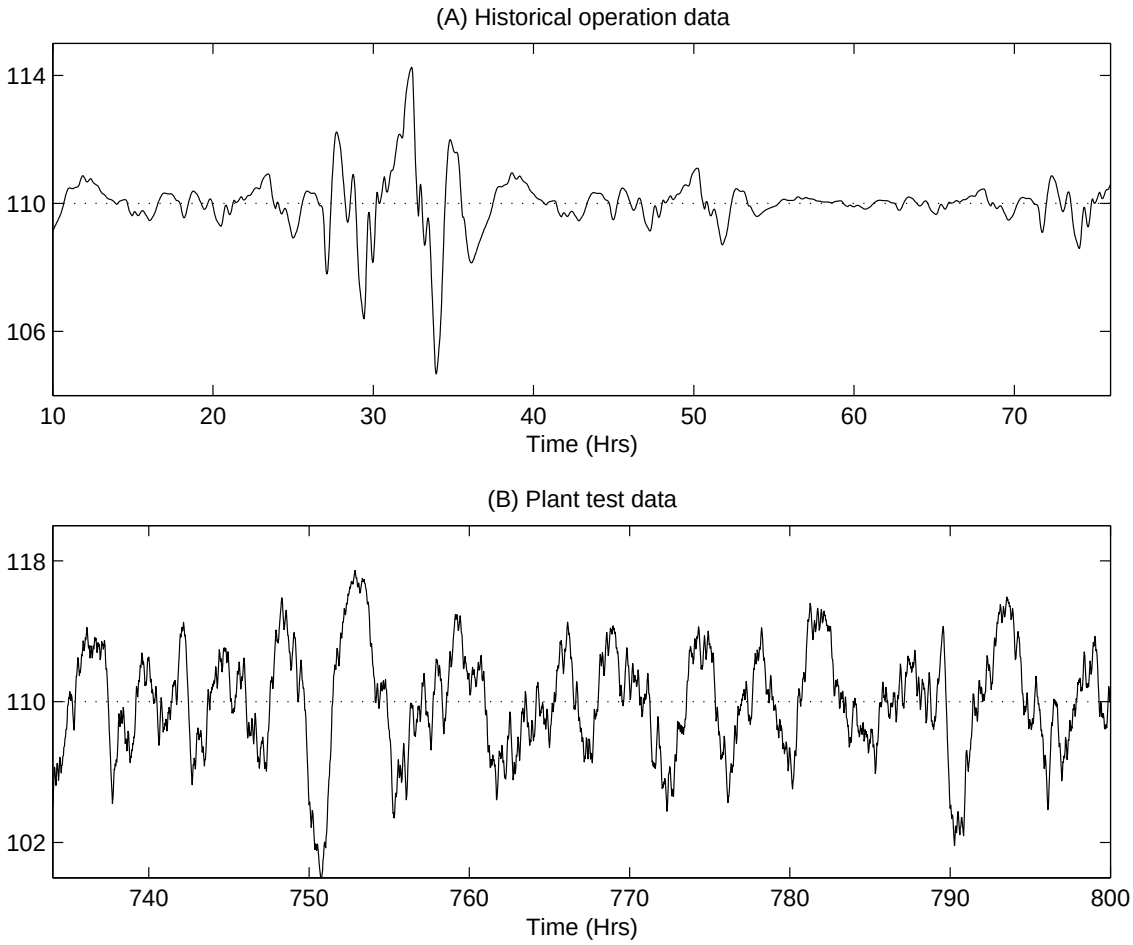

Figure 4.3: The combined database for AZ.

shows the historical operation data and plant test data for the Azeotropic (AZ) tower. Here, the fifth tray temperature, recorded over about 66 hours of normal plant operation (historical operation data) and about 66 hours of plant test data, is shown for illustrative purposes. A disturbance of varying magnitude in the feed to the column is the cause of the variability in the bottoms product quality ($x_B(H_2O)$) and this disturbance also affects the fifth tray temperature.

Once the historical operations data ($\mathbf{Y_H}$) are available, the next step is to carry out plant testing. Pseudo Random Binary Sequence (PRBS) inputs are used in an experiment designed specifically to obtain data which can be useful in identifying a dynamic process

model. Issues pertaining to plant testing are tied to the dynamics of the process and hence the choice of the sampling time (Δt), the length of the experiment (N_e), the magnitude of the perturbations, etc. are explained in the dynamic model identification section (i.e. section 4.4.1). The selection of secondary measurements assumes that the historical operation data and the plant test data (i.e. a record of the manipulated variable moves $\mathbf{u}$ and the measurements $\mathbf{Y_P}$ during the identification experiment) are available. This step is discussed next.

4.2 Measurement selection

McAvoy et al [33] developed a nonlinear inferential parallel cascade and Chen et al [6] developed a lagged PCA model as a part of a multivariate statistical (i.e. score) controller for on-line quality improvements for the Tennessee Eastman (TE) process. In both cases the same 11 measurements were chosen to form the model used for control. For the same process, McAvoy [31] chose 5 measurements to demonstrate the MP-SPC methodology. In all three studies, the choice of measurements was not too difficult a task because a steady state process model and prior knowledge about the dynamics of the process under consideration, the Tennessee Eastman (TE) process, was available. However the selection of measurements could be a challenge for new, lesser understood and complicated processes. Hence, it is necessary to generate systematic rules for selecting measurements, particularly when a steady state or dynamic simulation is not available. In such cases statistical analysis of process measurements could be used to make this selection.

Some general guidelines for selecting secondary measurement are given in section 1.4.2. Although the measurement selection procedure described next is specifically for MP-

SPC and related controllers, the approach is quite general and with slight adjustments, can be applied to most inferential control applications based on PCA/PLS. The measurement selection procedure presented here consists of two stages. The first stage involves the manual screening of measurements. The second stage involves an iterative algorithm and can be implemented as a computer code for automatic selection.

4.2.1 First stage

In the first stage, measurements having the certain characteristics detrimental to MP-SPC performance are excluded from the candidate list. This screening procedure is conducted manually. It is recommended that the following measurements should not be considered as candidate secondary measurements:

1. Do not consider measurements that do not show adequate sensitivity to the manipulated variable (MV) moves. For example, when D-feed SP is being used as a MV and is perturbed by ±1% (i.e. ±36.64 gmol/min), the condenser cooling water temperature changes from a mean value of 77.2970 °C to a maximum value of 77.3056 °C. Clearly, this measurement is insensitive to the manipulated variable changes and it would make a poor choice as a secondary measurement. So it should be eliminated as a candidate.

2. Do not consider a measurement that is part of a lower loop whose setpoint is being used as a MV for MP-SPC. This is done to avoid incorporating any correlation between the measurements used to be built the model and the MVs. Hence, for the Tennesse Eastman (TE) case, when D-feed SP is being used as a MV, D-feed flow should be removed from the candidate measurement list.

3. Do not consider measurements that exhibit inverse response with respect to the

manipulated variable. For example for TE, as seen in Figure 4.4, A-feed flow shows inverse behavior when D-feed SP is used as a MV and so A-feed flow should not be considered as a candidate measurement.

4. Do not consider measurements which exhibit inconsitency in their cross-correlaion with respect to the quality variable in the presence of disturbances and MV forcing. For example, for TE, as seen in Figure (4.5), %E in the reactor feed is negatively correlated with G/H ratio when D-feed is used as a manipulated variable. However, %E in the reactor feed is positively correlated with G/H ratio in the presence of IDV(8). Hence, the cross-corelation between %E in the reactor feed and the product G/H ratio is not consistent and so this measurement should not be considered as a candidate.

4.2.2 Second stage

Candidate secondary measurements that are not eliminated in the first manual screening stage are evaluated further in the second atage. The second stage of the measurement selection procedure is based upon striking a compromise between maximizing the cross-correlation and the lead-time between the collinear scores and the quality variables. As shown in Figure 4.6, the second stage of the measurement selection procedure consists of two steps.

In the first step, the cross-correlation coefficient ($r_{i,q,\tau}$) between each secondary measurement candidate (i) and the quality variable (q) is evaluated at each time-shift (τ). Process measurement data is available as a time series. At any time shift τ, the

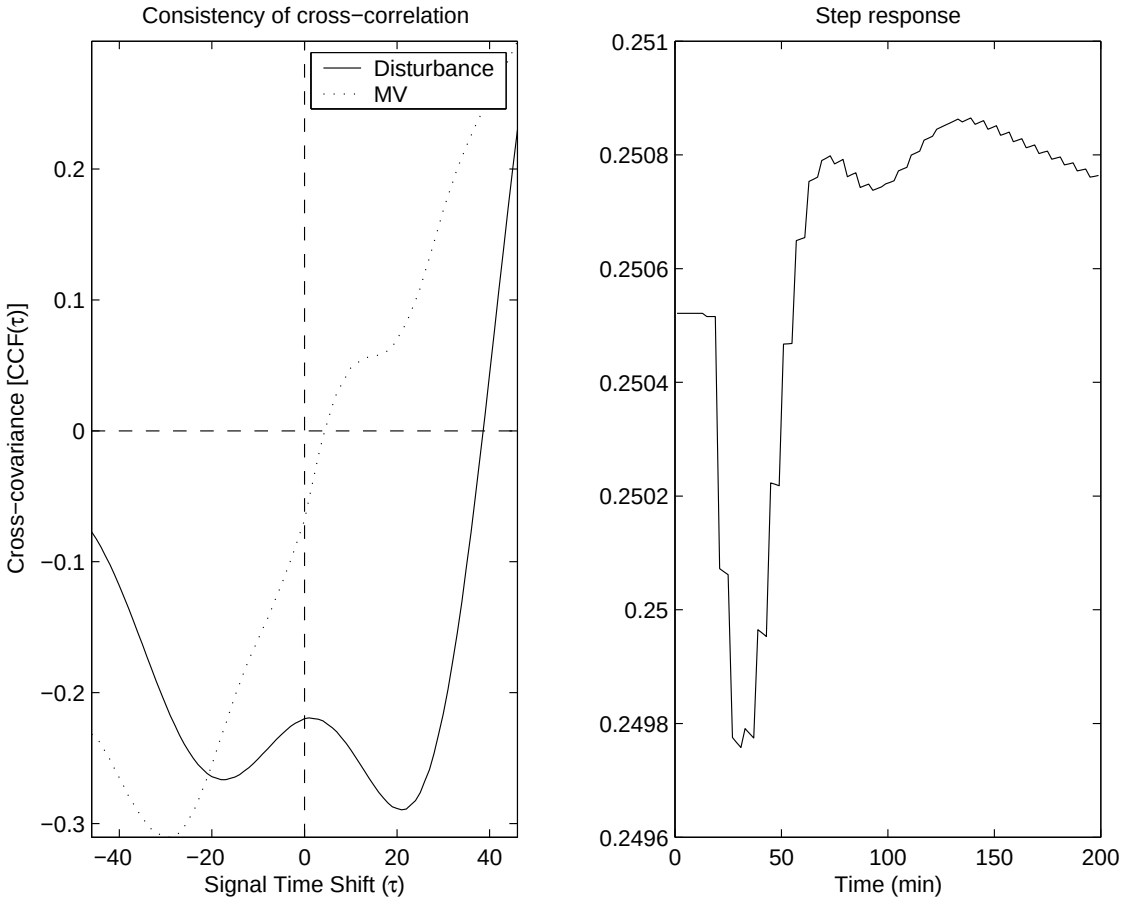

Figure 4.4: Manual screening for consistency of correlation and inverse response: A-feed flow for the Tennesse Eastman (TE) case.

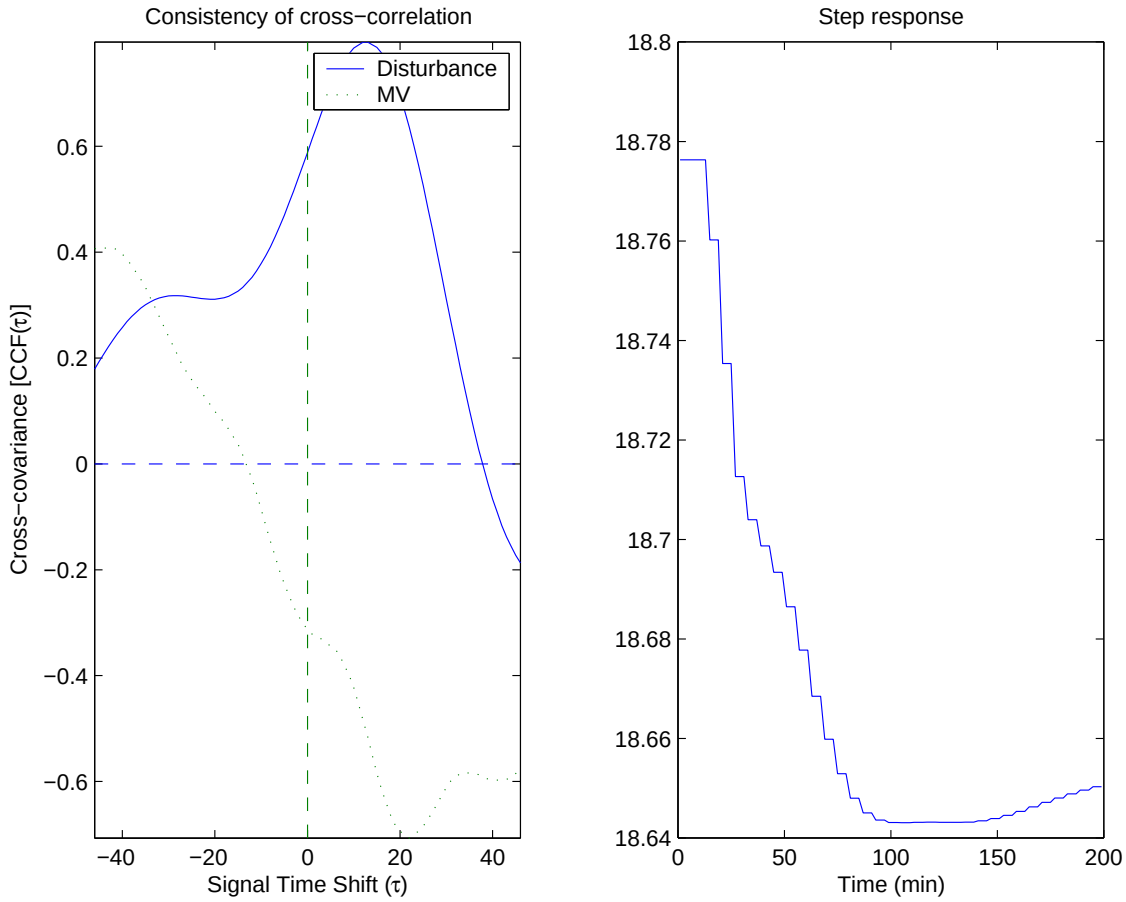

Figure 4.5: Manual screening for consistency of correlation and inverse response: %E in reactor feed for Tennesse Eastman (TE) case.

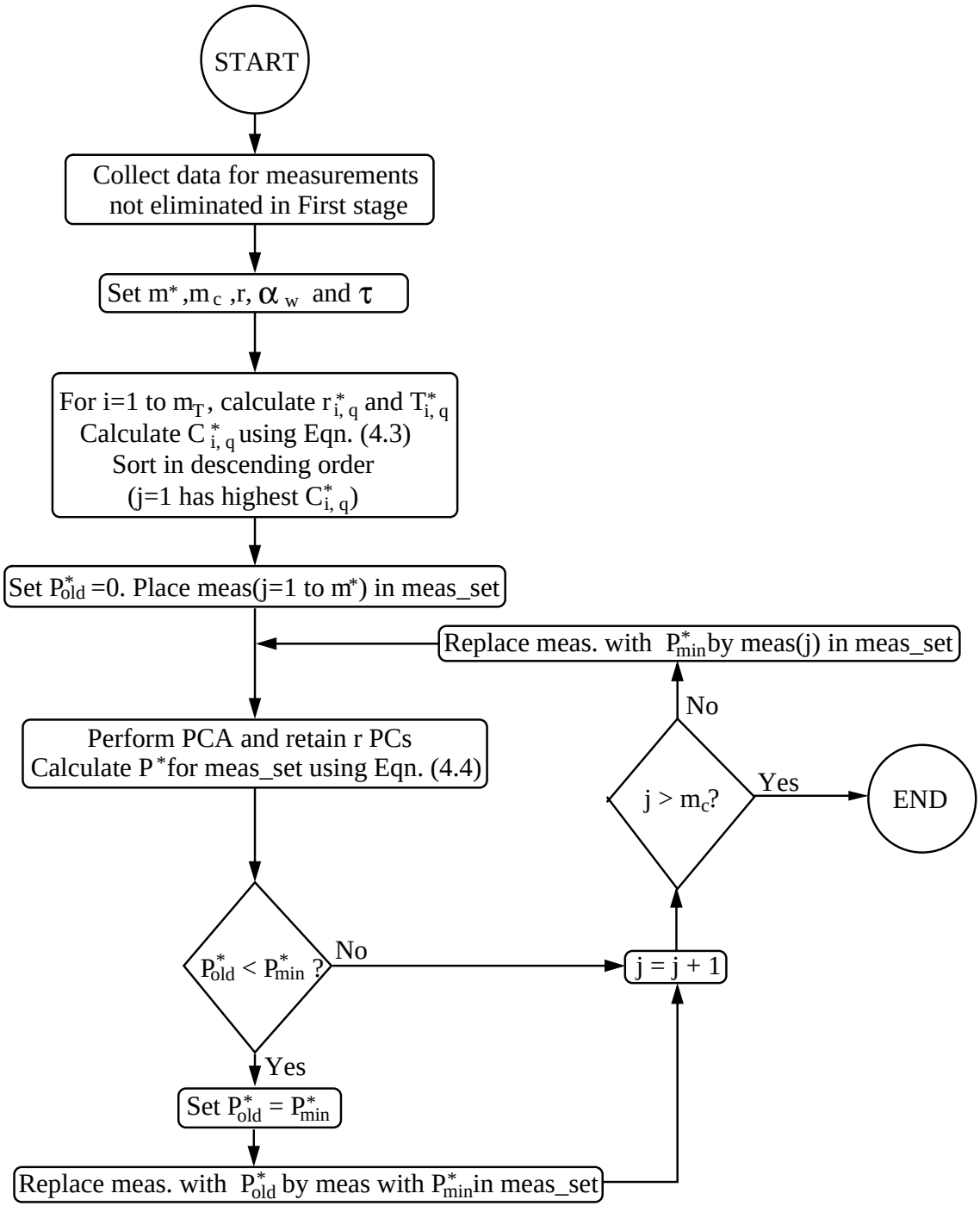

Figure 4.6: Algorithm for the second stage of measurement selection.

cross-correlation coefficient $r_{i,q,\tau}$ is defined as:

$$r_{i,q,\tau} = \frac{\sum_{m=1}^{n_t}(y_i(m \pm \tau) - \bar{y_i})(y_q(m) - \bar{y_q})}{\left(\sqrt{\sum_{m=1}^{n_t}(y_i(m \pm \tau) - \bar{y_i})^2}\right)\left(\sqrt{\sum_{m=1}^{n_t}(y_q(m) - \bar{y_q})^2}\right)} \tag{4.1}$$

where $n_t + 2\tau$ is the total number of time series data points available for analysis. Since both positive and negative time-shifts are considered, the time vector is of size $2\tau + 1$. The maximum of the absolute value of the cross-correlation ($r^*_{i,q} = max(abs(r_{i,q,\tau}))$) and the time-shift ($T^*_{i,q}$) at which this maximum occurs is noted. The absolute value of the cross-correlation is taken because the two measurements could be positively or negatively correlated. Only the measurements that lead the quality variable measurements are chosen.

The $C^*_{i,q}$ of a measurement i is an attempt to combine the effects of lead time and cross-correlation. The aim is to rank the measurements based on their $C^*_{i,q}$. One way to rank the measurements is to evaluate $C^*_{i,q}$ using:

$$C^*_{i,q} = r^*_{i,q} + \alpha_w * T^*_{i,q} \tag{4.2}$$

and then to arrange them in a descending order. However, it is possible that a particular measurement has a very large lead time but is un-correlated with the quality variable. As a result the large $T^*_{i,q}$ could completely dwarf the smaller $r^*_{i,q}$. This measurement could end up having a large $C^*_{i,q}$ and thereby getting a high rank. This problem can be overcome by defining $C^*_{i,q}$ as:

$$C^*_{i,q} = (r^*_{i,q})^{\alpha_w} * T^*_{i,q} \tag{4.3}$$

The parameter α_w can be used as a tuning parameter to adjust the relative importance of cross-correlation versus lead time. A large value of α_w would mean that more importance is given to the cross-correlation while a small value of α_w would mean that more importance

is given to the lead time. In this study, Equation 4.3 is used to evaluate $C^*_{i,q}$ for each measurement and then the measurements are rearranged in descending order.

The second step involves making sure that the chosen measurements contribute adequately to the PCA model. The absolute value of the loadings is a measure of how much a particular measurement participates in the model. In other words, a measurement with an absolute loading value close to 0 contributes very little information and should not be used. Since collinear scores are being used in MP-SPC, the loadings are also transformed accordingly. The transformation matrix used to obtain the collinear scores from the ordinary PC scores is used transform the loadings. For each measurement i, the maximum of the absolute value of it's transformed loadings is termed it's "contribution" to the PCA model and is denoted as p^*_i:

$$\left.\begin{aligned} \mathbf{P}^* &= [\mathbf{p}^*_1\ \mathbf{p}^*_2\ \cdots\ \mathbf{p}^*_{m^*}] = \mathbf{P}\ \mathrm{orth}(\mathbf{\Theta}^T\mathbf{u}) \\ p^*_i &= \|\mathbf{p}^*_i\|_\infty \end{aligned}\right\} \tag{4.4}$$

The dimension of the original loadings matrix (denoted as $\mathbf{P}$) is $m^* \times r$. After projecting the original loadings onto the transformation matrix, the dimension of the loadings matrix becomes $m^* \times n_u$ and is denoted by $\mathbf{P}^*$. Here, m^* is the number of measurements used to build the PCA model, n_u is the number of inputs (i.e. the MVs) and r is the number of latent variables retained in the PCA model (take $r = n_u + 1 = 3$ when two MVs are being used). In the implementation adopted in this study, the largest loading associated with each measurements is used for deciding it's contribution to the model however the sum of the absolute values of the loadings could also be used instead of the maximum value.

Figure 4.6 shows the iterative evaluation of the ranked measurements. Here m_T is the total number of candidate measurements available after the first stage, Contributions for the first m^* measurements are calculated using Equation 4.4.

