MULTIVARIATE STATISTICAL PROCESS MONITORING AND CONTROL

A Thesis Submitted in Partial Fulfillment for the Award of the Degree

Of

MASTER OF TECHNOLOGY
In
CHEMICAL ENGINEERING

By

SESHU KUMAR DAMARLA

Chemical Engineering Department
National Institute of Technology
Rourkela 769008

May 2011

MULTIVARIATE STATISTICAL PROCESS MONITORING AND CONTROL

A Thesis Submitted in Partial Fulfillment for the Award of the Degree

Of

MASTER OF TECHNOLOGY
In
CHEMICAL ENGINEERING

By

SESHU KUMAR DAMARLA

Under the guidance of

Dr. Madhusree Kundu

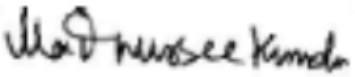
Chemical Engineering Department
National Institute of Technology
Rourkela 769008

May 2011

Department of Chemical Engineering
National Institute of Technology
Rourkela 769008 (ORISSA)

CERTIFICATE

*This is to certify that the thesis entitled "**Multivariate Statistical Process Monitoring and Control**", being submitted by Sri Seshu Kumar Damarla for the award of M.Tech. degree is a record of bonafide research carried out by him at the Chemical Engineering Department, National Institute of Technology, Rourkela, under my guidance and supervision. The work documented in this thesis has not been submitted to any other University or Institute for the award of any other degree or diploma.*



Dr. Madhushree Kundu
Department of Chemical Engineering
National Institute of Technology
Rourkela - 769008

ACKNOWLEDGEMENT

I would like to express my sincere gratefulness to my father who has struggled every minute of his life to keep me at higher level and shown a path in which I am traveling now. My each and every success always dedicates to him.

The second person l love and respect is Prof. Madhusree Kundu who has brought me from low level to certain level. She encourages and supports me in all the way. Her innovative thoughts and valuable suggestions made this dissertation possible. She inspires me through her attitude and dedication towards research. I always obliged to her. I never forget her. I wish and love to work with Prof. Madhusree Kundu wherever I will be.

I would like to express my indebtedness to Prof. Palash Kundu who has given wonderful ideas and an amazing path to work. Without his help we could not have proceeded further. He is hero off the screen.

I express my gratitude and indebtedness to Professor S. K. Agarwal and Professor K. C. Biswal, Head, Chemical Engineering Department, for providing me with the necessary computer laboratory and departmental facilities.

I express my gratitude and indebtedness to Prof. Basudeb Munshi, Prof. Santanu Paria, Prof. S. Khanam, Prof. A. Sahoo, Prof. A. Kumar, Prof. S. Mishra of the Department of Chemical Engineering, for their valuable suggestions and instructions at various stages of the work.

Today I am in some respectful position because of my two brothers- Pavan and Prabhas. They support and motivate me in every moment of my life.

Throughout these two years of journey I have wonderful friendship with Pradeep, Kasi, Satish garu, Jeevan garu, Tarangini akka.

I always responsible to my sisters and their families, who scarified their education and desires for my brother and me.

Finally, I express my humble regards to my mom who always thinks and worships God about us. She scarifies everything for us. Her tireless endeavors keep me good.

SESHU KUMAR DAMARLA
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008(ORISSA)

CONTENTS

ABSTRACT

Application of statistical methods in monitoring and control of industrially significant processes are generally known as statistical process control (SPC). Since most of the modern day industrial processes are multivariate in nature, multivariate statistical process control (MVSPC), supplanted univariate SPC techniques. MVSPC techniques are not only significant for scholastic pursuit; it has been addressing industrial problems in recent past.

. Monitoring and controlling a chemical process is a challenging task because of their multivariate, highly correlated and non-linear nature. Present work based on successful application of chemometric techniques in implementing machine learning algorithms. Two such chemometric techniques; *principal component analysis* (PCA) & *partial least squares* (PLS) were extensively adapted in this work for process identification, monitoring & Control. PCA, an unsupervised technique can extract the essential features from a data set by reducing its dimensionality without compromising any valuable information of it. PLS finds the latent variables from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor and response variables even if it is extended to time series data. In the present work, new methodologies; based on clustering time series data and moving window based pattern matching have been proposed for detection of faulty conditions as well as differentiating among various normal operating conditions of Biochemical reactor, Drum-boiler, continuous stirred tank with cooling jacket and the prestigious Tennessee Eastman challenge processes. Both the techniques emancipated encouraging efficiencies in their performances.