4.2.3 Results and discussion

The result of using this approach for the Tennessee Eastman (TE) case when D-feed SP is used as one of the manipulated variables (MVs) is that only measurements Y(7) - Reactor Pressure, Y(13) - Separator Pressure, Y(16) - Stripper Pressure, Y(21) - Reactor cooling water temperature and Y(22) - Condenser cooling water temperature remain after the first stage. During the manual screening, the remaining 31 measurements are eliminated either because their cross correlation with G/H is not consistent or because they exhibit inverse response. Out the five measurements chosen to build the reduced order predictive model for MP-SPC, Figure 4.7 shows plots for the cross-corelaion between four of these measurements and the quality variable (G/H ratio). It can be seen that all the four plots look similar and that the plots for the three pressures are almost identical. Although the maximum absolute value of the correlation coefficient is about 0.5, subsequent processing (i.e. orthogonal PCA calculations) will increase the value of the correlation coefficient between the scores and the quality variable to about 0.8. When the Reactor level is used as one of the MVs, all the measurements exhibit inverse response.

For the sake of illustrating the measurement selection algorithm, consider the Tennessee Eastman (TE) case and using D-feed SP as the manipulated variable, assume that all the measurements pass the first, manual screening stage. If this is the case, then for each measurement i, it's $C^*_{i,q}$ is calculated using Equation 4.3. The α_w parameter in this case is chosen to be 2. It should be noted that the measurements have already been scaled to zero mean and unit variance. Since in the case of the Tennessee Eastman (TE) process, there is slow drift component in the historical operation data due to the inclusion of IDV(13) (i.e. reaction kinetics), the mean values will not be the true steady state value. If the mean of the combined data (historical operation data + plant test data) is used, this

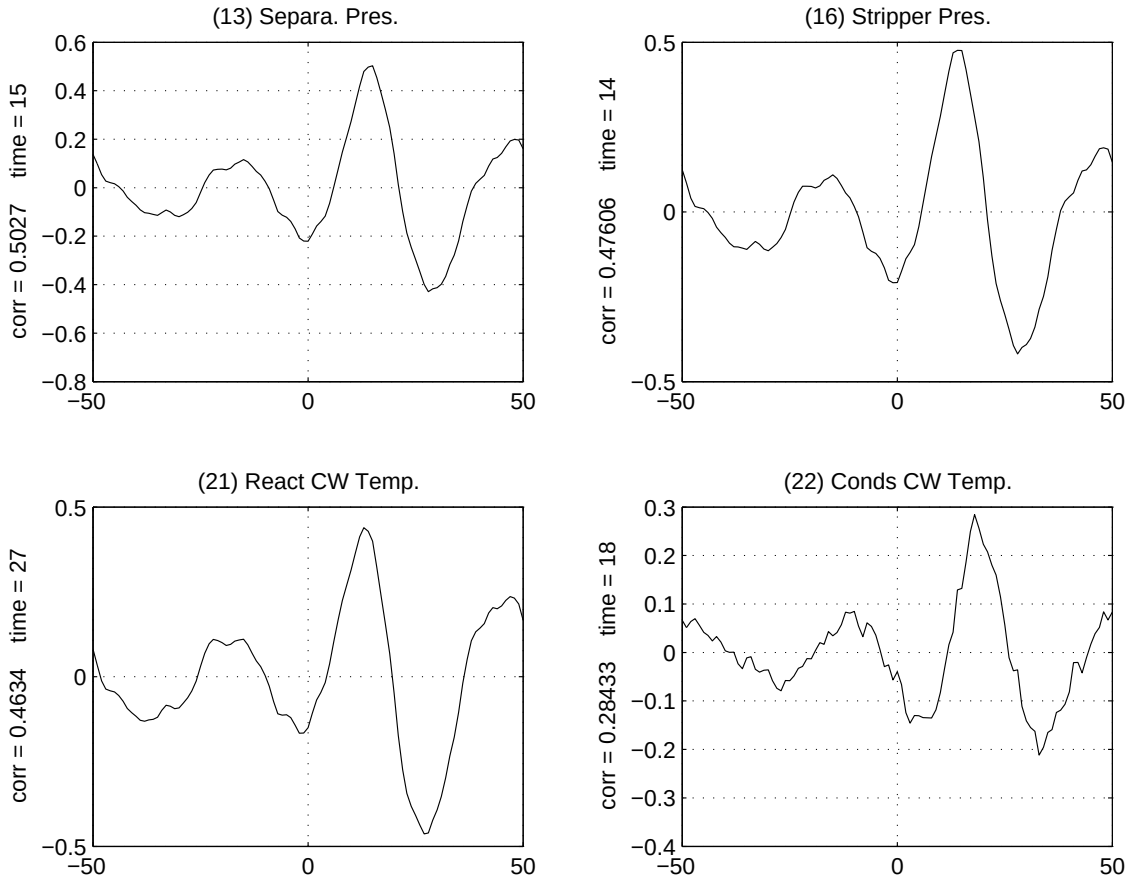

Figure 4.7: Cross correlation plots for Tennessee Eastman (TE) process.

Table 4.2: Ranking of candidate measurements as per their $C^*_{i,q}$ values for the Tennessee Eastman (TE) case.

Rank	Measurement (i)	$C^*_{i,q}$
1	(20) Compressor Work	0.1094
2	(5) Recycle Flow	0.0871
3	(21) React CW Temperature	0.0833
4	(11) Separator Temp.	0.0654
5	(6) Reactor Feed	0.0516
6	(15) Stripper Level	0.0452
7	(18) Stripper Temperature	0.0389
8	(17) Product Flow	0.0334
9	(9) Reactor Temperature	0.0334
10	(36) Purge Gas %H	0.0260
11	(8) Reactor Level	0.0225
12	(3) E-Feed Flow	0.0202
13	(34) Purge Gas %F	0.0166
14	(32) Purge Gas %D	0.0163
15	(35) Purge Gas %G	0.0155

could lead to an offset in the subsequent dynamic model identification calculations. The mean of just the plant test data is used to scale the measurements instead.

Next, the measurements are rearranged in descending order as per their $C^*_{i,q}$ values. The number of measurements used to form the PCA model (denoted by m^*) is 5 and the maximum number of candidate measurement sets to be evaluated (denoted by m_c) is 10. Hence, only the top 15 ranked measurements need to be examined further and these are listed in Table (4.2). The first m^* (in this case 5) measurements are taken up and an ordinary PCA model for the previously scaled combined database (historical operation data + plant test data) is generated. Since only the D-feed SP was used as the manipulated variable, the number of PCs retained (i.e. r) is 2. Using the loadings ($\mathbf{P_{P+H}}$) for this ordinary PCA model and the plant test data alone, the scores for the plant test data (denoted by $\theta_{\mathbf{P}}$) are generated. As illustrated in Equation 4.4, contributions

(denoted as p_i^*) for each measurement i in the measurement set are evaluated using the same transformation matrix as that used to generate the collinear scores. For the first measurement set, the contributions are:

$$\begin{bmatrix} Y(20) & 0.0749 \\ Y(5) & 0.3910 \\ Y(21) & 0.2043 \\ Y(11) & 0.7525 \\ Y(6) & 0.4833 \end{bmatrix}$$

It can be seen that measurement number 20 (i.e. Y(20) - Compressor Work) has the smallest contribution (i.e. 0.0749) and so it has the potential to be eliminated from the measurement set. Another measurement set is constructed in which the next highest ranked measurement (number 6: Y(15) Stripper Level) is inserted in the measurement set in place of Y(20). Contributions for this measurement set are calculated as before:

$$\begin{bmatrix} Y(21) & 0.4954 \\ Y(5) & 0.5009 \\ Y(6) & 0.4961 \\ Y(11) & 0.5069 \\ Y(15) & 0.0254 \end{bmatrix}$$

The measurement with the smallest contribution (i.e. 0.0254) for this set is Y(15). This contribution value means that compared to Y(20), Y(15) does not impart any extra information (0.0254 versus 0.0749). So Y(20) is left in the measurement set. In the third iteration, the next highest ranked measurement Y(18) - Stripper Temperature is included

in the measurement set. Contributions for this set are:

$$\begin{bmatrix} Y(21) & 0.4373 \\ Y(5) & 0.4553 \\ Y(6) & 0.4471 \\ Y(11) & 0.4552 \\ Y(18) & 0.4409 \end{bmatrix}$$

For this set, the smallest contribution (i.e. 0.4409) is that of Y(18). However, this contibution is larger than that for Y(20) - Compressor Work. Hence, Y(20) is eliminated in favour of Y(18) for all future considerations. However, Y(18) could still be eliminated from the measurement set. In order to check whether Y(18) is eliminated, the next highest ranked measurement Y(17) Product Flow is inserted in the measurement set and the contributions for this set are calculated:

$$\begin{bmatrix} Y(18) & 0.4082 \\ Y(6) & 0.5143 \\ Y(11) & 0.5273 \\ Y(5) & 0.5200 \\ Y(17) & 0.1430 \end{bmatrix}$$

For this measurement set, the smallest contribution (i.e. 0.1430) is that for Y(17). Since this is smaller than the 0.4409 for Y(18), Y(18) is retained. In this way, a total of 10 measurement sets are iteratively evaluated and a set of measurements that individually contribute the highest amount of information to the PCA model is retained. In the case under consideration, no new measurements are added to the measurement set after the third iteration and so this set is the final combination of measurements used to form the PCA model.

Table 4.3: Iterative measurement selection for the Tennessee Eastman (TE) process.

Iteration number	Measurement set (meas_set)					smallest contribution (p^*_{min})
1	20	5	21	11	6	0.0749
2	21	5	6	11	15	0.0254
3	**21**	**5**	**6**	**11**	**18**	**0.4373**
4	18	6	11	5	17	0.1430
5	18	6	11	5	9	0.1803
6	18	6	11	5	36	0.3077
7	18	6	11	5	8	0.0094
8	18	6	11	5	3	0.0220
9	18	6	11	5	34	0.2768
10	18	6	11	5	32	0.0986

The entire iterative procedure is given in Table (4.3). The set of measurements finally selected is highlighted in bold. It can be seen that the smallest contribution (p_*min) for the set of measurements finally selected is largest among all the set of measurements evaluated. The entire selection/rejection process is tranparent and one can see what are the factors contributing to a certain measurement making/not making the list.

As far as the Azeotrpic (AZ) tower is concerned, only tray temperatures 1 to 7 make it past the first stage. For the size of the MVs considered, the temperatures of trays 8 to 20 show very small deviations. Owing to the highly sensitive and nonlinear nature of the tower, larger size inputs could destabilize the tower. Figure 4.8 shows the cross-correlation plots for four of the chosen measurements. It can be seen that the maximum absolute value of the correlation coefficient is about 0.8.

4.3 Orthogonal PCA model development

Section 1.3.4 introduces Orthogonal PCA and the two step approach for obtaining scores collinear with the manipulated variables (MVs) in an efficient manner. This section

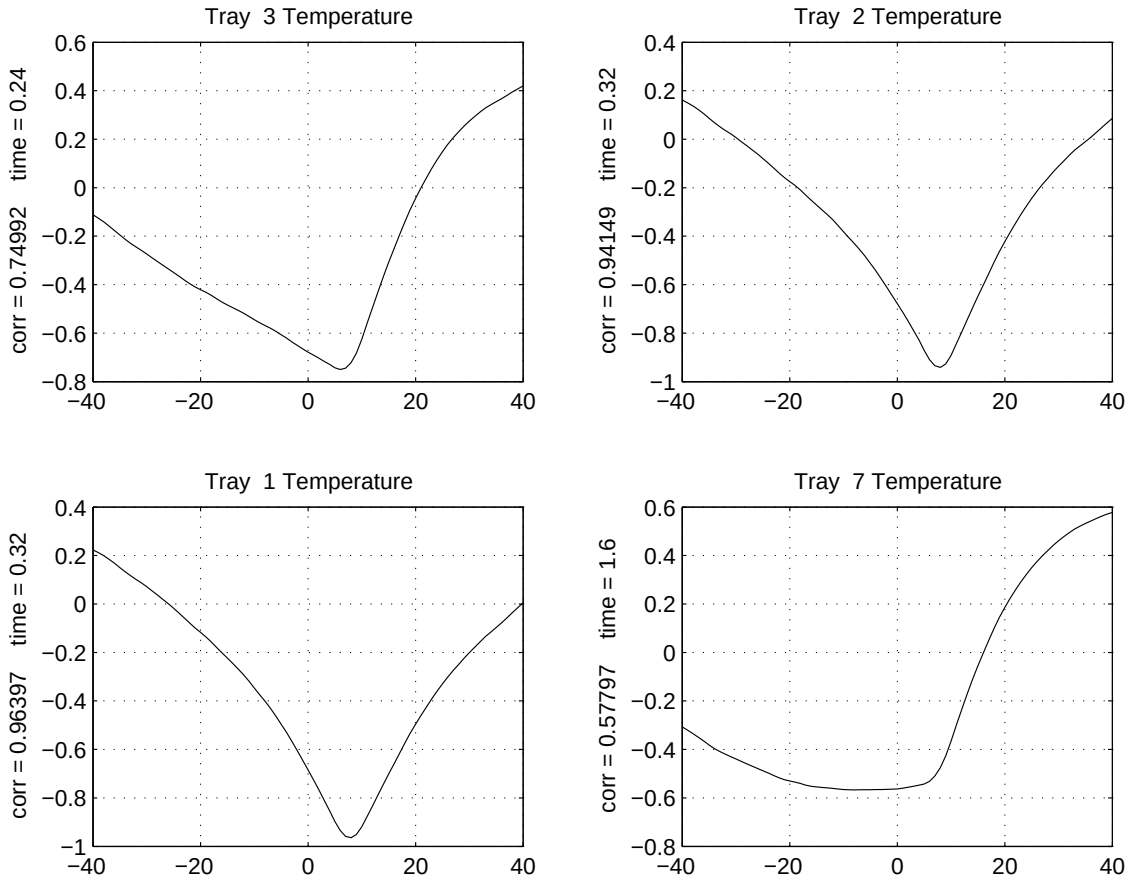

Figure 4.8: Cross correlation plots for Azeotropic (AZ) tower.

discusses this two step approach in greater detail. First issues related to the calculation of ordinary PCA for the combined database (historical operations data + plant test data) are discussed. This step helps reduce the dimensionality of the measurement variable space from m^* to r. Here, m^* is the number of measurements used to form the PCA model while r is the number of principal components retained in the PCA model. In the subsequent step, the ordinary PC scores are transformed so that they are collinear with the manipulated variable (MV) inputs. This operation further reduces the dimension of the variable space from r to n_u. Here, n_u is the number of MVs. This two step procedure was suggested by Zheng [59] and Zheng et al [60]. In order to better understand orthogonal PCA, Section 4.3.1 develops a connectivity between orthogonal PCA and PLS.

One of the critical decisions to be made during PCA model development is the determination of the number of leading principal components (PCs) to retain (i.e. r). PCA helps in reducing the dimensionality by eliminating the redundancy (i.e. linear dependance) in process measurements. PCA also filters out the noise and retains only the important information (i.e. useful dynamics) contained in the measurements. The choice of r dictates how efficiently these two goals are met. Too large a value of r would result in noise being incorporated in the model and as a result the model would be inaccurate. Too small a value of r would mean that the model does not contain enough useful information. Several rules of thumb that are commonly used to determine r are listed in section 1.3.4. One of the rules mentioned is that a knowledge of the number of sources of variation during database generation could be used to determine r. The historical operation data for the Tennessee Eastman (TE) process incorporated disturbances of many different types and directions. The corresponding data for the Azeotropic (AZ) tower however incorporated only the feed disturbances. During the plant testing of both processes, over and

above a small feed disturbance, the primary source of variation is the PRBS forcing of the manipulated input(s). Keeping these factors in mind, in this study, r was determined by adding 1 to the number of manipulated variables (MVs) perturbed during the identification experiment. In other words, $r = 2$ for the univariate cases and $r = 3$ when two MVs were being used.

4.3.1 Understanding orthogonal PCA

In order to understand the orthogonal PCA calculation used to obtain scores collinear with the manipulated variable (MV) inputs, the following discussion is restricted to single MV cases. The aim of the discussion is to demonstrate a connectivity between orthogonal PCA and partial least squares (PLS). Although the result obtained in the following analysis is strictly applicable only to single MV cases, it does provide a perspective that is helpful in understanding cases involving more than one MV.

Consider $\mathbf{Y}$, the $n_s \times n_y$ matrix of measurements and $\mathbf{u}$, the $n_s \times 1$ vector of manipulated inputs. In order to perform orthogonal PCA as per the two-step approach suggested by Zheng [59] and Zheng et al [60], the first step is to calculate ordinary PCA. As mentioned in section 1.3.4, if all the principal components are retained then ordinary PCA is equivalent to singular value decomposition (SVD):

$$\mathbf{Y} = \mu\Sigma V^T \tag{4.5}$$

where $\mu\Sigma$ are the ordinary PC scores and $\mathbf{V}^T$ are the ordinary PC loadings. The second step involves projecting the ordinary PC scores onto the range of $(\mu\Sigma)^T\mathbf{u}$ in order to obtain $\mathbf{z}$, the $n_s \times 1$ vector of scores collinear with the MV:

$$\mathbf{z} = (\mu\Sigma)\ \mathrm{orth}\left\{(\mu\Sigma)^T\mathbf{u}\right\} \tag{4.6}$$

Since $(\mu\Sigma)^T\mathbf{u}$ is an $n_y \times 1$ vector, the orthonormal basis for it's range space is given by:

$$\text{orth}\left\{(\mu\Sigma)^T\mathbf{u}\right\} = \frac{(\mu\Sigma)^T\mathbf{u}}{\|(\mu\Sigma)^T\mathbf{u}\|} \tag{4.7}$$

Hence, the $n_s \times 1$ vector of collinear scores is given by:

$$\mathbf{z} = (\mu\Sigma)\frac{(\mu\Sigma)^T\mathbf{u}}{\|(\mu\Sigma)^T\mathbf{u}\|} \tag{4.8}$$

Observing that $\mathbf{V}$ is an orthogonal matrix (i.e. $\mathbf{V}^T\mathbf{V} = \mathbf{I}$),

$$\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T = (\mu\Sigma\mathbf{V}^T)(\mathbf{V}\Sigma^T\mu^T)\mathbf{u}\mathbf{u}^T = \mu\Sigma\,(\mu\Sigma)^T\,\mathbf{u}\mathbf{u}^T \tag{4.9}$$

Substituting Equation 4.7 and Equation 4.8 into Equation 4.9:

$$\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T = \mu\Sigma\,\text{orth}\left\{(\mu\Sigma)^T\mathbf{u}\right\}\,\|(\mu\Sigma)^T\mathbf{u}\|\mathbf{u}^T = \mathbf{z}\,\|(\mu\Sigma)^T\mathbf{u}\|\mathbf{u}^T \tag{4.10}$$

and so

$$\mathbf{z} = \frac{\mathbf{Y}\mathbf{Y}^T\mathbf{u}}{\|(\mu\Sigma)^T\mathbf{u}\|} \tag{4.11}$$

In order to determine the eigenvalues and eigenvectors of $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$, $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$ is multipied by an arbitrary vector $\mathbf{b}$:

$$\left(\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T\right)\cdot\mathbf{b} = \mathbf{z}\,\|(\mu\Sigma)^T\mathbf{u}\|\mathbf{u}^T\mathbf{b} \tag{4.12}$$

It follows from Equation 4.12 that the orthogonal complement of $\mathbf{u}$ (i.e. all vectors perpendicular to $\mathbf{u}$) form the eigenspace of $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$ with eigenvalue zero. Moreover, $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T\mathbf{b}$ is parallel to $\mathbf{z}$ for all vectors $\mathbf{b}$. Therefore, $\mathbf{z}$ is an eigenvector of $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$ with eigenvalue $\|(\mu\Sigma)^T\mathbf{u}\|\mathbf{u}^T\mathbf{z}$. Using Equation 4.11 this eigenvalue can be written as $\mathbf{u}^T\mathbf{Y}\mathbf{Y}^T\mathbf{u} = \|\mathbf{u}^T\mathbf{Y}\|^2$. In summary, $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$ has only one non-zero eigenvalue $\|\mathbf{u}^T\mathbf{Y}\|^2$ with eigenvector $\mathbf{z}$. The other eigenvalue (of multiplicity $n_s - 1$) is zero. The eigenspace for this eigenvalue is the ($n_s - 1$ dimensional) orthogonal complement of $\mathbf{u}$. Note that it

has been implicity assumed that $\mathbf{u}^T\mathbf{Y} \neq 0$ in this discussion. This is a reasonable assumption physically because this condition amounts to choosing only such an MV that has a substantial effect on the measurements.

It has been reported (see for example de Jong [10]) that the first (i.e. leading) vector for the Y-block scores resulting from a partial least squares (PLS) computation is the first eigenvector of $\mathbf{Y}\mathbf{Y}^T\mathbf{X}\mathbf{X}^T$. If for a single MV case, the vector of manipulated input values (i.e. $\mathbf{u}$) is chosen as the X-block and the matrix of measurements (i.e. $\mathbf{Y}$) is chosen as the Y-block, then the first vector for the Y-block scores is the leading eigenvector of $\mathbf{Y}\mathbf{Y}^T\mathbf{u}\mathbf{u}^T$. It should be noted that for PLS calculations performed in this fashion, the weight vector $\mathbf{w} = 1$ and the X-block loadings $\mathbf{p} = 1$. Moreover, the X-block scores, $\mathbf{t}$, are the eigenvectors of $\mathbf{u}\mathbf{u}^T\mathbf{Y}\mathbf{Y}^T$ and so $\mathbf{t} = \mathbf{Y}$.

Hence, for cases involving one MV, collinear scores obtained using orthogonal PCA calculations such that all the principal components are retained are equivalent to the scores for the Y-block elements from a PLS calculation considering the $\mathbf{u}$ vector as the X-block and the measurements $\mathbf{Y}$ as the Y-block. Although this connectivity is strictly applicable only to cases involving a single MV, the analysis does demonstrate that orthogonal PCA is a 2-block procedure and is closely related to PLS. This characteristic of orthogonal PCA is in contrast with ordinary PCA, which a single block procedure for reducing the dimensionality of the measurement space.

4.4 Dynamic score model identification

Model identification is used to describe the process of determining a dynamic, input-output process model from empirical test data collected using a designed experiment conducted on the process. It is the most time consuming step in the implementation of

MPC and MPC-type controllers. For the controllers developed in this study, a discrete time linear dynamic model relating the response of collinear PC scores to the dynamics of the manipulated variables (MVs) is required. Specifically, a dynamic matrix (denoted as $\mathbf{A_D}$), to be used as an explicit part of the controller, is constructed from a finite step response (FSR) model which in turn is obtained from a finite impulse response (FIR) model. General guidelines for the design of the plant tests necessary to obtain the input-output data are given next.

4.4.1 Design of plant tests

Plant tests involve perturbing the process, running at steady state, using Pseudo Random Binary Sequence (PRBS) inputs and recording the outputs. The design of these experiments is dictated by the dynamics of the process and so it is a good idea to have a "ballpark" idea of the dynamics first. The first step is to determine the truncation number (n_t) and the sampling time (Δt). These two quantities are often determined simultaneously from the settling time of the process. The truncation number and sampling times are chosen such that the error in the steady state gain is less than 5% and the number of data points in the model is 30 to 50. Too small a truncation number, with respect to the true settling time of the process, could lead to model errors while too large a value would require a longer PRBS experiment and the size of the subsequent calculations (usually linear least squares) would be larger. Too short a sampling time would result in too many unnecessary manipulated variable changes which in turn would cause excessive control valve movement leading to equipment wear and hardware failure. The size of the model, in terms of the number of data points, would also be very large. Too large a sampling time could result in not enough information being captured and hence the control action

might not be enough. A disastrous consequence of too large a sampling time is that faster dynamics might not be captured at all and could also lead to *aliasing*. Aliasing is a consequence of the fact that more than one continuous signal, of different frequencies, can result in exactly the same sample values (see Franklin et al [15]).

The following two rules of thumb, which lead to similar results, are often used:

1. Select Δt as $\frac{1}{10}$th the dominant time constant of the process.

2. Select $\Delta t = \max \{0.003 \times \text{settling time}, 0.3 \times \text{dead time}\}$.

For higher order systems, for example distillation columns with multiple trays, it is a good idea to approximate the system as first order plus time delay (FOPTD). The settling time of the process would then be the sum of the first order time constant and the dead time. Once the truncation number (n_t) and the sampling time (Δt) are decided, one can decide the length of the identification experiment (denoted as N_e). In terms of data points, as a rule of thumb, $N_e \geq 10 \times n_t$.

These rules of thumb are not directly applicable to multirate and multivariable systems. When dealing with multirate systems, the sampling time is determined by the fastest loop and the truncation number is determined by the slowest loop. When dealing with multivariable systems, the identification calculations are done in a multi-input, single-output (MISO) fashion. It is possible, for example in a case with two manipulated variables (MVs), that the response of the output (i.e. measurement) to the first input (i.e. MV) is much larger in magntude (i.e. has a higher gain) than that to the second MV. In such a situation, if linear least squares is used to fit the experimental data to a model, the response for the MV with the smaller gain will have much larger relative errors. The errors for the response to the MV with the larger gain will dominate the calculations used

to fit the data to the model. As a result the relative errors for the MV with the larger gain will be smaller while the errors for the MV with the smaller gain will be larger. Since in this study, scaled MVs are being used instead of their absolute values, it is recommended that the MV that has a larger gain be perturbed at a lower frequency during the PRBS experiment. If this reduction in the frequency is not possible then it is recommended that a variance scaling smaller than that necessary for unit variance be used for the MV with the larger gain.

Another critical decision to be made is the size of the perturbation in the MVs. Too large an input could lead to non-linearities being incorporated and hence give rise to inaccurate models. Too large an input could also destabilize the process and this would obviously be unacceptable from economic and safety perspectives. If the size of the MV movement is too small, it's effect on the measurements would be very small and it is possible that the field sensors are unable to pick up the resulting deviations. Computer simulations predict deviations that are unrealistic to measure in the real world and so do not expose the problem of sensor insensitivity. Hence, it is not a good idea to decide the input sizes using simulations alone. However, computer simulations do indicate process gains and coupled with the sensor manufacturer's specifications, can be helpful in determining sensor insensitivity. If the size of the perturbation is too small, this would also have a very poor signal to noise ratio. Signal to noise ratio (S/N ratio) is defined as:

$$S/N = \frac{\sigma_s^2}{\sigma_N^2} \tag{4.13}$$

where, σ_s is the standard deviation of the signal while σ_N is the standard deviation of the noise. Practitioners recommend a S/N of at least 6 (see Brosilow and Joseph [3]).

4.4.2 Least squares calculations for model identification

For a multivariable system, a multi-input single-output (MISO) model is constructed for each output and these models are then arranged accordingly. The rth output, i.e. measurement ($r = 1, 2, \ldots, n_y$, where n_y is the number of output measurements) at the kth sampling instant can be predicted using:

$$\hat{y}_r(k) = \sum_{m=1}^{n_u} \sum_{i}^{ns} h_{r,m,i} u_m(k-i) \tag{4.14}$$

Writing the above equation for time k and $k-1$ and upon differencing, the following relation between the output measurement ($y_r(k)$) and it's prediction ($\hat{y}_r(k)$) is obtained:

$$\Delta y_r(k) = \Delta \hat{y}_r(k) + \nu_r(k) = \sum_{m=1}^{n_u} \sum_{i}^{ns} h_{r,m,i} \Delta u_m(k-i) + \nu_r(k) \tag{4.15}$$

The $\nu_r(k)$ included in Equation 4.15 collects any effects on the rth output measurement not described by the model. It is assumed that ν_r is well represented as white noise (a sequence of random variables with zero mean). A bias in the model estimates results when the white noise is intergrated, i.e. passes through a process element that exhibits integrating behavior and shows up at the output. The differencing carried out in Equation 4.15 is done to tackle this problem of bias by converting intergrated white noise into white noise with zero mean. It should be noted that in this study, collinear PC scores are used as outputs and their dimension is equal to the dimension of the manipulated inputs, i.e. $n_y = n_u$. The values of the measured past outputs from $n_t + 1$ to $N_e + n_t$ are collected and rearranged into a vector:

$$\mathbf{y}_{N,r} = [\Delta y_r(n_t+1)\ \Delta y_r(n_t+2)\ \cdots\ \Delta y_r(n_t+N_e)]^T \tag{4.16}$$

Note that the first n_t data points for the collinear scores will not be used in the calculations. The unmeasured noise is given by:

$$\nu_{N,r} = [\nu_r(n_t+1) \;\; \nu_r(n_t+2) \;\; \cdots \;\; \nu_r(n_t+N_e)]^T \tag{4.17}$$

The vector of parameters to be identtified (i.e. the vector of impulse response coefficients corresponding to the rth output measurement denoted by $\mathbf{\Theta_r}$) is given by Equation 4.21. A matrix of the known past manipulated variable inputs ($\mathbf{\Phi_N}$) is constructed using Equation 4.22.

Hence, in order to obtain $\mathbf{\Theta_r}$, the following set of linear equations have to be solved:

$$\mathbf{y}_{N,r} = \mathbf{\Phi_N \Theta_r} + \nu_{N,r} \tag{4.18}$$

Equation 4.18 is the most basic equation used to obtain a finite impulse response (FIR) model from experimental data. The objective of the calculation, referred to as a linear least squares (LLS) fit, is to select the $\mathbf{\Theta_r}$ that minimize the square of the norm of the residual vector $\nu_{N,r}$:

$$\min_{\mathbf{\Theta_r}} \; \nu_{N,r}^T \nu_{N,r} \tag{4.19}$$

Let the length of the experiment N_e be greater than the total number of parameters in $\mathbf{\Theta_r}$ (i.e. greater than $n_u \times n_t$). Then the solution of Equation 4.18, i.e. the choice of $\mathbf{\Theta_r}$ that minimizes the quantity $(y_{N,r} - \mathbf{\Phi_N \Theta_r})^T (y_{N,r} - \mathbf{\Phi_N \Theta_r})$ can be found from:

$$\hat{\mathbf{\Theta}}_\mathbf{r} = \left[\mathbf{\Phi_N}^T \mathbf{\Phi_N}\right] \mathbf{\Phi_N Y_N} \tag{4.20}$$

Here the LLS calculation used to find the estimate, $\hat{\mathbf{\Theta}}_\mathbf{r}$, weighs all the samples equally and gives equal importance to all the coefficients (i.e. data points) in the FIR model. The next section discusses an augmented form of fitting the experimental data to the model, the weighted linear least squares calculation.

$$\mathbf{\Theta_r} = [h_{r,1}(1)\ h_{r,1}(2)\ \cdots\ h_{r,1}(n_t)\ h_{r,2}(1)\ h_{r,2}(2)\ \cdots\ h_{r,2}(n_t)\ \cdots h_{r,n_u}(1)\ h_{r,n_u}(2)\ \cdots\ h_{r,n_u}(n_t)]^T \tag{4.21}$$

$$\mathbf{\Phi_N} = \begin{bmatrix} \Delta u_1(n_t) & \cdots & \Delta u_1(1) & \Delta u_2(n_t) & \cdots & \Delta u_2(1) & \cdots & \Delta u_{n_u}(n_t) & \cdots & \Delta u_{n_u}(1) \\ \Delta u_1(n_t+1) & \cdots & \Delta u_1(2) & \Delta u_2(n_t+1) & \cdots & \Delta u_2(2) & \cdots & \Delta u_{n_u}(n_t+1) & \cdots & \Delta u_{n_u}(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta u_1(n_t+N_e-1) & \cdots & \Delta u_1(N_e) & \Delta u_2(n_t+N_e-1) & \cdots & \Delta u_2(N_e) & \cdots & \Delta u_{n_u}(n_t+N_e-1) & \cdots & \Delta u_{n_u}(N_e) \end{bmatrix} \tag{4.22}$$

4.4.3 Weighting methods

In keeping with the philosophy of developing a framework that is simple yet effective, the following three points are noted:

1. The primary aim of the MP-SPC and related controllers is to deal with stationary disturbances. As a result, the models used need not have a very high steady state accuracy. In other words, greater emphasis should be placed on short term prediction during model identification.

2. The model that needs to be identified relates the dynamics of the PC scores to the manipulated variables. There is a Q-residual associated with each sample used to generate the PC model and this quantity is partly indicative of the process noise, the underlying non-linearities and loss of accuracy due to dimension reduction. Residual information can be used in two ways:

   (a) During the data pre-procesing step, samples with large values of the Q-residual can be considered as outliers and removed from the data set used to generate the PCA model.

   (b) Samples (i.e. data points) with small Q values are likely to be more accurate and should be weighted more and those with large Q-values should be weighted less during the identification calculations.

3. There is also a Hotelling's statistic (also known as the T^2 statistic) associated with each sample. A sample with a high value of T^2 statistic has a greater leverage (i.e. influence) on the PCA model and so such a sample should be weighed more during the calculations.

In order to incorporate these three objectives, a weighted, linear least squares identification procedure is adopted. The first weighting matrix (Λ_θ) penalizes the magnitudes of the impulse response coefficients (i.e. sharp changes in the step response coefficients). As a result, there is a trade-off: a smoother step response is obtained at the expense of accuracy. Small values in the initial portion of the impulse response coefficients (i.e. from 1 through P, where P is the most likely prediction horizon) and large values in the latter portion of the FIR (i.e. from $P+1$ through n_t, where n_t is the truncation number) would generate an impulse response model which emphasizes short term prediction at the expense of steady state accuracy. The second weighting matrix (Λ_v) penalizes the weighted squares of the residuals. Higher weights are assigned to data points that are believed to be more accurate i.e. have a smaller Q residual. Hence, a matrix of impulse response coefficients ($\mathbf{\Theta_r}$) is found such that the quantity

$$(y_{N,r} - \mathbf{\Phi_N \Theta_r})^T \Lambda_v^T \Lambda_v (y_{N,r} - \mathbf{\Phi_N \Theta_r}) + \mathbf{\Theta_r}^T \Lambda_\theta^T \Lambda_\theta \mathbf{\Theta_r}$$

is minimized. This minimization can be performed by solving the following set of linear equations:

$$\begin{bmatrix} \Lambda_v Y_N \\ \emptyset \end{bmatrix} = \begin{bmatrix} \Lambda_v \Phi_N \\ \Lambda_\theta \end{bmatrix} \Theta$$

The above set of linear equations can be solved using multiple linear regression (MLR) or partial least squares (PLS). However, since for this study the manipulated input moves have been designed to be independent of each other and the measured outputs are PC scores which have the property of being orthogonal to one another, i.e. independent, both MLR and PLS would give identical results.

4.4.4 Results and discussion

Using the guidelines mentioned in section 4.4.1 and reasonable starting guesses, the sampling time (Δt), the truncation number (n_t) and the size of the perturbations for the identification of finte impulse response (FIR) models are determined iteratively. The sampling time used for both the TE process and the AZ case studies is 2 minutes. It should be noted that McAvoy [31] used a sampling time of 5 minutes when he applied MP-SPC to the TE process. However, some measurements, for example A-feed flow, exhibit aliasing for this large a sampling time. Due to the multirate nature of the processes, the truncation number for both cases is more than the recommended 30 to 50. The FIR for the TE process has 60 data points while that for the AZ tower has 400. The size of the inputs was decided based on their steady state values: 3664 kg/hr for D-feed SP and 75% for R-level SP in case of the TE process. The corresponding values for the AZ tower are 4.985 kmol/min for FR-SP and 110 °C for T5-SP. The input perturbations are ±1% of their steady state values for both case studies. However, for the TE process, an input size of 2% was necessary when a single MV was used because an input size of 1% did not provide enough excitation.

Plant test data for the Tennessee Eastman (TE) process and the Azeotropic (AZ) are generated as follows. Designed experiments are conducted on the two test case processes by varying the manipulated variables (MVs) in a pseudo random binary sequence (PRBS). These tests are summarized in Table 4.4 and Table 4.5. A small disturbance in the feed is present during the experiment. During the subsequent calculations this disturbance is not modeled and it is treated as noise. In order to study the effectiveness of the various weighting methods in dealing with different data-sets, the experiments conducted have varying degrees of noise levels and are of a longer or shorter duration. The noise

Table 4.4: Plant tests conducted on the Tennessee Eastman (TE) using PRBS inputs.

Experiment Number	PRBS input		Duration (min)	S/N ratio
	D-feed SP (kg/hr)	R-level SP (%)		
1	± 36.64	± 0.75	1260	45.9688
1a	± 36.64	± 0.75	1260	3.6524
1b	± 36.64	± 0.75	660	2.5464
2	± 73.28	0	1260	29.9051
2a	± 73.28	0	1260	1.6110
2b	± 73.28	0	660	1.7437
3	0	± 1.5	1260	254.4559
3a	0	± 1.5	1260	12.5255
3b	0	± 1.5	660	5.2434

level is varied by changing the size of the feed disturbanace. A signal to noise ratio (i.e. S/N ratio) greater than 6 is considered good. An experiment is considered longer if the duration of that experiment is greater than the minimum duration recommended. For the TE process (see Table 4.4), the D-feed SP (Experiment numbers 2, 2a and 2b) and the Reactor-level SP (Experiment numbers 3, 3a and 3b) are the two single MV cases while the two MVs used in combination (Experiment numbers 1, 1a and 1b) is the multivariable case. Correspondingly, for the AZ tower (see Table 4.5), the Reflux flow SP (Experiment numbers 2, 2a and 2b) and the fifth tray temperature SP (Experiment numbers 3, 3a and 3b) are the two single MV cases while the two MVs used in combination (Experiment numbers 1, 1a and 1b) is the multivariable case. For each of the three MV cases, the first experiment (numbers 1, 2 and 3) is of longer duration with a low noise level, the second experiment (numbers 1a, 2a and 3a) is of a longer duration with a high noise level and the third experiment (1b, 2b and 3b) has a shorter duration and a high noise level. In this way, for each of the two case studies, nine PRBS experiments have been conducted.

Figure 4.9 and Figure 4.10 compare the various weighting methods used during the weighted linear least squares (WLLS) calculation. The notation used is detailed in Table

Table 4.5: Plant tests conducted on the Azeotropic (AZ) tower using PRBS inputs.

Experiment Number	PRBS input		Duration (min)	S/N ratio
	FR-SP (kmol/min)	T5-SP (°C)		
1	± 0.05	± 1.1	8000	3579.6
1a	± 0.05	± 1.1	8000	135.63
1b	± 0.05	± 1.1	4000	128.56
2	± 0.05	0	8000	3675.7
2a	± 0.05	0	8000	168.28
2b	± 0.05	0	4000	145.22
3	0	± 1.1	8000	24.06
3a	0	± 1.1	8000	1.79
3b	0	± 1.1	4000	1.8

Table 4.6: Notation used to describe different weighting methods used during WLLS calculations.

Notation	Weight	Description
V=0	$\Lambda_v = \mathbf{I}$	All samples are equally important (no weighting).
V=1	$\Lambda_v = diag(\mathbf{T^2})$	Samples having a larger Hotelling's statistic are more important.
V=2	$\Lambda_v = diag(\mathbf{Q^{-1}})$	Samples having a smaller Q-residual are more important.
T=0	$\Lambda_\theta = \emptyset$	All coefficients are equally important (no weighting).
T=1	$\Lambda_\theta = 30 \times \mathbf{I}$	Uniformly penallizing sharp changes in coefficients.
T=2	$\Lambda_{\theta,1} = 10 \times \mathbf{I} \quad \Lambda_{\theta,2} = 30 \times \mathbf{I}$	Smaller penalty $\Lambda_{\theta,1}$ for sharp changes in initial values of coefficients

4.6. A summary of the results that helps compare the various weighting methods is givem in Table 4.7 and Table 4.8. The root mean square (RMS) of fitness is reflective of how well the model fits the experimental data while the RMS of validation reflects how accurately the model fits the true step response. It can be seen that:

1. There is no consistent trend regarding the superiority of a particular weighting method over others. However, on certain occasions, an improved model is obtained even when the signal to noise ratio is poor and/or when the length of the identification experiment is short. For example, for the Tennessee Eastman (TE) process, for experiment numbers 1a and 1b, using the T^2 statistic as the sample weight results in a model with the smallest RMS of validation eventhough the corresponding RMS of fitness value is worse. Similarly, for experiment numbers 2a and 2b for the Tennessee Eastman (TE) process and experiment number 3a for the Azeotropic tower (AZ), using the Q-residual as a sample weight results in the smallest RMS of validation.

2. The Q-residual incorporates both the effect of the loss of information due to reduction in dimensionality and the presence of noise in a particular sample. If the Q-residual accounts for primarily the noise, then using it's inverse would be a good idea but if the Q-residual accounts for the loss of information due to a reduction in dimensionality then it should not be used as a sample weight at all. It is not possible to separate out the two effects and so the use of the Q-residual as a sample weight has mixed results. Hence, in certain situations it is advantageous to use it's inverse as a sample weight but on certain occasions the predictive ability of the model gets diminished if the inverse of Q-residual is used as a sample weight.

3. For short experiments and when the signal to noise ratio is poor, weighting the

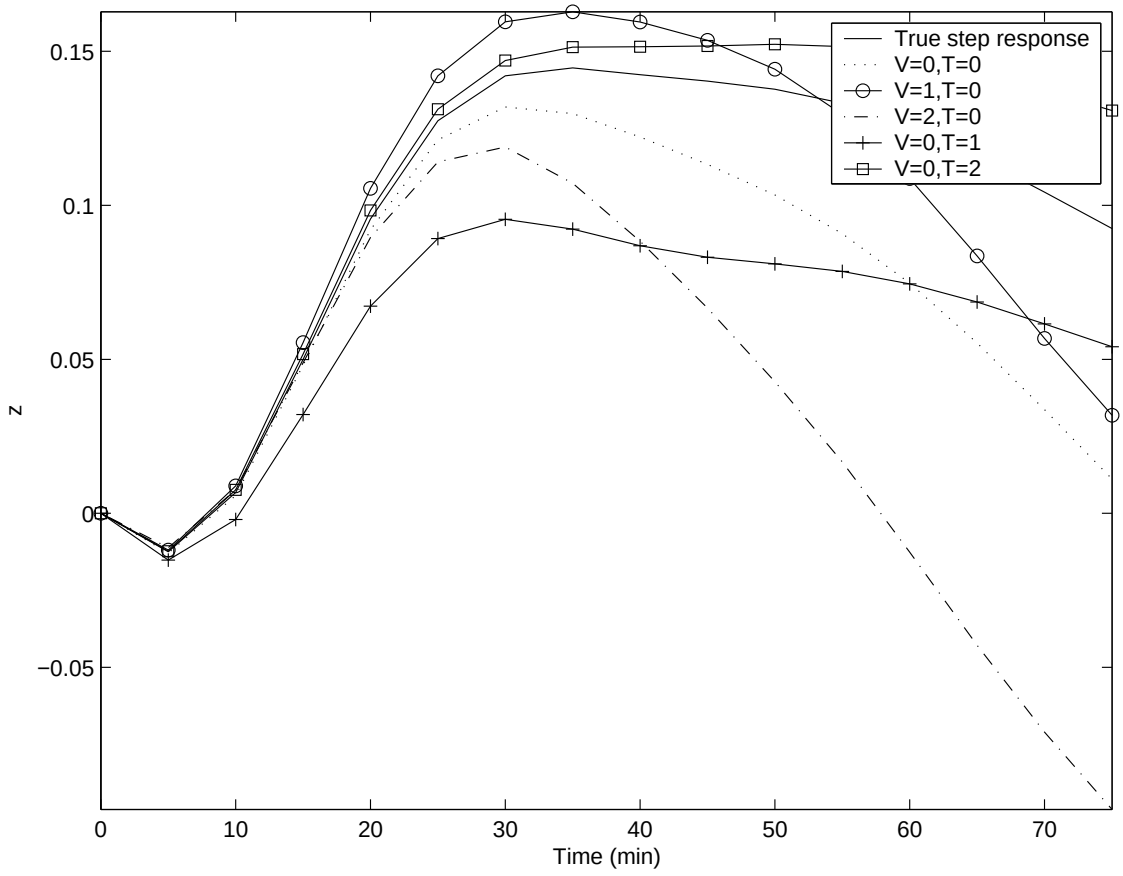

Figure 4.9: Comparison of weighting methods for WLLS identification for TE.

impulse response coefficients improves the model predictions. This is because, for such situations, noise begins to get incorporated into the impulse response coefficient values. As a result the coefficients are overpredicted. This effect can be countered by weighting the coefficients.

Since MP-SPC is being implemented on top of a base control system already in place, all the step response models exhibit underdamped second order behavior. All the models for control are those identified using data from long experiments with good S/N ratios and without any weighting. Figure 4.11 shows the finite step response (FSR) models for the two single manipulated variable (MV) cases for the Tennessee Eastman (TE) process. For both MVs, a comparison of the identified FSR model and the "true" model obtained

Table 4.7: Comparison of weighting methods for the Tennessee Eastman (TE) process.

Exp. No.	Duration	S/N ratio	V	T	RMS(Fitness)	RMS(Validation)
1	1260	45.97	0	0	0.3326	0.0132
			1	0	0.3732	0.0052
			2	0	0.3577	0.0085
			0	1	0.3177	0.0178
			0	2	0.3191	0.0137
1a	1260	3.65	0	0	1.0299	0.0711
			1	0	1.0862	0.0742
			2	0	1.1096	0.0667
			0	1	1.0138	0.0708
			0	2	1.0085	0.0780
1b	660	2.5	0	0	1.7404	0.1035
			1	0	1.7674	0.1021
			2	0	1.9527	0.1002
			0	1	1.1639	0.0910
			0	2	1.0971	0.1359
2	1260	29.9	0	0	0.2100	0.0073
			1	0	0.2281	0.0294
			2	0	0.2411	0.0145
			0	1	0.2757	0.0330
			0	2	0.2824	0.0543
2a	1260	1.61	0	0	1.0692	0.0346
			1	0	1.0878	0.0182
			2	0	1.3209	0.2525
			0	1	1.1148	0.0389
			0	2	1.1039	0.0464
2b	660	1.74	0	0	1.1308	0.0919
			1	0	1.2057	0.0654
			2	0	1.3727	0.1598
			0	1	1.1612	0.0747
			0	2	1.1277	0.0304
3	1260	254.5	0	0	0.4699	0.0014
			1	0	0.4922	0.0017
			2	0	0.5045	0.0016
			0	1	0.4657	0.0014
			0	2	0.4655	0.0014
3a	1260	12.5	0	0	0.4003	0.0132
			1	0	0.5795	0.0156
			2	0	0.5690	0.0120
			0	1	0.4005	0.0133
			0	2	0.4025	0.0132
2b	660	5.24	0	0	0.3816	0.0217
			1	0	0.4240	0.0182
			2	0	0.4659	0.0183
			0	1	0.3793	0.0217
			0	2	0.3771	0.0215

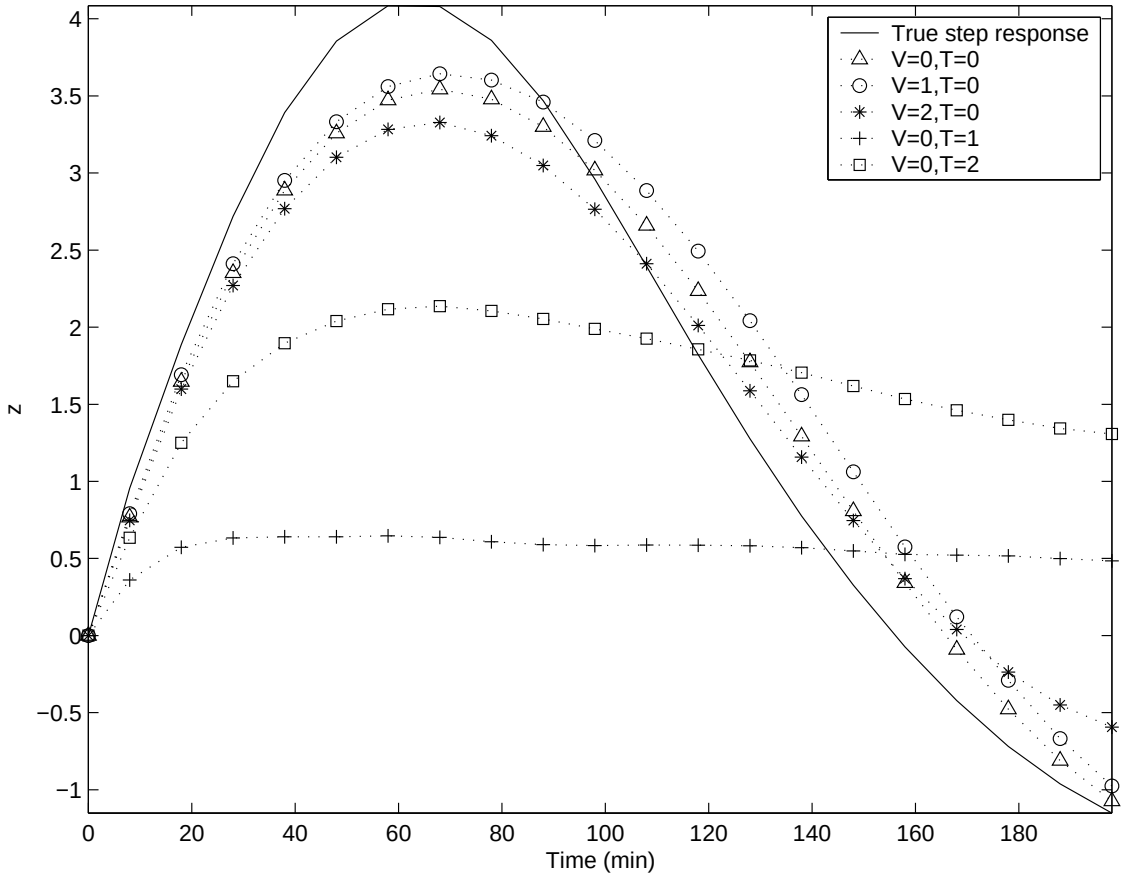

Figure 4.10: Comparison of weighting methods for WLLS identification for AZ.

Table 4.8: Comparison of weighting methods for the Azeotropic (AZ) tower.

Exp. No.	Duration	S/N ratio	V	T	RMS(Fitness)	RMS(Validation)
1	8000	3579.6	0	0	0.31196	0.1245
			1	0	1.1628	1.0611
			2	0	0.6249	0.2939
			0	1	1.484	0.8115
			0	2	1.0682	0.93
1a	8000	135.63	0	0	0.7511	0.9521
			1	0	3.0411	1.3782
			2	0	1.2897	0.5790
			0	1	2.6877	0.9421
			0	2	1.9159	1.4744
1b	4000	128.56	0	0	0.5594	0.7893
			1	0	2.9479	1.412
			2	0	1.1304	1.2397
			0	1	2.9052	0.8290
			0	2	1.9159	1.4744
2	8000	3675.7	0	0	0.5679	0.9461
			1	0	1.618	2.9544
			2	0	0.6953	0.7523
			0	1	2.5335	2.6564
			0	2	1.7632	2.5763
2a	8000	168.28	0	0	0.5742	5.0681
			1	0	1.0255	5.2351
			2	0	0.6740	5.0832
			0	1	2.5545	3.2843
			0	2	1.7427	4.3066
2b	4000	145.2	0	0	0.2758	1.2027
			1	0	0.5624	1.235
			2	0	0.6168	1.5294
			0	1	2.4974	2.7204
			0	2	1.7438	2.6264
3	8000	24.06	0	0	0.0212	0.0084
			1	0	0.0345	0.0239
			2	0	0.0924	0.236
			0	1	0.1996	0.2821
			0	2	0.1256	0.1591
3a	8000	1.79	0	0	0.0676	0.2348
			1	0	0.1600	0.1851
			2	0	0.5371	1.7298
			0	1	0.2474	0.2815
			0	2	0.1851	0.1317
3b	4000	1.8	0	0	0.0912	0.0382
			1	9	0.1788	0.1772
			2	0	0.5775	1.6031
			0	1	0.2738	0.2991
			0	2	0.2056	0.1964

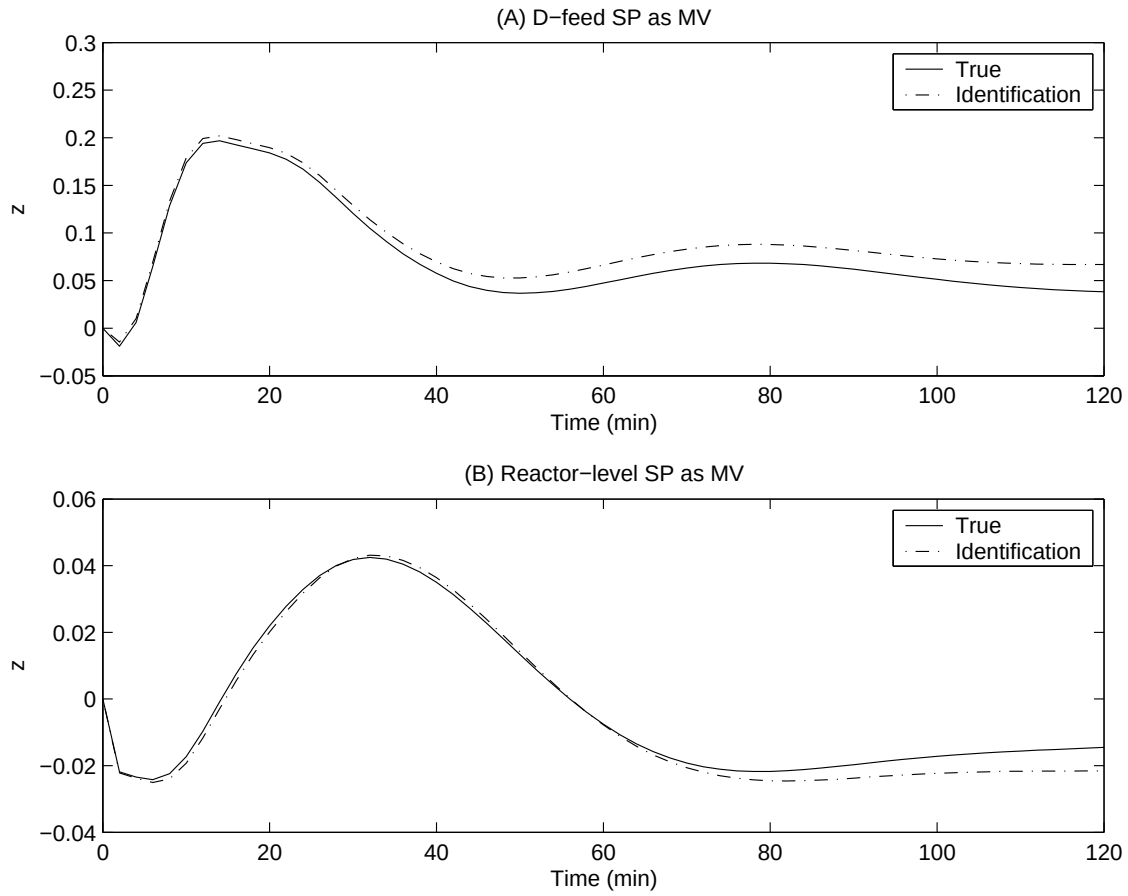

Figure 4.11: Identification of FSR models for Tennessee Eastman (TE) process: One manipulated variable cases.

using a hypothetical step input shows very good agreement. It can be seen that when the Reactor-level SP is being used as the MV, a very significant inverse response is seen. As seen earlier, this inverse response severely retards the performance of MP-SPC when the Reactor-level SP is being used as the MV. Figure 4.12 shows the single MV FSR models for Azeotropic (AZ) tower. It is seen that the process gain is very small compared to the very large overshoot when the Reflux-flow SP (FR-SP) is being used as the MV. Also, there is a very significant inverse response in this case. On the other hand, when the fifth tray temperature SP (T5-SP) is being used as the MV, there is no inverse response.

Figure 4.13 shows the FSR models for the Tennessee Eastman (TE) process when

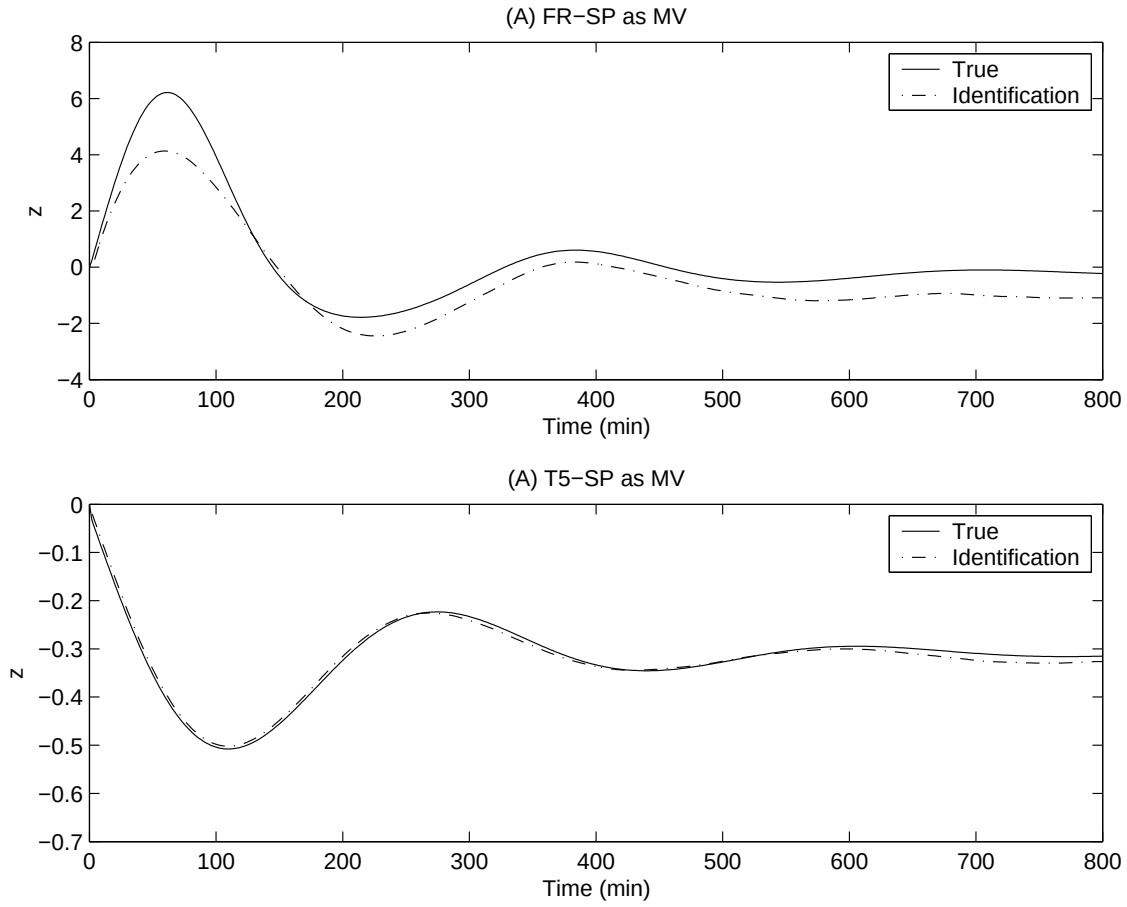

Figure 4.12: Identification of FSR models for Azeotropic (AZ) tower: One manipulated variable cases.

both the MVs are being used in combination. It can be seen that compared to the single MV cases, the model fit is not as good. However, the identified model agrees with the true response in the intial period (0 to 40 minutes) reasonably well and this accuracy is sufficient enough for MP-SPC. It should also be noted that there is a significant "wrong-way" behavior in both the score variables, in the response to the Reactor-level SP. As a result, as seen earlier, these two channels are detuned using a very large value of the corresponding move suppression coefficients (i.e. $\mathbf{\Lambda}$). The two MV FSR models for the Azeotropic (AZ) tower are shown in Figure 4.14. It can be seen that the response of $z(1)$ to Reflux-flow SP (FR-SP) exhibits inverse behavior and that of $z(2)$ has a very large overshoot but a very small gain. Hence both channels related to FR-SP are detuned. The response to the fifth tray temperature SP (T5-SP) is acceptable but the model fit is not very good even for some of the initial period. As a result, for this case study, the addition of a second MV does not impart any performance advantages but due to increased mismatch betweeen the model and the true response, the performance of MP-SPC actually deteriorates.

4.5 Conclusions

This chapter deals with secondary measurement selection, collinear PC score model development and dynamic model identification issues. The implementation of a novel approach that combines historical operation data with data from plant testing to build a database that could be used for PCA modeling is demonstarted.

The measurement selection procedure adopted here consists of two stages. The first stage involves manually screening out measurements which (1) do not exhibit adequate sensitivity to the manipulated input, (2) are not consistently correlated with the quality

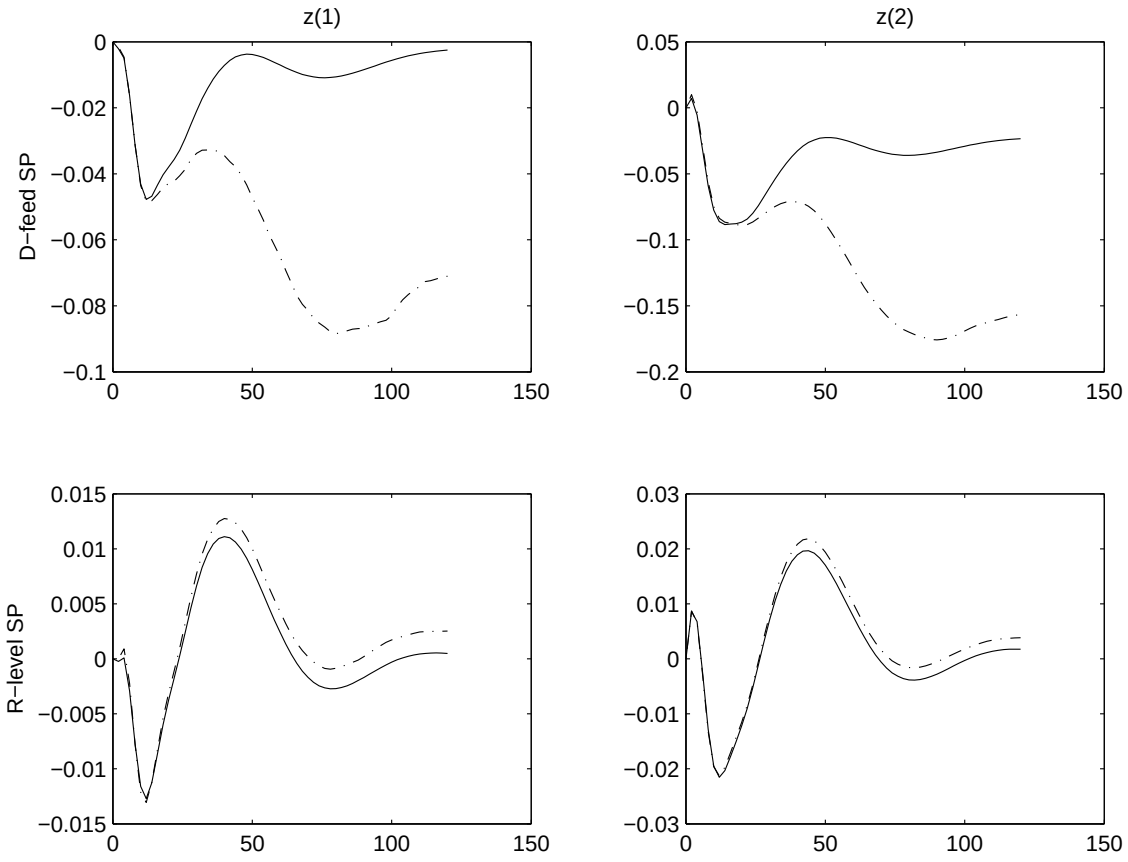

Figure 4.13: Identification of FSR models for Tennessee Eastman (TE) process: Two manipulated variable case.

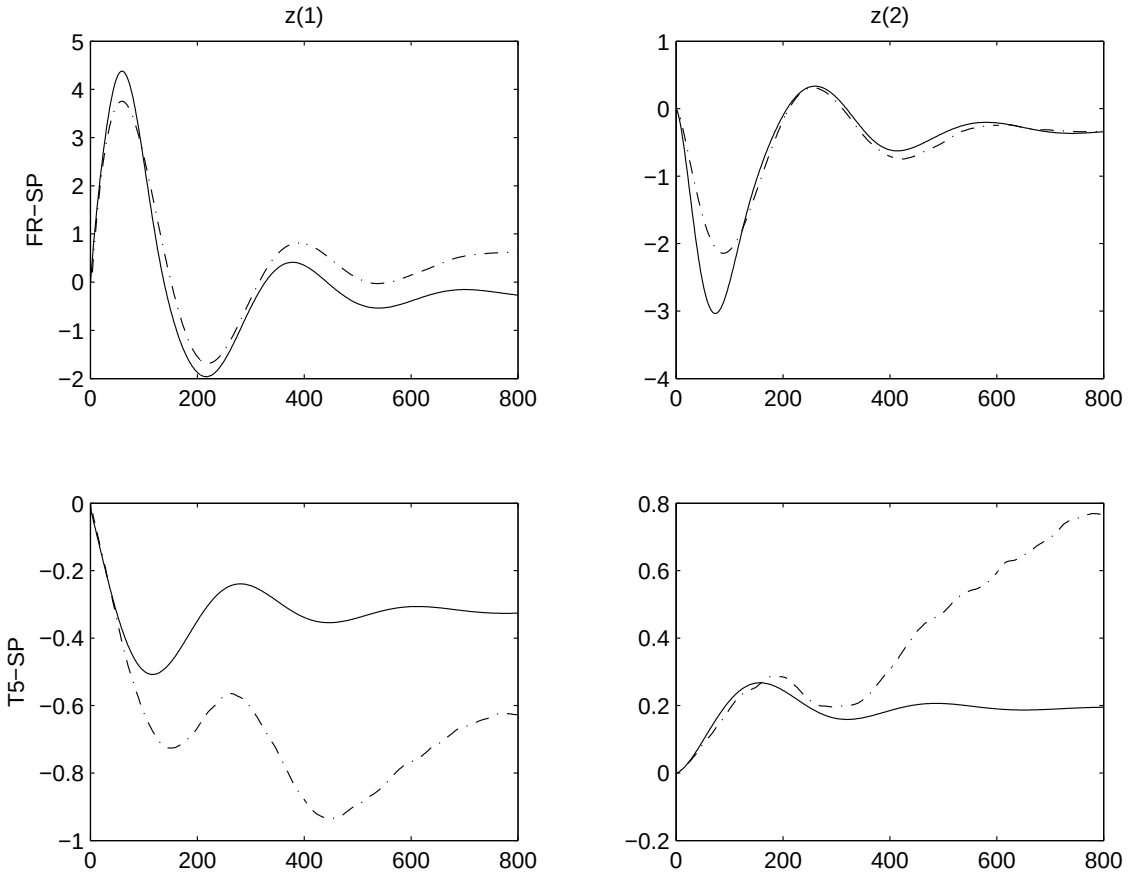

Figure 4.14: Identification of FSR models for Azeotropic (AZ) tower: Two manipulated variable case.

variable, and (3) exhibit inverse response. The second stage consists of two steps. In the first step, measurements are ranked according to the maximum value of their cross-correlation coefficient (with respect to the quality variable) and the time at which this maximum occurs. The second step involves an iterative procedure in which higher ranked measurements are retained in the measurement set as long as their contribution to the PCA model is sufficiently large.

The aim of the identification step is to develop a finite impulse response (FIR) model that relates the dynamics of the collinear PC scores to the inputs, i.e. MV moves. It is shown how a weighted linear least squares calculation can be adopted to develop accurate models even when the S/N ratio for the plant test data is very poor.

Chapter 5

Alternative controller formulations and extensions

A promising new approach to quality control for continuous processes in the absence of on-line quality or disturbance measurements, termed model predictive statistical process control (MP-SPC), is outlined in Chapter 3. In this chapter, an alternative to the basic MP-SPC formulation for the case when an on-line analyzer for the product composition is available is discussed first. The influence of disturbance size, frequency and direction on the performance of different controller formulations is analyzed next. Finally, extensions to the basic controller formulation that help deal with special disturbance scenarios are discussed.

5.1 On-line analyzer cascaded on score controller (MP-SPC + ACSC)

The basic model predictive statistical process controller (MP-SPC) formulation is developed under the premise that neither the on-line analyzer for detecting the source of the disturbance (i.e. fluctuations in the feed compositions) nor the one for detecting the effect of the disturbance (i.e. fluctuations in the product compositions) is available. The alternative formulation being studied in this section, henceforth referred to as MP-SPC + ACSC, relaxes the earlier premise and assumes that an on-line analyzer that measures the compostion of the product stream is available. This measurement is used to alter the setpoint for the score controller (i.e. MP-SPC) in a cascaded proportional integral (PI) control arrangement. The PI controller is implemented in the velocity form:

$$\Delta u(t) = K_c \left[\varepsilon(t) - \varepsilon(t-1) + \varepsilon(t) \frac{\Delta t}{T_R} \right] \tag{5.1}$$

Table 5.1: Comparative performance of Base control, Base control + analyzer, MP-SPC and MP-SPC + ACSC in dealing with stationary disturbances for the Tennessee Eastman (TE) process.

Formulation	D-feed SP as MV	R-level SP as MV	Both
Base control	2.9353×10^{-4}		
Base control + analyzer	3.4479×10^{-4}		
MP-SPC	1.1414×10^{-4}	2.5353×10^{-4}	7.2544×10^{-5}
MP-SPC + ACSC	1.0054×10^{-4}	2.5170×10^{-4}	6.4770×10^{-5}

where $\Delta u(t)$ is the change in the manipulated variable while $\varepsilon(t)$ is the error at time t. The two controller tuning parameters are K_c, the controller gain, and T_R, the reset (i.e. integral) time. Here Δt refers to the integration time step.

The performance of the MP-SPC + ACSC configuration is evaluated as follows. For the Tennessee Eastman (TE) process, the response of various controller configurations to IDV(8), i.e. a random fluctuation in the C-feed (i.e. stream 4) is simulated for 2500 minutes. Two univariate cases involving D-feed SP alone or Reactor level SP alone being used as the manipulated variable (MV) and one multivariate case where both MVs are used in combination, are considered. Table 5.1 gives a comparison of the variance in the product quality variable (G/H ratio in the product) for the Base control, Base control + analyzer, MP-SPC and MP-SPC + ACSC formulations. Similarly, in order to consider the scenario involving non-stationary disturbances, a step change in C-feed composition, termed IDV(1), is assumed to become active after 10 hours (i.e. 600 minutes). This disturbance is over and above the IDV(8). Table 5.2 summarizes the performance of various configurations simulated for 3100 minutes. It can be seen that, in general, adding an analyzer to MP-SPC is advantageous and that the average % improvement over and above that provided by MP-SPC is about 10%.

As regards the Azeotropic (AZ) tower, the responses of the Base control system

Table 5.2: Comparative performance of Base control, Base control + analyzer, MP-SPC and MP-SPC + ACSC in dealing with non-stationary disturbances for the Tennessee Eastman (TE) process.

Formulation	D-feed SP as MV	R-level SP as MV	Both
Base control	4.0444×10^{-4}		
Base control + analyzer	4.5419×10^{-4}		
MP-SPC	1.7263×10^{-4}	3.4341×10^{-4}	7.3140×10^{-5}
MP-SPC + ACSC	1.9434×10^{-4}	3.4045×10^{-4}	1.9764×10^{-4}

alone, the Base control system + on-line analyzer, MP-SPC and MP-SPC + ACSC configurations to stationary and non-stationary random disturbances in the feed to the column are simulated. The Reflux-flow setpoint (FR-SP) and the Fifth tray temperature setpoint (T5-SP) are the two manipulated variables (MVs) for the univariate configurations while for the multivariate case, both of these MVs are used in combination. Table 5.3 summarizes 2500 minutes of the response for the stationary case. Again, it can be seen that the MP-SPC + ACSC configuration performs about 15% better than MP-SPC. As mentioned in Section 3.2.3, when the Azeotropic (AZ) tower encounters a non-stationary disturbance, the tower shifts to a new steady state. MP-SPC continues to perform better than the Base control system (with or without the analyzer for the bottoms composition) at this new steady state. However, if the MP-SPC + ACSC configuration is used in this case, the PI controller that is cascaded on top of MP-SPC and has been tuned for the stationary disturbance shifts the setpoint for the score variable in such a way that this action ultimately results in destabillizing the tower. Hence, it is not a good idea to use MP-SPC + ACSC in cases where the process shifts to a new steady state and there is a potenial for instability due to the highly nonlinear nature of the process. In such situations, it is a good idea to detect the non-stationary and to switch off the controller during the transient period. Section 5.3.2 discusses alternative approaches to deal with such special scenarios.

Table 5.3: Comparative performance of Base control, Base control + analyzer, MP-SPC and MP-SPC + ACSC in dealing with stationary disturbances for the Azeotropic (AZ) tower.

Formulation	FR-SP as MV	T5-SP as MV	Both
Base control	3.0249×10^{-5}		
Base control + analyzer	3.0249×10^{-5}		
MP-SPC	2.9507×10^{-5}	1.5447×10^{-5}	1.6710×10^{-5}
MP-SPC + ACSC	2.7764×10^{-4}	1.2084×10^{-5}	9.9118×10^{-6}

The MP-SPC + ACSC configuration is tuned in two steps. In the first step, the MP-SPC is tuned. This tuning is discussed in Section 3.3 and the tuning parameters for the case studies being considered in this study are given in Table 3.3. The second step is to determine the appropriate K_c and T_R for the ACSC portion. An interesting feature about the tuning of MP-SPC + ACSC is that one has to strike a compromise between the possible reduction in variance and the possibility of increasing deviations from the mean. The deviation in the mean is despite the fact that both an on-line analyzer and integral action in the controller are available. The classical approach for tuning PI controllers relies on the fact that integral action minimizes the steady state offset for setpoint and load disturbances that are step changes. However, the disturbances encountered by MP-SPC and MP-SPC + ACSC are randomo fluctuations with a significant frequency content and so the speed of response becomes the most important issue rather than steady state offset. As a result, the classical rules for tuning PI controllers are not directly applicable. K_c and T_R values for the ACSC portion are determined using trial and error for the different controller configurations. Optimal values for the two case studies have been listed in Table 5.4. It should be noted that the K_c values corresponding to the case when both the manipulated variables are used for the Tennessee Eastman (TE) case is $[-100 \;\; 100]$. The difference in sign means that one of the score variables is postively correlated with

Table 5.4: MP-SPC + ACSC tuning parameters for the Tennessee Eastman (TE) process and the Azeotropic (AZ) tower cases.

Case		KCZ	TRZ
Tennessee Eastman (TE) process	D-feed SP as MV	-120	1020
	R-level SP as MV	-4.5	0.05
	Both	[-100 100]	[50 50]
Azeotropic (AZ) tower	FR-SP as MV	190	5
	T5-SP as MV	2	0.08
	Both	[950 950]	[37 37]

the G/H ratio while the other is negatively correlated. However, when the MP-SPC for this case was setup, both the score variables were chosen to be positively correlated with G/H ratio. This means that the cross-correlation between the second score variable and G/H ratio has changed in the presence of the disturbance. This change does not affect the controller performance drastically because the multivariable controller tuning parameters for MP-SPC, $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$, have been chosen in such a way that the channel for which the cross-correlation changes has been severely detuned.

In order to study the effect of the two tuning parameters (i.e. K_c and T_R) on controller performance, various controller configurations were simulated for the two case studies. In this parametric study, one of the parameters was kept constant and the other parameter was varied about it's optimal value. Figure 5.1 shows the effect of the controller gain (K_c) on the variance and the mean value of G/H ratio in the product stream for the Tennessee Eastman (TE) process when the D-feed SP is being used as the manipulated variable (MV). It can be seen that the most appropriate K_c value is -120 because although the mean value continues to approach 1.226 as K_c increases, the variance in G/H ratio begins to increase beyond $K_c = -120$. Hence, the optimal value of K_c is a compromise between decreasing the variance and increasing the deviations from the desired mean value. Figure 5.2 shows that integral time (T_R) has little effect on controller performance.

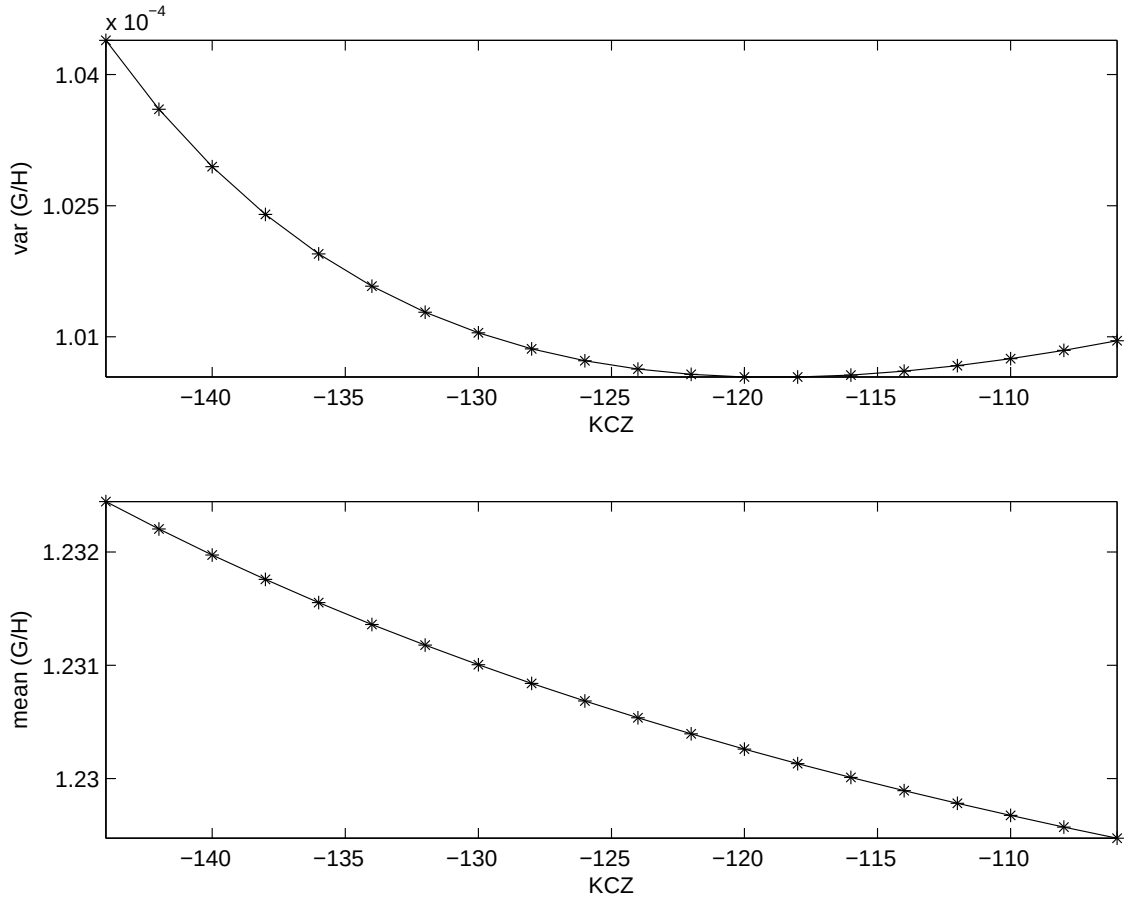

Figure 5.1: Tuning of controller for MP-SPC + ACSC formulation (effect of K_c).

Similar trends are observed for the effect of K_c and T_R on the mean and the variance of the quality variable for the case when R-level is used as the MV or for the multivariate case involving both being used as MVs. Moreover, the three configurations involving the Azeotropic (AZ) tower case study also have similar characteristics.

5.2 Influence of disturbance characteristics on controller performance

As mentioned before, it is a well known fact that controller performance is dependent on the type and direction of the inputs. Thus far, controller performances were evaluated for both stationary and non-stationary disturbances for the Tennessee Eastman (TE)

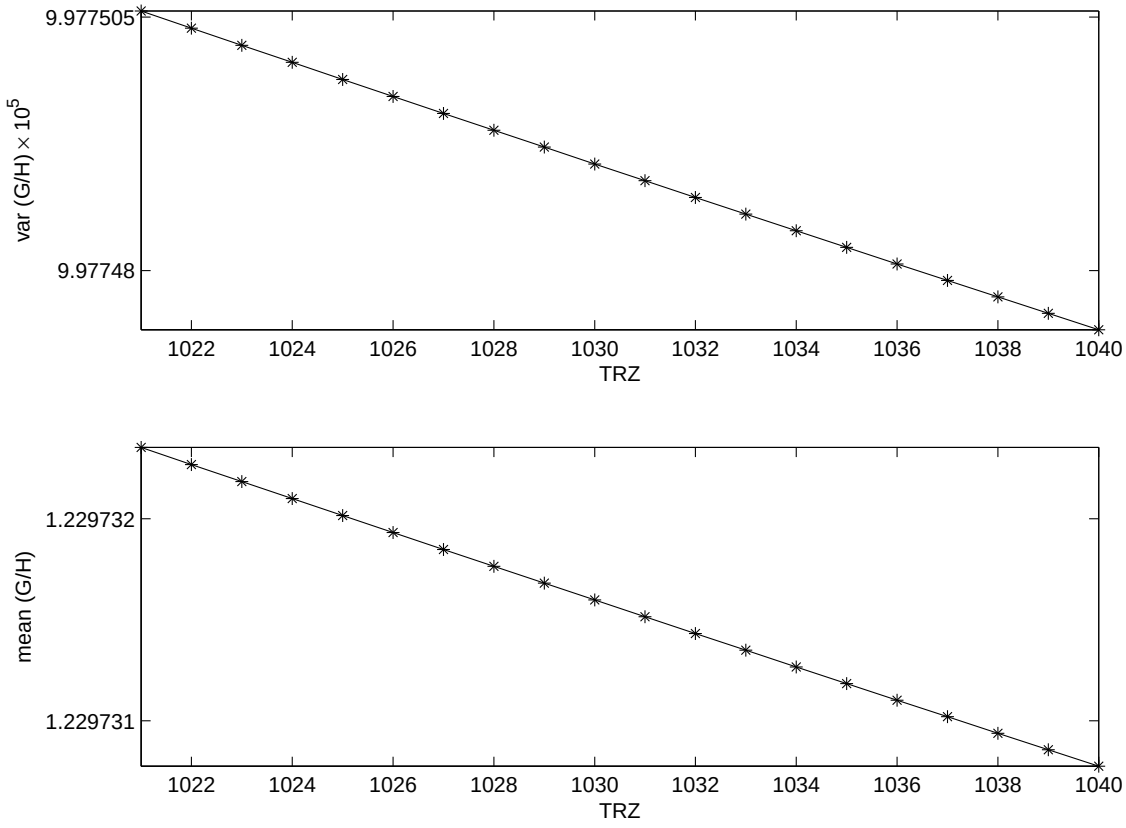

Figure 5.2: Tuning of controller for MP-SPC + ACSC formulation (effect of T_R).

process and the Azeotropic (AZ) tower. The non-stationary disturbances were constructed by overlaying a step disturbance on top of a stationary disturbance. IDV(8), a random fluctuation in the C-feed was the stationary disturbance for the Tennessee Eastman (TE) process while a random fluctuation in the feed to the tower was the stationary disturbance for the Azeotropic (AZ) tower case study. As discussed in Section 2.1 and shown in Figure 2.3, IDV(8) is composed of a wide range of frequencies and the nature of this disturbance (i.e. frequency content and magnitude) is also time varying. Similarly, Figure 2.6 shows that the disturbance in the feed to the Azeotropic (AZ) tower also encompasses a range of frequencies. Hence, it would be worthwhile to investigate controller performance at a fixed disturbance frequency for a range of frequencies.

In order to examine controller performance at a fixed frequency, the original FORTRAN code for the Tennessee Eastman (TE) process was modified such that the mole fraction of A in stream 4 is varied in a sinusoidal fashion. Controller performance, in terms of the variance in the G/H ratio, is observed when the process is upset with sinusoids of different frequencies. The range of frequencies covered spans the same range as that for IDV(8). The magnitude of the sinusoids is $\pm$ 2% of the steady state value. It should be noted that for sinusoids of $\pm$ 4%, the plant shuts down for low frequeny upsets. This is due to an imbalance in the stoichiometry of the components that cannot be compensated for by the control systems and a subsequent build up of a particular components causes the shutdown.

Figure 5.3 shows a comparison of the frequency response for four different controller formulations for the Tennessee Eastman (TE) process. It can be seen that MP-SPC performs better than all other formulations for a major portion of the frequency range. This portion of the frequency range, i.e. frequencies greater than 10^{-4} rad/s, are the frequen-

cies of interest. It can also be seen that at very low frequencies (i.e. those less than 10^{-4} rad/s) the base control system with the analyzer loop (G/H to D/E setpoint) activated performs the best (i.e. leads to the lowest variance in the product quality variable). This is because at lower frequencies, the validity of the relationship between the score variable and the quality variable becomes less accurate. The inacuracy of the model used in MP-SPC reduces the advantage provided by the lead time available to MP-SPC. As a result, the Base control + analyzer configuration performs slightly better than MP-SPC. It is also seen that the MP-SPC + ACSC formulation does worse than MP-SPC alone at very low and at very high frequencies. The tuning parameters for the PI controller (i.e. the ACSC portion) of the MP-SPC + ACSC formulation have been optimized for IDV(8) and the mean frequency content of IDV(8) is about 5×10^{-4} rad/s. For this frequecny, MP-SPC + ACSC performs the best. However, at frequencies farther from 5×10^{-4} rad/s, the K_c and T_R values are no longer optimal and the performance of MP-SPC + ACSC deteriorates as a result. Hence, the manner in which the ACSC portion has been tuned (i.e. the frequency for which the tuning parameters have been optimized) has a major impact on the controller's performance. Another feature of the frequency response plot for the Tennessee Eastman (TE) process is that the curves go through a maxima which is close to the ultimate frequency of the system (i.e. 10^{-3} rad/s). It should be noted that these plots are the responses of the non-linear systems and they are not the Bode plots of the linearized model. Similar trends are observed for the AZ tower. However, since the ultimate frequency for the Base control system and that for MP-SPC is significantly lower, the curves do not exhibit a peak. However, the curve for MP-SPC + ACSC system does exhibit a maximum because the PI controller for the ACSC portion has been aggressively tuned.

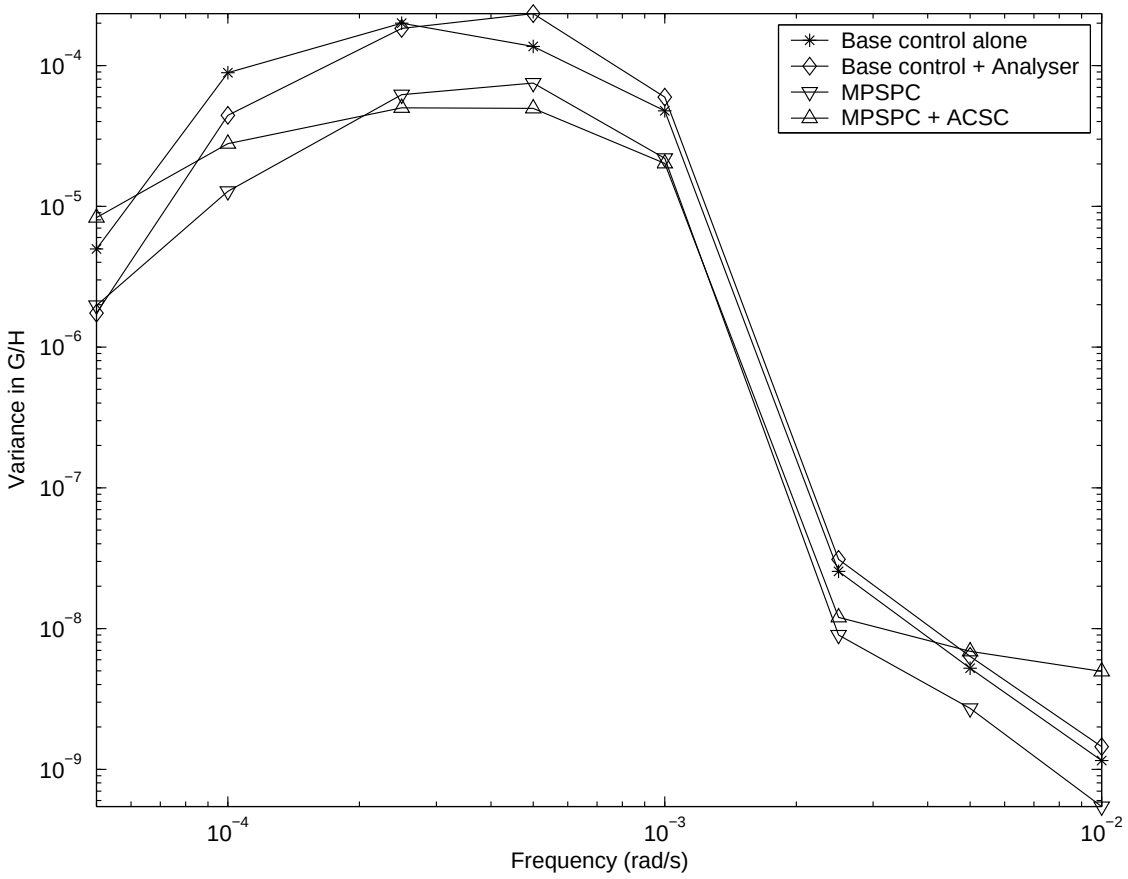

Figure 5.3: Frequency response of different controller formulations for the Tennessee Eastman (TE) process.

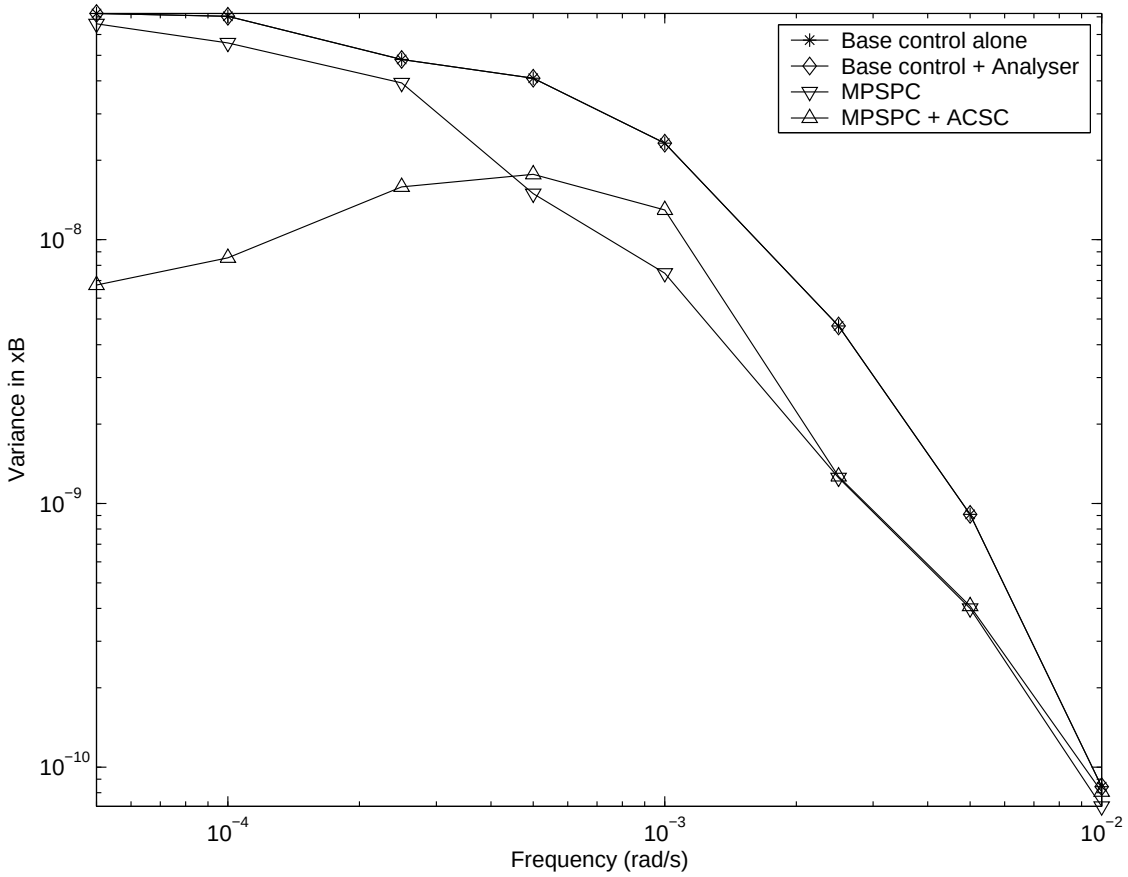

Figure 5.4: Frequency response of different controller formulations for AZ.

5.3 Alternative approaches for dealing with non-stationary disturbances

The ability of the basic formulation of MP-SPC in dealing with non-stationary distubances was discussed in Section (3.2.3). In the case of the Tennessee Eastman (TE) process there is an intial period immediately after the onset of the step disturbance that has large deviations from the target setpoint. After this initial period, the Base control system alone is capable of bringing the quality variable back to oscillating about it's original steady state. MP-SPC was able to provide an improved response, when compared to the Base control system alone, by not only reducing the variability of the product but also by eliminating the initial period of large deviations. In the case of the Azeotropic (AZ) tower, a non-stationary disturbance in the feed results in an offset in the composition of the bottoms product for both control systems, but MP-SPC continues to perform better than the Base control system (with or without the on-line analyzer).

In both of these cases, MP-SPC performed better than the Base control system. However, in some situations, it is possible that MP-SPC with a fixed setpoint for the score variable performs worse than the Base control system. For example, when McAvoy [31] applied MP-SPC to the Tennessee Eastman (TE) process, a non-stationary disturbance resulted in a deterioration in the performance of MP-SPC and an offset for the G/H ratio. In order to prevent the deterioration in the performance of MP-SPC during non-stationary disturbances it is necessary to adjust the setpoint for the score variables to compensate for this change. There are two ways in which the setpoints can be adjusted:

5.3.1 Using steady-state process models (MP-SPC + SS-MPC)

The approach adopted by McAvoy [31] to tackle such situations is as follows. A steady state model predictive controller (SS-MPC) is used on top of the dynamic score

controller. This SS-MPC changes the D/E ratio based on lab results analyzing the G/H ratio in the product. The control action is determined by solving:

$$\left.\begin{array}{r} \min \\ \left(\Delta\left(\frac{d}{e}\right)\right)_k \end{array} \left[\left(\frac{\hat{g}}{h}\right)^2 + w * \Delta\left(\frac{d}{e}\right)_k^2\right] \right\} \quad (5.2)$$
$$\text{subject to: } \left(\frac{\hat{g}}{h}\right) = m\Delta\left(\frac{d}{e}\right)_k + \Delta\left(\frac{g}{h}\right)_k$$

where,

$$\left.\begin{array}{l} \left(\frac{\hat{g}}{h}\right) = \left(\frac{G}{H}\right)^{\text{predicted}} - 1.226 \\ \Delta\left(\frac{g}{h}\right)_k = \left(\frac{G}{H}\right)_k^{\text{measured}} - 1.226 \\ \Delta\left(\frac{d}{e}\right)_k = \left(\frac{d}{e}\right)_k - \left(\frac{d}{e}\right)_{k-1} = \left(\frac{D}{E}\right)_k - \left(\frac{D}{E}\right)_{k-1} \end{array}\right\} \quad (5.3)$$

Also w is a scalar weight. As discussed earlier, based on the stoichiometry,

$$\frac{g}{h} \propto \frac{d}{e} = m * \frac{d}{e} = 1.5 * \frac{d}{e}$$

and so,

$$\Delta\left(\frac{d}{e}\right)_k = -\frac{\left[m * \Delta\left(\frac{g}{h}\right)_k\right]}{[m^2 + w]} \quad (5.4)$$

It is assumed that lab results for G/H are available every 8 hours for cumulative samples taken during this time period i.e. G/H values averaged over 8 hours are available. It should be noted that the Tennessee Eastman (TE) process has been extensively studied and a dynamic simulation for the process is available. Moreover, the process is simple enough to make engineering judgements for the design of the steady state MPC controller based solely on the stoichiometry. However, intermittent lab samples and linear models relating the manipulated variable to the quality variable are not always available and so alternative approaches need to be explored.

5.3.2 Using tests for signal stationarity (MP-SPC_W)

In this formulation, the Wilcoxon test is used to determine whether the score variable is at a steady state or not. During controller initialization, it is assumed that only stationary disturbances are present and so the setpoint for the score variable is zero. Upon encountering a step disturbance, the score variable starts shifting to a new steady state. The Wilcoxon test detects this departure and the MP-SPC is switched OFF and is kept OFF during the entire transient time period. Once a new steady state is reached, the controller is re-initiallized and brought back on-line. Re-initiallization involves determining the new set-point for the score variable. The mean value of the score variable at this new steady state is the new set-point for the score controller. A proper utilization of the Wilcoxon test is critical to the effectiveness of this approach and so this aspect is discussed next.

The Wilcoxon test, also referred to as the Mann-Whitney test, is a statistical test to determine whether two populations of numbers are identical (or whether their medians are equal). It is non-parametric, i.e. it makes no assumptions about the distribution of the data and so is more versatile than other tests. However it does assume that the data are distributed symmetrically around their median. In order to determine whether the score variable is at a steady state, an observation window of a suitable size is chosen. This window is divided into two equal halves and the two halves are compared using the Wilcoxon test. If the test determines that the process is not at a steady state, the controller is switched OFF (or kept OFF in case it already is OFF). If the test determines that the process is at a steady state, then the moving average of the score variable in the latter half is considered its new steady state value. This new steady state value is used as the new set-point for the score variable by the controller.

As mentioned in Section 1.3.1, process measurements can be categorized as matched samples or as independent samples. In the present implementation, since we are comparing the same attribute/measurement (i.e. the score variable) twice, the samples are matched and hence the `signrank` function available in the MATLAB Statistics Toolbox is used. Wilcoxon tests, like other statistical hypothesis tests, require the specification of the level of significance (α). From this study's perspective this parameter represents the compromise between missed correct alarms and false alarms. Too low a value of α would lead the controller to incorrectly think that the system is not at steady state when in reality it is. As a result it would switch ON and OFF too often. On the other hand, too high a value of α would lead the controller to incorrectly think that the system is at steady state when in reality it is not. Thereby, the controller would remain ON even during the unsteady state period and the performance will be degraded. In this study, three values of α are considered: $\alpha = 0.05$, 0.01 and 0.005.

The second handle/tuning parameter in this formulation is the size of the window of observation, henceforth referred to as Win_size. It should be noted that too small a value of Win_size would make the controller incorrectly think that the system is not at steady state when in fact it is. However, too large a value of Win_size would create a larger lag time before the controller would take action thereby making it inefficient.

In order to better understand the interplay of α and Win_size on controller behavior, a series of experiments has been conducted. For the Tennessee Eastman (TE) process, the behavior of MP-SPC_W is observed under IDV(8) alone. Four values of Win_size are considered: Win_size = 20, 30, 40 and 50 data points. With a sampling interval of 2 minutes, these window sizes reflect time spans of 40, 60, 80 and 100 minutes respectively. The variations in the score variable are stationary and so ideally the controller should

Table 5.5: Number of times controller switches ON/OFF for IDV(8) alone

Win_size (No. of data points)	Time interval (min.)	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$
20	1:1000	33	30	24
	1001:2000	25	25	25
	2001:3000	29	27	27
	3001:4000	23	24	24
30	1:1000	17	9	9
	1001:2000	19	17	17
	2001:3000	21	21	19
	3001:4000	22	20	20
40	1:1000	15	9	9
	1001:2000	15	13	13
	2001:3000	15	16	16
	3001:4000	17	17	17
50	1:1000	7	7	5
	1001:2000	13	13	12
	2001:3000	14	12	12
	3001:4000	11	10	11

remain ON all the time. Table 5.5 shows the number of times the controller switches ON and OFF for different values of α and Win_size. It can be seen that controller switching increases as the value of α increases and Win_size decreases. However we cannot afford to have too low a value of α or too high a value of Win_size.

5.4 Conclusions

This chapter discusses several extensions to the basic formulation for MP-SPC. First, the assumption regarding the unavailability of the on-line analyzer for the composition of the product stream is relaxed. The alternative score control formulation termed MP-SPC + ACSC, in which the on-line analyzer is utilized, provides an improved performance over the basic MP-SPC formulation. Tuning of MP-SPC + ACSC is also discussed and it is shown how the right choice of the tuning parameters K_c and T_R can enable striking a

balance between a reduction in the variability and achieving the right mean value of the product composition. An analysis of the frequency response of the different controller formulations shows how MP-SPC and MP-SPC + ACSC performed over a wide frequency range of interest and hence demonstrates the robustness of the methodology. Enhancements to the basic MP-SPC formulation that help deal with certain non-stationary disturbance scenarios where the performance of MP-SPC is worse than the Base control system were discussed. The steady state model predictive controller cascaded on top of MP-SPC (i.e. MP-SPC + SS-MPC) uses intermittent laboratory results coupled with a knowledge of the underlying stoichiometry to adjust the D/E ratio in the Tennessee Eastman (TE) process. The MP-SPC_W formulation uses the Wilcoxon test to ensure that the score controller is active only when the process measurements are fluctuating about a steady state.

Chapter 6

Conclusions

This PhD study involves inferential model predictive control of continuous processes using statistical tools. The basic idea is that difficult to measure quality variables are *inferred* from easier to measure secondary variables. These inferential measuremnts are used in a *model predictive control* formulation to reduce the variability in the product quality. *Statistical tools* are used to develop reduced order predictive models (PCA), to deal with non-stationary disturbances (Wilcoxon test) and to select the most appropriate secondary measurements (correlation analysis).

6.1 Summary of Results

In this study, a detailed analysis of the feedback based score control methodology, termed model predictive statistical process control (MP-SPC), is conducted. The most important feature of this approach is that quality control is achieved without using real-time quality variable or disturbance measurements. Controller performance results are based on industrially benchmarked, nonlinear simulations of the following two case studies: the Tennessee Eastman (TE) process and the Azeotropic (AZ) distillation tower of the Vinyl acetate process. The following is a summary of the main results:

1. A multivariable formulation (i.e. using multiple manipulated variables) for model predictive statistical process control (MP-SPC) is presented. Significant improvements in product quality control (as reflected in a reduced variance of the quality variable) is achieved.

2. A study of the controller's performance in dealing with disturbances of different frequencies, sizes and directions and it's ability to deal with non-stationary disturbances reveals the robustness of the approach.

3. Strategies for tuning MP-SPC are discussed. The penalty on manipulated variable moves is the most efffective tuning parameter.

4. A novel strategy for multivariate statistical analysis of databases constructed from historical operation data and plant test data is proposed.

5. The proposed measurement selection method considers only those measurements that are consistent in their cross-correlation with the quality variable and those that do not exhibit inverse response. A selection algorithm ranks measurements according to their cross-correlation with the quality variable and lead time. Higher ranked variables are chosen as long as they make sufficiently large contributions to the model.

6. The connectivity between collinear PC scores and PLS is illustrated using the simple configuration involving one manipulated variable (MV). For the single MV case, collinear scores obtained using orthogonal PCA calculations such that all the principal components are retained are equivalent to the scores for the Y-block elements from a PLS calculation considering the $\mathbf{u}$ vector as the X-block and the measurements $\mathbf{Y}$ as the Y-block.

7. Several approaches for identifying dynamic score models are proposed. One approach is to put greater emphasis on short term predictions while another is to utilize residual information to remove outliers or to incorporate the inverse of the residual as a sample weight.

6.2 Principal contributions

1. A model predictive score controller involving multiple manipulated variables. The code is in MATLAB script.

2. A measurement selection algorithm. The code is in MATLAB script.

3. Three approaches for model identification. The code is in MATLAB script.

4. A modified version of the Tennesse Eastman Challenge problem such that the user can externally input stream 4 composition disturbances and also extract out IDV(8). This is a modification of the original FORTRAN code and the MEX interface for MATLAB.

6.3 Recommendations for Future Work

The results obtained in this study are promising enough to motivate further study. Some suggestions for further investigations are:

1. Alternative formulations and extensions:

   (a) State-space formulations: The MP-SPC controller discussed in this study is based upon the DMC algorithm. The major limitation of this approach is it's limited ability to handle different noise and disturbance models. The implicit assumption there is that the disturbances are random steps (Wiener process) added at the process output and that the measurement noise is white. Lately several authors (for example Lee et al [25], Qin and Badgwell [41, 42]) have pointed out the advantages of using a state-space approach to model predictive control.

(b) Non-linear formulations: PCA is a linear method and most chemical processes, the kind of applications MP-SPC is well suited for, are nonlinear. Since most systems exhibit linear behavior when perturbed within a small range, linear PCA is adequate for most chemical process applications. Sometimes, when the non-linearities are strong, it would be a good idea to try non-linear PCA (NPCA). NPCA, proposed by Dong and Mc Avoy [11], is based on principal curves and neural networks. It is a more efficient way of capturing the nonlinear relationships among process variables than linear PCA. It should be noted that if NPCA is used to build score models, the controller formulations would no longer be linear and it would not be possible to obtain closed form solutions as before.

(c) On-line model updating: Chemical processes tend to drift due to phenomena such as changes in the feed or upstream processing, catalyst deactivation, fouling and aging of equipment (particularly heat exchangers), changes in analytical instrumentation, etc. This drift can change the relationship between variables and could cause the PCA/PLS models to become invalid. Persistent high values of Q-residuals, T^2 or a bias in the estimate is an indication that the model needs some kind of updating. PCA/PLS models can be updated on-line provided reference values are measured. Recursive techniques with or without forgetting, recalculating the model using the entire history or a moving window are some of the possible approaches that can be taken.

(d) Missing values: Process sensors are prone to failure. They are also quite often taken off-line for routine maintenance. If one of the secondary measurements used to build the soft-sensor becomes un-available, it maybe assumed that dur-

ing this time period the correlation structure of the measurements is preserved and the the PCA/PLS model used to infer the quality variable.

2. Tuning: It is obvious that not only is the tuning of the MP-SPC and similar controllers vital to their success but also that an improperly tuned controller, for example too much aggressiveness in the control action, could lead to instabilities. The tuning strategy adopted here was one of trial and error using simulations; a luxury not always available. What is required is a reliable strategy that generates very good initial estimates for the tuning parameters, using a closed form expression if possible, so that very little fine tuning is required during on-line implementation. One possible approach could be the one suggested by Sridhar and Cooper [49, 50].

3. Measurement selection: The secondary measurement selection procedure developed in this study is quite general and applicable to many different forms of PCA/PLS based inferential estimators. However, two questions remain unanswered: the determination of the number of measurements to use in the PCA model (m^*) and the number of candidate measurements to be iteratively evaluated according to their rank (m_c). When determining m^*, it is important to assess whether the benefit of a better fit is worth the additional complexity involved in using more measurements. In this study, m^* and m_c were arbitrarily chosen to be 5 and 10 respectively. A systematic way of determining these quantities needs to be developed. One approach could be the use of the Akaike's Information Theoretic Criterion (AID), Akaike's Final Prediction Error (FPE) or Rissanen's Minimum Description Length (MDL):

$$AIC = V\left[1 + 2\left(\frac{d}{N}\right)\right] \tag{6.1}$$

$$FPE = V\frac{N+d}{N-d} \tag{6.2}$$

$$MDL = V\left[1 + log(N)\left(\frac{d}{N}\right)\right] \tag{6.3}$$

here V is the quadratic loss function, d is the total number of parameters in the structure in question (in our case $d = m^*$) and N is the number of data points used for the estimation. The aim is to choose models with smaller values of AIC, FPE or MDL. For details regarding these approaches see Ljung [27].

4. PCA model development: A key question during PCA model development is: How many PCs to retain? Although this issue has been well studied for situations where PCA is applied on static data, systematic approaches to make this decision for dyanmic model development need to be developed.

5. Dynamic score model identification: Better identification techniques need to be investgated. One could also incorporate a disturbance or noise model and thereby make the predicions more accurate.

6. Additional case studies and pilot-plant evaluations: In order to firmly establish the effectiveness of the MP-SPC approach, it's applicability to several other case studies needs to be investigated. Besides the test case processes used in this study, a test case involving a Propane-propylene splitter has also been developed. With the computing resources currently available, the run times for this case study are too long (about four hours for one simulation experiment). This case study may be considered once faster computing power becomes available. Nonlinear dynamic simulations for the Recyle process proposed by Reyes-DeLeon and Luyben [45] and others listed in Chapter 2 are also available and so these would make good candidate case studies. The next step would be to evaluate the approach on a pilot plant before commercial scale deployment which is the ultimate goal.

Appendix A

Linear algebra

This appendix reviews some preliminary topics in linear algebra relevant to this study.

Two subspaces $\mathbf{V}$ and $\mathbf{W}$ are orthogonal if every vector $\mathbf{v}$ in $\mathbf{V}$ is orthogonal to every vector $\mathbf{w}$ in $\mathbf{W}$, i.e. $\mathbf{v}^T\mathbf{w} = 0$ for all $\mathbf{v}$ and $\mathbf{w}$. Now consider a matrix $\mathbf{A} \in \Re^{m\times n}$ and let $\mathbf{U}$ be the echelon matrix obtained by performing elimination on $\mathbf{A}$. Then the rank r of $\mathbf{A}$ is equal to the number of non-zero rows of $\mathbf{U}$. Similarly, the dimension of the row space of $\mathbf{A}$ (the space spanned by the rows of $\mathbf{A}$), denoted by $\mathcal{R}(\mathbf{A}^T)$, is equal to r. Moreover, since elementary row operations leave the row space unchanged, the rows of $\mathbf{U}$ form a basis for the row space of $\mathbf{A}$, i.e. they span the same space. The column space of $\mathbf{A}$ (the space spanned by the columns of $\mathbf{A}$) is also referred to as the range of $\mathbf{A}$. It is denoted as $\mathcal{R}(\mathbf{A})$ and also has dimension r. The nullspace of $\mathbf{A}$, denoted by $\mathcal{N}(\mathbf{A})$, is of dimension $n - r$. It is the space of $\Re^n$ not spanned by the rows of $\mathbf{A}$. Similarly, the nullspace of $\mathbf{A}^T$, denoted by $\mathcal{N}(\mathbf{A}^T)$, is also known as the left nullspace of $\mathbf{A}^T$. It has dimension $m - r$ and is the space of $\Re^m$ not spanned by the columns of $\mathbf{A}$.

The nullspace $\mathcal{N}(\mathbf{A})$ and the row space $\mathcal{R}(\mathbf{A}^T)$ are orthogonal subspaces of $\Re^n$. Similarly, the left nullspace $\mathcal{N}(\mathbf{A}^T)$ and the column space $\mathcal{R}(\mathbf{A})$ (i.e. the range of $\mathbf{A}$) are orthogonal subspaces of $\Re^m$.

The *Singular Value Decomposition* (SVD) of $\mathbf{A}$ involves factoring it as:

$$\mathbf{A} = \mu * \mathbf{\Sigma} * \mathbf{V}^T \tag{A.1}$$

where, μ is an orthogonal $m \times m$ matrix (i.e. $\mu\mu^T = \mathbf{I} \in \Re^{m\times m}$). It contains the eigen-

vectors of $\mathbf{A}\mathbf{A}^T$. $\mathbf{V}$ is an orthogonal $n \times n$ matrix (i.e. $\mathbf{V}\mathbf{V}^T = \mathbf{I} \in \Re^{n \times n}$). It contains the eigenvectors of $\mathbf{A}^T\mathbf{A}$. It forms an orthonormal basis for the range of $\mathbf{A}$. That is, the columns of $\mathbf{V}$ span the same space as the columns of $\mathbf{A}$, and the number of columns of $\mathbf{V}$ is the rank of $\mathbf{A}$. $\mathbf{\Sigma}$ is a diagonal $m \times n$ matrix. It's non-negative, diagonal entries are referred to as the singular values, σ_i. The singular values are the square roots of the eigenvalues of $\mathbf{A}^T\mathbf{A}$. They decrease monotonically from the upper left to the lower right of $\mathbf{\Sigma}$, i.e. $\sigma_1 \geq \sigma_2 \geq \sigma_3 \cdots \geq \sigma_n \geq 0$. It should be noted that although the decomposition is not unique, the singular values are.

One of the most useful applications of SVD in process control is described next. Let $\mathbf{A}$ be defined as $\mathbf{A} \equiv (\mathbf{X})^T\mathbf{S}$, where the matrix of measurements, $\mathbf{X}$, is of dimension $m \times n$ and the matrix of set-points $\mathbf{S}$ is of dimension $m \times s$ such that the rank of $\mathbf{S}$ is s and $n > s$. Then the SVD of $\mathbf{A}$ yields μ which is used to transform $\mathbf{X}$ into two sets. This is depicted in Figure (A.1) and can be represented as:

$$\mathbf{X} * \mu = \mathbf{X} * [\text{Orth } \mathbf{A}, \text{Null } \mathbf{A}] = [\mathbf{C}, \mathbf{U}]$$

where the subspace $\mathbf{C}$ denotes the range of $\mathbf{A}$, i.e. the matrix of transformed measurements that are correlated with $\mathbf{S}$ (i.e. $\mathbf{C}^T\mathbf{S} \neq 0$). It results from the first s columns of μ. $\mathbf{U}$ defines the null space of $\mathbf{A}$, i.e. the matrix of transformed measurements that are not correlated with $\mathbf{S}$ (i.e. $\mathbf{U}^T\mathbf{S} = 0$). It results from columns $s + 1$ to n of μ.

Appendix B

The Propane-Propylene column (PP)

An additional test-bed problem, the third, specifically developed for this study involves a Propane-Propylene (PP) separation column (C_3-Splitter).

Superfractionators are industrial distillation columns used to separate very close

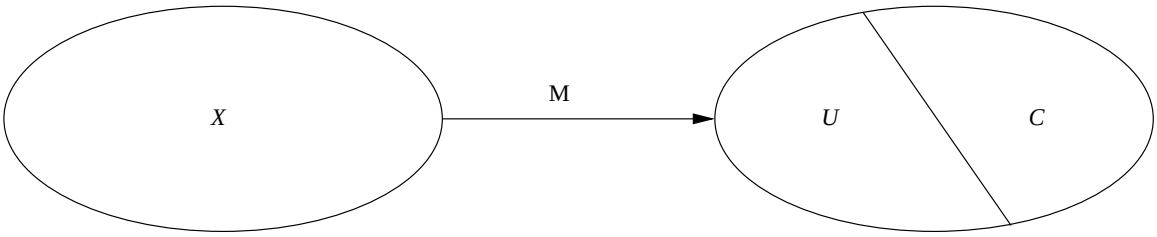

Figure A.1: Orthogonal transformation using SVD.

boiling mixtures such as alcohol isomers, C_5 and C_6 isomers, ethyl benzene/styrene and propane/propylene. Owing to the low relative volatility of these mixtures, a comparatively large number of stages is required to separate them. The reflux ratio and energy requirements are also very large. Consequently, such columns are sluggish with open loop response times of 5 to 30 hours. Control of this class of columns is important because improved control could lead to substantial savings in energy costs and/or reduction in the variability of the final products.

With an aim to use industrially relevant models as case studies, a dynamic model for the Propane-Propylene column ($C3$-Splitter) has been developed by comparing it with a dynamic model described by Gokhale et al [17, 18] and Hurowitz [20]. Their model (henceforth referred to as the benchmarked model for brevity) was successfully benchmarked against dynamic step data from an industrial C_3-Splitter. The benchmarked model had several simplifying assumptions and also did not consider off-key components. The model developed and used in this study not only exhibits similar dynamics but the thermodynamic calculations used are more rigorous. The added detail is obtained by incorporating parameters from a case-study provided with the ChemCAD process simulation software package [5]. Hence, it may be claimed that the model developed and used in this study provides a better match to industrially observed behavior. It has the following features:

1. Our model considers a feed consisting of four components: Ethane, Propylene, Propane and n-Butane. The benchmarked model did not consider the off-key components Ethane and n-Butane and so effectively treated this as a binary system. It is worth noting that when studying inferential control applications in distillation columns, it is essential that one considers situations where a direct one-to-one relationship betwen tray temperature and tray compositions is absent. Hence, binary systems do not provide a true test for inferential control strategies.

2. In the benchmarked model, the vapor liquid equilibria (VLE) was calculated using a relative volatility that was a quadratic function of liquid phase propylene mole fraction. The quadratic coefficients were linear functions of pressure. Hence the tray temperatures aren't functions of the tray compositions. In other words, the benchmarked model is inadequate for inferential control studies. In our model, the Peng Robinson Equation Of State (PR-EOS) is used to determine the bubble point temperature and vapor compositions.

3. Our column model has real 148 trays. The total condenser and partial reboiler act as stages too and so the column has 150 stages in all. The benchmarked model had 232 real trays and hence 234 stages.

4. Our model considers a Murphre tray efficiency of 90% while the benchmarked model's trays are 85% efficient. This difference in tray efficiences and thermodynamic calculations are responsible for the large differences in the number of trays required for similar degress of separation.

5. Our model uses a linear wier equation to describe the liquid dynamics with a hydraulic time constant of 3 seconds which is the same as that used in the benchmarked

Table B.1: Physical and thermodynamic property data for C3-Splitter.

Component index	Ethane 1	Propylene 2	Propane 3	n-Butane 4
Molecular weight (MW)	30.07	42.081	44.097	58.124
Critical Temperature (T_c)	305.42	364.76	369.82	425.18
Critical Pressure (P_c)	4880.1	4612.6	4249.2	3796.9
Accentric factor (ω)	0.099	0.1424	0.1516	0.1931
Specific gravity (Sp. Gr.)	0.548	0.612	0.582	0.579

model.

The PR-EOS is given by:

$$P = \frac{RT}{V - b_{mix}} - \frac{a_{mix}(T)}{V(V + b_{mix}) - b_{mix}(V - b_{mix})}$$

Bubble point temperature calculation using the PR-EOS are performed as follows. First, the following pure component quantities for each component q are evaluated:

$$\left.\begin{aligned} \kappa_q &= 0.37464 + 1.54226\omega_q - 0.26992\omega_q^2 \\ b_q &= 0.07780R\frac{T_{c,q}}{P_{c,q}} \\ ac_q &= 0.45724R^2\frac{T_{c,q}^2}{P_{c,q}} \\ a_q &= a_{c,q}\left(1 + \kappa_q\left(1 - \sqrt{\frac{T}{T_{c,q}}}\right)\right)^2 \end{aligned}\right\} \tag{B.1}$$

The Van der Waal one-fluid mixing rules are used to calculate the parameters for the mixture and the EOS of the mixture in terms of the compressibility.

$$a_{mix} = \sum_{i=1}^{4}\left(\sum_{j=1}^{4}(x_i x_j\sqrt{a_{i,i}a_{j,j}(1 - k_{i,j})}\right) \tag{B.2}$$

$$b_{mix} = \sum_{i=1}^{4} x_i b_i \tag{B.3}$$

$$A = \frac{a_{mix}P}{(RT)^2} \tag{B.4}$$

$$B = \frac{Pb_{mix}}{RT} \tag{B.5}$$

The following cubic equation is solved for the compressibility roots using Cardano's method. All imaginary roots are set to zero. The largest root is the vapor compressibility ($Z_v = \mathcal{Z}_{max}$) while the smallest one is the liquid compressibility ($Z_l = \mathcal{Z}_{min}$).

$$\mathcal{Z}^3 + (B-1)\mathcal{Z}^2 + (A - 3B^2 - 2B)\mathcal{Z} + (-AB + B^2 + B^3) = 0$$

$$ln\left(\frac{f_i}{x_i P}\right) = \frac{(\mathcal{Z}-1)b_i}{b} \quad -ln(\mathcal{Z}-B) \quad - \frac{a\left(\frac{2\sum_{j=1}^{4}(z_j\sqrt{a_i a_j}(1-k_{i,j})}{a} - \frac{b_i}{b}\right) ln\left[\frac{(\mathcal{Z}+(1+\sqrt{2}B)}{(\mathcal{Z}+(1-\sqrt{2}B)}\right]}{2\sqrt{2}bRT} \tag{B.6}$$

The procedure consits of two nested iteration loops. The inner loop iterates on the vapor phase composition until the vapor phase species fugacities (f_vs) are equal to the liquid phase species fugacities (f_ls) for every species within a prespecified tolerance. Once the inner loop converges, the outer loop iterates on the temperature until the vapor phase mole fractions sum to 1 within a prespecified tolerance. When the outer loop iteration is done, the final temperature and vapor composition are the Bubble point temperature and vapor composition.

Figure B.1 shows the process flowsheet along with the controller pairings for the base control system. The Base control system consists of the Reflux-drum level being controlled by the distillate flow rate while the Reboiler-level is being controlled by the heat input to the Reboiler. If analyzers for the compostion of the top and bottoms streams are available, then the distillate composition can be controlled using the Reflux-flow rate while the bottoms compositon can be controlled using the bottoms flow rate. This arrangement is typical of situations involving a high boilup ratio (i.e. high V/B) and is known as the

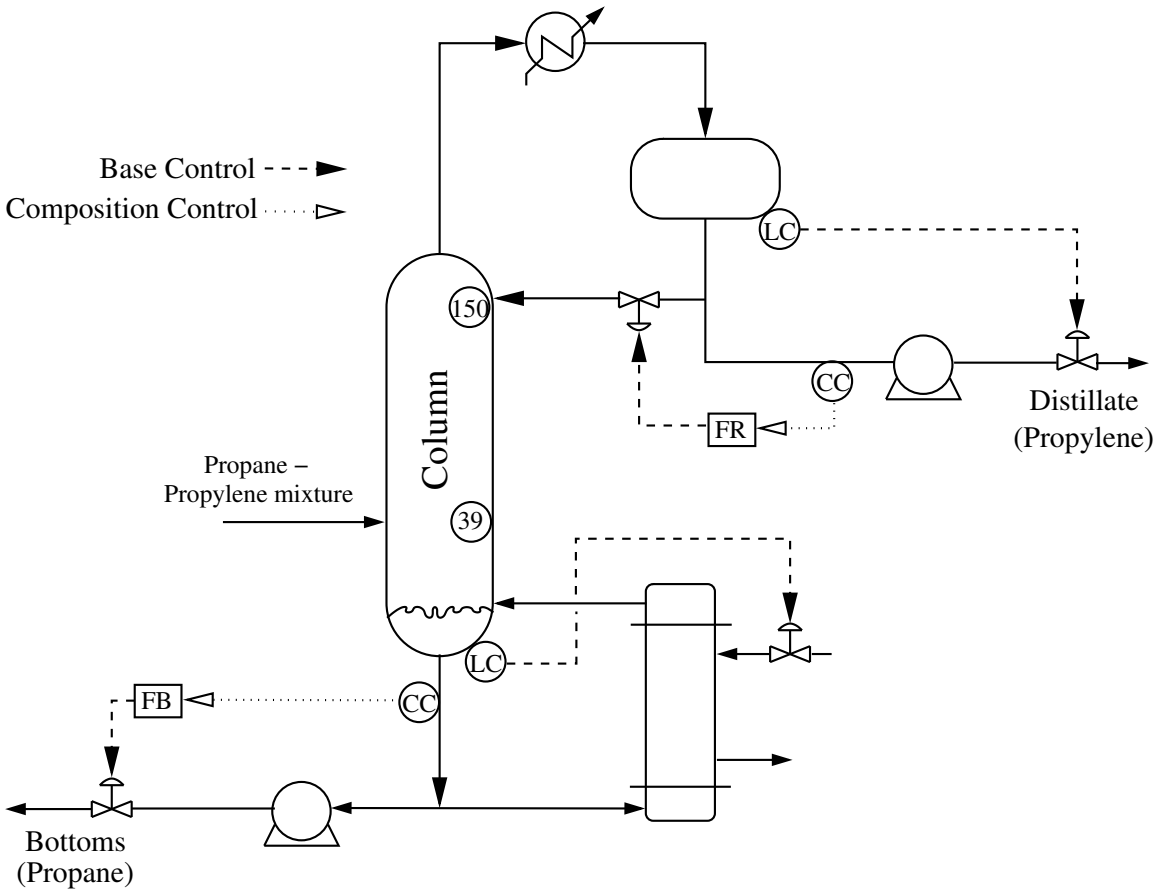

Figure B.1: The Superfractionator (C3_Splitter) with it's base control system.

"L-B" configuration. It should be noted that such an arrangement has the potential for inverse response. Table B.2 provides the controller tuning parameters used.

Table B.2: Control structure, controller parameters, setpoints and steady state values for Propane-Propylene splitter.

Loop No.	Controlled variable		Manipulated variable		Type	K_c	T_R (min)
		setpoint		S.S. value			
1	% C_3H_6 in distillate ($xT(2)$)	0.954	Reflux flow rate (FR)	74546	PI	0.5	60
2	% C_3H_6 in bottoms	0.046	Bottoms flow rate (FB)	1602	PI	20	30
3	Condenser drum level (LT)	50%	distillate flow rate (FD)	4108	P	1	
4	Column bottom level (LB)	50%	Vapor flow rate (VB)	78654	P	1	

B.1 Computer code

This is the computer code for the Propane-propylene splitter. First set up the C-Mex compiler in MATLAB and then use `compile_dll.m` to compile all the files into a single DLL file that can be called from within the MATLAB environment using:

```
     dstatedt_Col, T_Column, TB, TT, y, yB, yT] = C3Splitter(states,
P_Column, FL, xF, TFL, PF, FR, VB, FD, FB, T_guess, TB_guess, TT_guess,
yy_guess, yB_guess, yT_guess, Eff, NT_column, NF_column, hydtau_column,
M0_column, L0_column, Reboiler_drum_Working_Level_Volume,
Reflux_drum_Working_Level_Volume);
```

B.1.1 compile_dll.m

```
     mex C3Splitter.c BubblePR_T_y.c PRparamsPure.c PRmixParams.c
PRsolvEOS.c PRfug.c InnerIterate.c cubic_solve.c
```

B.1.2 C3Splitter.c

```
     #include <stdio.h>
     #include <math.h>
     #include "mex.h"
     void C3Splitter(double *dstatedt, double *T, double *TB, double *TT,
double *yy, double *yB, double *yT, double states[], double P[],
double FL, double xF[], double TFL, double PF, double FR, double VB,
double FD, double FB, double T_guess[], double TB_guess, double TT_guess,
double yy_guess[], double yB_guess[], double yT_guess[], double Eff, int NT,
int NF, double hydtau, double M0[], double LZERO[],
double bottom_working_volume, double top_working_volume)
     {
     double PropMatrix[3][4]=
     {
     {305.42, 364.76, 369.82, 425.18},
     {4880.1, 4612.6, 4249.2, 3796.9},
     {0.099, 0.1424, 0.1516, 0.1931},
     };
     double kij[4][4]=
     {
     {0., 0.0089, 0.0011, 0.0096},
     {0.0089, 0., 0.0074, 0.},
     {0.0011, 0.0074, 0., 0.0033},
     {0.0096, 0., 0.0033, 0.},
     };
     double MW[4] = {30.07, 42.081, 44.097, 58.124};
     double SpG[4] = {0.548, 0.612, 0.582, 0.579};
     double R = 8.314; /*Gas constant, l.kPa/(gmol.K) or J/(gmol.K)*/
```

```
     double Dw = 999.; /*density of water at 0 degC, kg/m3 or g/l*/
     double T_ref = 273.15; /*Reference Temperature, K*/
     double x[148][4], y[148][4], y_guess[148][4], M[148], L[148];
     double dMdt[148], dMxdt[148][3], dxdt[148][3],dxBdt[3],dxTdt[3];
     double xB[4], xT[4], xx[4], yyy_guess[4];
     double temp_T[1]=0.;
     double temp_P;
     double bottom_conc, top_conc, bottom_density, top_density,
Volume_bottom, Volume_top;
     double sum1 = 0.;
     double sum2 = 0.;
     double dMdt_top = 0.;
     double dMdt_bottom = 0.;
     double L_dot_top = 0.;
     double L_dot_bottom = 0.;
     int NC = 4;
     int i = 0;
     int j = 0;
     int n = 0;
     for (i=0;i<NT; i++)
     y_guess[i][0]=yy_guess[i];
     if (y_guess[i][0]<0.)
     y_guess[i][0]=0.;
     y_guess[i][1]=yy_guess[i+NT];
     if (y_guess[i][1]<0.)
     y_guess[i][1]=0.;
     y_guess[i][2]=yy_guess[i+2*NT];
     if (y_guess[i][2]<0.)
     y_guess[i][2]=0.;
     y_guess[i][3]=1-y_guess[i][0]-y_guess[i][1]-y_guess[i][2];
     if (y_guess[i][3]<0.)
     y_guess[i][3]=0.;
     }
     //mexPrintf("%12.20f\n", P[120]);
     //mexPrintf("%12.20f\n", P[121]);
     for (i=0;i<NT; i++)
     {
     x[i][0]=states[i];
     if (x[i][0]<0.)
     x[i][0]=0.;
     x[i][1]=states[i+NT];
     if (x[i][1]<0.)
     x[i][1]=0.;
     x[i][2]=states[i+2*NT];
     if (x[i][2]<0.)
     x[i][2]=0.;
     x[i][3]=1-x[i][0]-x[i][1]-x[i][2];
     if (x[i][3]<0.)
```

```
x[i][3]=0.;
M[i]=states[i+3*NT+6];
if (M[i]<0.)
{
M[i]=0.;
mexPrintf("Tray %d hold-up negative!\n", i)
}
}
xB[0] = states[3*NT];
xB[1] = states[3*NT+1];
xB[2] = states[3*NT+2];
xB[3] = 1 - xB[0] - xB[1] - xB[2];
xT[0] = states[3*NT+3];
xT[1] = states[3*NT+4];
xT[2] = states[3*NT+5];
xT[3] = 1 - xT[0] - xT[1] - xT[2];
for (i=0;i<NC; i++)
{
if (xB[i]<0.)
xB[i]=0.;
if (xT[i]<0.)
xT[i]=0.;
}
sum1=0.;
sum2=0.;
for (i=0;i<NC; i++)
{
sum1+=xB[i]*MW[i]/SpG[i];
sum2+=xB[i]*MW[i];
};
bottom_conc=Dw/sum1;
bottom_density=sum2*bottom_conc;
Volume_bottom=states[4*NT+6]*bottom_working_volume;
sum1=0.;
sum2=0.;
for (i=0;i<NC; i++)
{
sum1+=xT[i]*MW[i]/SpG[i];
sum2+=xT[i]*MW[i];
};
top_conc=Dw/sum1;
top_density=sum2*top_conc;
Volume_top=states[4*NT+7]*top_working_volume;
temp_P=P[0];
BubblePR_T_y(&TB_guess, temp_P, xB, yB_guess, R, PropMatrix, kij, NC);
TB[0]=TB_guess;
for (i=0;i<NC; i++)
yB[i]=yB_guess[i];
```

```
temp_P=P[NT+1];
BubblePR_T_y(&TT_guess, temp_P, xT, yT_guess, R, PropMatrix, kij, NC);
TT[0]=TT_guess;
for (i=0;i<NC; i++)
yT[i]=yT_guess[i];
for (n=0;n<NT; n++)
{
temp_T[0]=T_guess[n];
temp_P=P[n+1];
for (i=0;i<NC; i++)
{
xx[i]=x[n][i];
yyy_guess[i]=y_guess[n][i];
};
BubblePR_T_y(temp_T, temp_P, xx, yyy_guess, R, PropMatrix, kij, NC);
T[n]=temp_T[0];
for (i=0;i<NC; i++)
y[n][i]=yyy_guess[i];
L[n] = LZERO[n] + (M[n]-M0[n])/hydtau;
//mexPrintf("%12.20f\n", M[n]);
if (L[n]<0)
mexPrintf("Tray %d liquid flow negative!\n", n);
};
for (i=0;i<NC; i++)
y[0][i] = Eff*(y[0][i]-yB[i]) + yB[i];
for (n=1;n<NT; n++)
{
for (i=0;i<NC; i++)
y[n][i] = Eff*(y[n][i]-y[n-1][i]) + y[n-1][i];
;
if (FB<0.)
mexPrintf("Bottom flow negative!\n");
dMdt_bottom= L[0] - FB - VB;
L_dot_bottom=dMdt_bottom/bottom_conc/bottom_working_volume;
if (FD<0.)
mexPrintf("Top flow negative!\n");
dMdt_top= VB - FR - FD;
L_dot_top=dMdt_top/top_conc/top_working_volume;
dMdt[0] = L[1]-L[0];
for (i=0;i<NC-1; i++)
dMxdt[0][i]= L[1]*x[1][i]+VB*(yB[i]-y[0][i])-L[0]*x[0][i];
for (i=1;i<NT-1; i++)
{
dMdt[i] = L[i+1]-L[i];
for (j=0;j<NC-1; j++)
dMxdt[i][j]=L[i+1]*x[i+1][j]-L[i]*x[i][j]+VB*(y[i-1][j]-y[i][j]);
};
dMdt[NF-1] = dMdt[NF-1]+FL;
```

```
     //printf("%12.20f\n", dMdt[1]);
     for (i=0;i<NC-1; i++)
     dMxdt[NF-1][i]= dMxdt[NF-1][i]+ FL*xF[i];
     dMdt[NT-1] = FR-L[NT-1];
     for (i=0;i<NC-1; i++)
     dMxdt[NT-1][i]=
     VB*(y[NT-2][i]-y[NT-1][i])+FR*xT[i]-L[NT-1]*x[NT-1][i];
     for (i=0;i<NT; i++)
     {
     for (j=0;j<NC-1; j++)
     dxdt[i][j]=(dMxdt[i][j]-x[i][j]*dMdt[i])/M[i];
     };
     for (i=0;i<NC-1; i++)
     {
     dxBdt[i] = (x[0][i]*L[0]-yB[i]*VB-xB[i]*FB-xB[i]*dMdt_bottom)/
bottom_conc/Volume_bottom;
     dxTdt[i] = (VB*(y[NT-1][i]-xT[i])-xT[i]*dMdt_top)/top_conc/Volume_top;
     };
     for (i=0;i<NT; i++)
     {
     yy[i]=y[i][0];
     yy[i+NT]=y[i][1];
     yy[i+2*NT]=y[i][2];
     };
     for (i=0;i<NT; i++)
     {
     dstatedt[i]=dxdt[i][0];
     dstatedt[i+NT]=dxdt[i][1];
     dstatedt[i+2*NT]=dxdt[i][2];
     dstatedt[i+3*NT+6]=dMdt[i];
     };
     dstatedt[3*NT]=dxBdt[0];
     dstatedt[3*NT+1]=dxBdt[1];
     dstatedt[3*NT+2]=dxBdt[2];
     dstatedt[3*NT+3]=dxTdt[0];
     dstatedt[3*NT+4]=dxTdt[1];
     dstatedt[3*NT+5]=dxTdt[2];
     dstatedt[4*NT+6]=L_dot_top;
     dstatedt[4*NT+7]=L_dot_bottom;
     }
     void mexFunction(int nlhs, mxArray *plhs[], int nrhs,
const mxArray *prhs[])
     {
     double *states;
     double *P;
     double FL;
     double *xF;
     double TFL;
```

```
double PF;
double FR;
double VB;
double FD;
double FB;
double *T_guess;
double TB_guess;
double TT_guess;
double *yy_guess;
double *yB_guess;
double *yT_guess;
double Eff;
int NT;
int NF;
double hydtau;
double *M0;
double *LZERO;
double bottom_working_volume;
double top_working_volume;
int mrows;
int ncols;
double *dstatedt;
double *T;
double *TB;
double *TT;
double *yy;
double *yB;
double *yT;
if(nrhs!=24)
mexErrMsgTxt("inputs not correct");
if(nlhs!=7)
mexErrMsgTxt("outputs not correct");
states = mxGetPr(prhs[0]);
P = mxGetPr(prhs[1]);
FL = mxGetScalar(prhs[2]);
xF = mxGetPr(prhs[3]);
TFL = mxGetScalar(prhs[4]);
PF = mxGetScalar(prhs[5]);
FR = mxGetScalar(prhs[6]);
VB = mxGetScalar(prhs[7]);
FD = mxGetScalar(prhs[8]);
FB = mxGetScalar(prhs[9]);
T_guess = mxGetPr(prhs[10]);
TB_guess = mxGetScalar(prhs[11]);
TT_guess = mxGetScalar(prhs[12]);
yy_guess = mxGetPr(prhs[13]);
yB_guess = mxGetPr(prhs[14]);
yT_guess = mxGetPr(prhs[15]);
```

```
     Eff = mxGetScalar(prhs[16]);
     NT = mxGetScalar(prhs[17]);
     NF = mxGetScalar(prhs[18]);
     hydtau = mxGetScalar(prhs[19]);
     M0 = mxGetPr(prhs[20]);
     LZERO = mxGetPr(prhs[21]);
     bottom_working_volume = mxGetScalar(prhs[22]);
     top_working_volume = mxGetScalar(prhs[23]);
     mrows = mxGetM(prhs[0]);
     ncols = mxGetN(prhs[0]);
     plhs[0] = mxCreateDoubleMatrix(mrows,ncols, mxREAL);
     dstatedt = mxGetPr(plhs[0]);
     mrows = mxGetM(prhs[10]);
     ncols = mxGetN(prhs[10]);
     plhs[1] = mxCreateDoubleMatrix(mrows, ncols, mxREAL);
     T = mxGetPr(plhs[1]);
     plhs[2] = mxCreateScalarDouble(0);
     TB = mxGetPr(plhs[2]);
     plhs[3] = mxCreateScalarDouble(0);
     TT = mxGetPr(plhs[3]);
     mrows = mxGetM(prhs[13]);
     ncols = mxGetN(prhs[13]);
     plhs[4] = mxCreateDoubleMatrix(mrows,ncols, mxREAL);
     yy = mxGetPr(plhs[4]);
     mrows = mxGetM(prhs[14]);
     ncols = mxGetN(prhs[14]);
     plhs[5] = mxCreateDoubleMatrix(mrows,ncols, mxREAL);
     yB = mxGetPr(plhs[5]);
     mrows = mxGetM(prhs[15]);
     ncols = mxGetN(prhs[15]);
     plhs[6] = mxCreateDoubleMatrix(mrows,ncols, mxREAL);
     yT = mxGetPr(plhs[6]);
     C3Splitter(dstatedt, T, TB, TT, yy, yB, yT, states, P, FL, xF, TFL, PF,
FR, VB, FD, FB, T_guess, TB_guess, TT_guess, yy_guess, yB_guess, yT_guess, Eff,
NT, NF, hydtau, M0, LZERO, bottom_working_volume, top_working_volume);
     }
```

B.1.3 BubblePR_T_y.c

```
     #include <stdio.h>
     #include <math.h>
     void BubblePR_T_y(double TT[], double P, double x[], double y[], double
R, double PropMatrix[][4], double k[][4], int NC)
     {
     double sumy=0.;
     double term=0.;
     double Tr[4]={0., 0., 0., 0.};
```

```
     double T=0.;
     double a[4]={0., 0., 0., 0.};
     double b[4]={0., 0., 0., 0.};
     double A[1]={0.};
     double B[1]={0.};
     double am[1]={0.};
     double bm[1]={0.};
     double ab[4]={0., 0., 0., 0.};
     double kap[4]={0., 0., 0., 0.};
     double ZL[1]={0.};
     double ZV[1]={0.};
     double fugl[4]={0., 0., 0., 0.};
     double fugv[4]={0., 0., 0., 0.};
     double lnzl[1]={0.};
     double lnzv[1]={0.};
     int i=0;
     int j=0;
     int test1=0;
     T=TT[0];
     while (test1!=1)
     {
     PRmixParams(Tr, a, b, A, B, am, bm, ab, kap, x, R, T, P, PropMatrix, k,
NC);
     PRsolvEOS(ZL, ZV, A[0], B[0]); PRfug(fugl, lnzl, x, a, b, B[0], k, am[0],
bm[0], ZL[0], R, T, P, NC);
     InnerIterate(Tr, x, y, kap, a, am[0], ab, b, bm[0], ZL, ZV, lnzv, k, R,
T, P, PropMatrix, fugl, NC);
     sumy=0.;
     for (i=0;i<NC; i++)
     sumy=sumy+y[i];
     if (fabs(sumy-1)>5e-6)
     T=T/(0.25*(sumy-1)+1);
     else
     test1=1;
     //mexPrintf("%12.20f\n", sumy);
     };
     TT[0]=T;
     }
```

B.1.4 PRparamsPure.c

```
     #include <stdio.h>
     #include <math.h>
     void PRparamsPure(double Tr[], double Pc[], double omega[], double a[],
double ab[], double b[], double ac[], double kap[], double T, double R, double
PropMatrix[][4], int NC)
     {
```

```
double Tc[4]={0., 0., 0., 0.};
int q=0;
for (q=0;q<NC; q++)
{
Tc[q]=PropMatrix[0][q];
Tr[q]=T/Tc[q];
Pc[q]=PropMatrix[1][q];
omega[q]=PropMatrix[2][q];
kap[q]=0.37464+1.54226*omega[q]-0.26992*pow(omega[q],2);
b[q]=0.07780*R*Tc[q]/Pc[q];
ac[q]=0.45724*pow(R*Tc[q],2)/Pc[q];
ab[q]=pow(1+kap[q]*(1-sqrt(Tr[q])),2);
a[q]= ac[q]*ab[q];
};
}
```

B.1.5 PRmixParams.c

```
#include <stdio.h>
#include <math.h>
void PRmixParams(double Tr[], double a[], double b[], double A[], double
B[], double am[], double bm[], double ab[], double kap[], double zz[], double
R, double T, double P, double PropMatrix[][4], double k[][4], int NC)
{
double Pc[4]={0., 0., 0., 0.};
double omega[4]={0., 0., 0., 0.};
double ac[4]={0., 0., 0., 0.};
double sum_n=0.;
int i, n;
PRparamsPure(Tr,Pc,omega,a,ab,b,ac,kap,T,R,PropMatrix,NC);
am[0]=0.;
for (i=0; i<NC; i++)
{
sum_n=0.;
for (n=0; n<NC; n++)
sum_n=sum_n+zz[n]*zz[i]*sqrt(a[i]*a[n])*(1-k[i][n]);
am[0] = am[0] + sum_n;
}
bm[0]=0.;
for (i=0; i<NC; i++)
bm[0]=bm[0]+zz[i]*b[i];
A[0] = am[0]*P/pow(R*T,2);
B[0] = P*bm[0]/R/T;
}
```

B.1.6 PRsolvEOS.c

```
#include <stdio.h>
#include <math.h>
void PRsolvEOS(double ZL[], double ZV[], double A, double B)
{
double C[4] = {0.,0.,0.,0.};
double ZZ[3] ={0.,0.,0.};
double ZZMin=0., ZZMax=0.;
C[2] = B-1;
C[1] = A-3*pow(B,2)-2*B;
C[0] = -A*B+pow(B,2)+pow(B,3);
cubic_solve(ZZ,C);
if (ZZ[0]>ZZ[1])
{
if (ZZ[0]>ZZ[2])
{
ZZMax = ZZ[0];
if (ZZ[1]>ZZ[2])
ZZMin=ZZ[2];
else
ZZMin=ZZ[1];
}
else
{
ZZMax = ZZ[2];
ZZMin = ZZ[1];
}
}
else
{
if (ZZ[1]>ZZ[2])
{
ZZMax = ZZ[1];
if (ZZ[0]>ZZ[2])
ZZMin = ZZ[2];
else
ZZMin = ZZ[0];
}
else
{
ZZMax = ZZ[2];
ZZMin = ZZ[0];
}
}
if (ZZMin==0)
ZZMin=ZZMax;
else
```

```
{
if (ZZMax==0)
ZZMax=ZZMin;
}
ZL[0] = ZZMin;
ZV[0] = ZZMax;
}
```

B.1.7 PRfug.c

```
#include <stdio.h>
#include <math.h>
void PRfug(double fug[], double lnz[], double zz[], double a[], double
b[], double B, double k[][4], double am, double bm, double ZZZ, double R, double
T, double P, int NC)
{
double f[4], phil[4];
double sum1=0.;
double sqrt2=0.;
int i,j;
sqrt2=pow(2,0.5);
lnz[0] = 1/(2*sqrt2*bm*R*T)*log((ZZZ+(1+sqrt2)*B)/(ZZZ+(1-sqrt2)*B));
for (i=0;i<NC; i++)
{
sum1 = 0.;
for (j=0;j<NC; j++)
sum1= sum1 + zz[j]*sqrt(a[i]*a[j])*(1-k[i][j]);
f[i]=(ZZZ-1)*b[i]/bm-log(ZZZ-B)-(2*sum1/am-b[i]/bm)*am*lnz[0];
phil[i]=exp(f[i]);
fug[i]=P*zz[i]*phil[i];
};
}
```

B.1.8 InnerIterate.c

```
#include <stdio.h>
#include <math.h>
void InnerIterate(double Tr[], double x[], double y[], double kap[], double
a[], double am, double ab[], double b[], double bm, double ZL[], double ZV[],
double lnzv[], double k[][4], double R, double T, double P, double PropMatrix[][4],
double fugl[], int NC)
{
double A[1]={0.};
double B[1]={0.};
double fugv[4]={0., 0., 0., 0.};
int i,test1[4] = 0, 0, 0, 0;
```

```
     while (!((test1[0]==1)&&(test1[1]==1)&&(test1[2]==1)&&(test1[3]==1)))
     {
     PRmixParams(Tr, a, b, A, B, &am, &bm, ab, kap, y, R, T, P, PropMatrix,
k, NC);
     PRsolvEOS(ZL, ZV, A[0], B[0]);
     PRfug(fugv, lnzv, y, a, b, B[0], k, am, bm, ZV[0], R, T, P, NC);
     for (i=0;i<NC; i++)
     {
     if ((fabs(1-fugl[i]/fugv[i])>1e-5)&&(fugv[i]>=1e-10))
     y[i]=y[i]*fugl[i]/fugv[i];
     else
     test1[i]=1;
     };
     };
     }
```

B.1.9 cubic_solve.c

```
     #include <math.h>
     void cubic_solve(double x[], double k[])
     {
     double a, b, c;
     double pi, oot, opf, three, srth;
     double srd, tmp, u, v;
     double p, q, D;
     double cosphi, phi, cf;
     a = k[2];
     b = k[1];
     c = k[0];
     pi = 3.14159265358979;
     oot = 1.0/3.0;
     opf = 1.5;
     three = 3.0;
     srth = sqrt(three);
     p = (3.0*b-pow(a,2))/3.0;
     q = c+2.0*pow(a,3)/27.0-a*b/3.0;
     D = pow((p/3.0),3)+pow((q/2.0),2);
     if (D>=0)
     srd = sqrt(D);
     tmp = -0.5*q+srd;
     u = pow(fabs(tmp),oot);
     if (tmp <0)
     u = -u;
     tmp = -0.5*q-srd;
     v = pow(fabs(tmp),oot);
     if (tmp<0)
     v = -v;
```

```
x[0] = -a/3.0+u+v;
x[1] = 0.;
x[2] = 0.;
}
else
{
cosphi = -0.5*q/pow((fabs(p)/3.0),opf);
phi = acos(cosphi);
cf = 2.0*sqrt(fabs(p)/3.0);
x[0] = -a/3.0+cf*cos(phi/3.0);
x[1] = -a/3.0-cf*cos((phi-pi)/3.0);
x[2] = -a/3.0-cf*cos((phi+pi)/3.0);
}
}
```

BIBLIOGRAPHY

[1] Amirthalingam, R., S. W. Sung and J. H. Lee, "Two-step procedure for data-based modeling for inferential control applications", *AICHE J.*, Vol. 46, No. 10, 1974 to 1988, (2000).

[2] Bequette, B. W., "Nonlinear control of Chemical Processes: A Review", *I. & E. C. Res.*, Vol. 30, 1391 to 1413, (1991).

[3] Brosilow, C. and B. Joseph, "Techniques of Model-Based Control", Prentice Hall, NJ, (2002).

[4] Cattell, R. B., "The scree test for the number of factors", *Multivariate Behav. Res.*, 1:245-76, (1966).

[5] Chemstations Inc., "PROPANE-PROPYLENE SPLITTER (PROPSPLT)" in *ChemCAD Version 5: Book of Examples* , p. 15, August (1999).

[6] Chen, G., T. J. McAvoy and M. J. Piovoso, "A multivariate statistical controller for on-line quality improvement", *J. Proc. Control* , Vol. 8, No. 2, pp 139 to 149, (1998).

[7] Chen, R., K. Dave, T. J. McAvoy and M. Luyben, "A Nonlinear dynamic model of a Vinyl Acetate process", *I. & E. C. Res.* , Vol. 42, No. 20, pp 4478 to 4487, (2003).

[8] Chen, R. and T. J. McAvoy, "Plantwide Control System Design: Methodology and Application to a Vinyl Acetate Process", *I. & E. C. Res.* , Vol. 42, No. 20, pp 4753 to 4771, (2003).

[9] Davis, J. F., M. J. Piovoso, K. A. Hoo and B. R. Bakshi, "Process Data Analysis and Interpretation", *Adv. Chem. Eng.*, Vol. 25, pp. 2 to 97, (2000).

[10] de Jong, S., "SIMPLS: An Alternative Approach to Partial Least Squares Regression," Chemometrics and Intelligent Laboratory Systems, 18, 251 -263, (1993).

[11] Dong, D. and T. J. Mc Avoy, "Nonlinear Principal Component Analysis - Based on Principal Curves and Neural Networks", *Computers Chem. Eng.*, Vol. 20, No. 1, pp. 65 to 78, (1996).

[12] Downs, J. J. and E. F. Vogel, "A Plantwide Industrial Control Problem", *Computers Chem. Eng.*, Vol. 17, No. 3, pp. 245 to 255, (1993).

[13] Douglas, J. M., "Conceptual Design of Chemical Processes", McGraw-Hill, New York, (1988).

[14] Downs, J. J., *Personal communication*, (2003).

[15] Franklin, G. F., J. D. Powell and M. L. Workman, "Digital Control of Dynamic Systems", 3^{rd} Ed., Prentice Hall, Upper Saddle River, N.J. (1997).

[16] Geladi, P. and B. R. Kowalski, "Partial Least Squares: A tutorial", *Analytica Chimica Acta*, Vol. 185, 1 to 17, (1986).

[17] Gokhale, V., S. Hurowitz, and J. B. Riggs, "A dynamic model of a superfractionator: A test case for comparing distillatio control techniques", *Proceedings of DYCORD '95*, 4th IFAC Symposium, Helsingor (Denmark), 311 to 316, (1995).

[18] Gokhale, V., S. Hurowitz, and J. B. Riggs, "A Comparison of Advanced Distillation Control Techniques for a Propylene/Propane Splitter", *I. & E. C. Res.*, Vol. 34, 4413 to 4419, (1995).

[19] Hunter J., "Exponentially weighted moving average", *J. Qual. Tech.* , Vol. 18, pp. 203 to , (1986).

[20] Hurowitz, S., E., "Superfractionator Process Control", Ph.D. Thesis, Texas Tech University, Lubbock, TX, (1998).

[21] Jolliffe, I. T., "Principal Component Analysis", Springer-Verlag, New York, (1986).

[22] Jones, G. C., "Process Data Mining", presented at CPACT conference, Glasgow, UK, (2001).

[23] Kookos, I., K. and J. D. Perkins, "A Systematic Method for Optimum Sensor Selection in Inferential Control Systems", *I. & E. C. Res.* , Vol. 38, No. 11, pp 4299 to 4308, (1999).

[24] Kresta, J., J. Macgregor and T. E. Marlin, "Multivariate Statistical monitoring of Process Operating Performance", *Can. J. Chem. Eng.*, Vol. 35, pp. 69 to , (1991).

[25] Lee, J. H., M. Morari and C. E. Garcia, "State Space Interpretation of Model Predictive Control", *Automatica*, Vol. 30, pp. 707 to 717, (1994).

[26] Levinson, W., "Understand the basics of Statistical Process Coontrol", *Chemical Engieering Progress* , (November 1990).

[27] Ljung, L., "System Identification - Theory For the User" (2nd ed), PTR Prentice Hall, Upper Saddle River, N.J., (1999).

[28] Luyben, M. L. and B. D. Tyreus, "An industrial design/control study for the vinyl acetate monomer process ", *Computers Chem. Eng.*, Vol. 22, No. 7-8, pp. 867 to 877, (1998).

[29] Luyben, W. L., B. D. Tyreus, and M. L. Luyben, *Plantwide Process Control*, McGraw-Hill, New York (1999).

[30] Marlin, T. E., *Process Control: Designing Processes and Control Systems for Dynamic Performance*, McGraw-Hill, New York, 2/e (2000).

[31] McAvoy, T. J., "Model Predictive Statistical Process Control of Chemical Plants", *I. & E. C. Res.*, Vol. 41, 6337 to 6344, (2002).

[32] McAvoy, T. J. and N. Ye, "Base Control for the Tennesse Eastman Problem", *Computers Chem. Eng.*, Vol. 18, No. 5, pp. 383 to 413, (1994).

[33] McAvoy, T. J., N. Ye and G. Chen, "Nonlinear Inferential Parallel Cascade Control", *I. & E. C. Res.* , Vol. 35, No. 1, pp. 130 to 137, (1996).

[34] Morari M. and J. H. Lee, "Model predictive control: past, present and future", *Computers Chem. Eng.*, Vol. 23, No. 4-5, pp. 667 to 682, (1999).

[35] Morari M., M. Baotic and F. Borrelli, " Hybrid Systems Modeling and Control", *European J. Control*, Vol. 9, No. 2-3, pp. 177 to 189, (2003).

[36] Muske, K. R., "Estimating the Economic Benefit from Improved Process Control", *I. & E. C. Res.*, Vol. 42, No. 20, 4535 to 4544, (2003).

[37] Pannocchia, G. and A. Brambilla, "Consistency of Property Estimators in Multicomponent Distillation Control", *I. & E. C. Res.*, Vol. 42, to , (2003).

[38] Pearson, K. "On lines and planes of closest fit to systems of points in space", *Philos. Mag.*, Vol. 2, pp. 559 to 572, (1901).

[39] Piovoso, M. and K. Kosanovich, "Applications of Multivariate Statistical Methods to Process Monitoring and Controller Design", *Int J. Control*, Vol. 59, pp. 743 to 765, (1994).

[40] Prett, D. M. and M. Morari, "The Shell Process Control Workshop", Butterworths, MA, (1987).

[41] Qin, S. J. and T. A. Badgwell, "An Overview of Industrial Model Predictive Control Technology", *CPC - V, Tahoe City, CA.*, (1996).

[42] Qin, S. J. and T. A. Badgwell, "A survey of industrial model predictive control technology", *Control Engineering Practice*, Vol. 11, No. 7, pp. 733 to 764, (1996).

[43] Qin S.J. and R. Dunia, "Determining the number of principal components for best reconstruction", *J. Proc. Control*, Vol. 10, No. 2-3, pp. 245 to 250, (2000).

[44] Rao, C., "The Use and Interpretation of Principal Component Analysis in Applied Research", *Sankhya A.*, Vol. 26, pp. 329 to 358, (1964).

[45] Reyes-DeLeon, F. and W. L. Luyben, "Design and Control of Tubular Reactor Systems with Both Gas and Liquid Recycles", *I. & E. C. Res.*, Vol. 40, pp. 4089 to 4101, (2001).

[46] Riggs, J., "Chemical Process Control", Ferret Publishing, USA, (1999).

[47] Savitzky, A. and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures", *Analytical Chemistry* , Vol. 36, No. 8, pp. 1627 to 1639, (1964).

[48] Shewart, W., "Economic control of quality of manufactured products", Van Nostrand, New York, (1931).

[49] Sridhar, R. and D. J. Cooper, "A Tuning Strategy for Unconstrained SISO Model Predictive Control", *I. & E. C. Res.* , Vol. 36, pp. 729 to 746, (1997).

[50] Sridhar, R. and D. J. Cooper, "A Tuning Strategy for Unconstrained Multivariable Model Predictive Control", *I. & E. C. Res.* , Vol. 37, pp. 4003 to 4016, (1998).

[51] Valle S., W. H. Li, S. J. Qin, "Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods", *I. & E. C. Res.* , Vol. 38, No. 11, pp. 4389 to 4401, (1999).

[52] Van Overschee, P. and B. De Moor, "Subspace identification for linear systems", Kluver Academic Publ. Boston, (1996).

[53] Wise, B. and N. Gallagher, "The Process Chemometrics Approach to Process Monitoring and Fault Detection", *J. Process Cont.*, Vol. 6, pp. 2 to 97, (1996).

[54] Wise, B. M., N. Gallagher, R. Bro and J. M. Shaver, "MATLAB PLS Toolbox 3.0", *Eigenvector Research, Inc.*, (2003).

[55] Wise, B. M., N.L. Ricker, D.F. Veltkamp and B.R. Kowalski, "The Process Chemometrics Approach to Process Monitoring and Fault Detection", *Process Control and Quality*, Vol. 1, No. 1, pp. 41 to 51, (1990).

[56] Woodward, R. and P. Goldsmith, "Cumulative sum techniques", Oliver and Boyd, London, (1964).

[57] Yu, C. C. "Autotuning of PID controllers", Springer-Verlag, London, (1999).

[58] Zafiriou, E., "Course Notes for ENCH 736: Model based process control", Univ. of Maryland College Park, MD, (1998).

[59] Zheng, L., "Applications of Multivariate Statistical Analysis in Batch Processes", M.S. Thesis, Univ. of Maryland College Park, MD, (1999).

[60] Zheng, L., T. J. McAvoy, Y. Huang and G. Chen, "Application of Multivariate Statistical Analysis in Batch Processes", *I. & E. C. Res.* , Vol. 40, pp 1641 to 1649, (2001).