The physics of data based model identification through PLS, and NNPLS, their advantages over other time series models like ARX, ARMAX, ARMA, were addressed in the present dissertation. For multivariable processes, the PLS based controllers offered the opportunity to be designed as a series of decoupled SISO controllers. For controlling non-linear complex processes neural network based PLS (NNPLS) controllers were proposed. Neural network; a supervised category of data based modeling technique was used for identification of process

dynamics. Neural nets trained with inverse dynamics of the process or direct inverse neural networks (DINN) acted as controllers. Latent variable based DINNS' embedded in PLS framework termed as NNPLS controllers. (2×2), (3×3), and (4×4) Distillation processes were taken up to implement the proposed control strategy followed by the evaluation of their closed loop performances.

The subject *plant wide control* deals with the inter unit interactions in a plant by the proper selection of manipulated and measured variables, selection of proper control strategies. Model based *Direct synthesis* and DINN controllers were incorporated for controlling brix concentrations in a multiple effect evaporation process plant and their performances were compared both in servo and regulator mode.

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION TO MULTIVARIATE STATISTICAL PROCESS MONITORING AND CONTROL

Chapter 1

INTRODUCTION TO MULTIVARIATE STATISTICAL PROCESS MONITORING AND CONTROL

1.1 GENERAL BACKGROUND

Application of statistical methods in monitoring and control of industrially significant processes are included in a field generally known as statistical process control (SPC) or statistical quality control (SQC). The most widely used and popular SPC techniques involve univariate methods, that is, observing and analyzing a single variable at a time. Industrial quality problems are multivariate in nature, since they involve measurements on a number of characteristics, rather than one single characteristic. As a result, univariate SPC methods and techniques provide little information about their mutual interactions. The conventional SPC charts such as *Shewhart chart* and *CUSUM chart* have been widely used for monitoring univariate processes, but they do not function well for multivariable processes with highly correlated variables. Most of the limitations of univariate SPC can be addressed through the application of Multivariate Statistical Process Control (MVSPC) techniques, which consider all the variables of interest simultaneously and can extract information on the behavior of each variable or characteristic relative to the others. MVSPC research is having high value in theoretical as well as practical application and is certainly be conducive to process monitoring, fault detection and process identification & control.

1.2 DATA BASED MODELS

Data based modeling is one of the very recently added crafts in process identification, monitoring and control. The black box models are data dependent and model parameters are determined by experimental results /wet labs, hence these models are called data based models or experimental models. Unlike the white box models derived from first principles, the black box/data based models or empirical models do not describe the mechanistic phenomena of the process; they are based on input-output data and only describing the overall behavior of the process. The data based models are especially appropriate for problems that are data rich, but hypothesis and/or information poor. Sufficient numbers of quality data points are required to propose a good model. Quality data is defined by noise free data; free of outliers and is ensured by data pre conditioning.

The phases in the Data based modeling are:

- System analysis
- Data collection
- Data conditioning
- Key variable analysis
- Model structure design
- Model identification
- Model evaluation

In this era of data explosion, rational as well as potential conclusions can be drawn from the data, and in this regard, we owe a profound debt to multivariate statistics. Several Data-driven multivariate statistical techniques such as *Principal Component Analysis* (PCA), *Canonical Correlation Analysis* (CCA), *Partial Least Squares* (PLS), *Principal component regression* (PCR) and *Canonical Variate* and *Subspace State Space* modeling have been proposed for these purposes.

Data based models can be divided in to two major categories namely:

- ***Unsupervised models:*** These are the models which try to extract the different features present in the data without any prior knowledge of the patterns present in the data. Examples are Principal Component Analysis (PCA), Hierarchical

Clustering Techniques (Dendrograms), non-hierarchical Clustering Techniques (*K*-means).

- ***Supervised models:*** These are the models which try to learn the patterns in the data under the guidance of a supervisor who trains these models with inputs along with their corresponding outputs. Examples include Artificial Neural Networks (ANN), Partial Least Squares (PLS) and Auto Regression Models etc.

Efficient data mining, hence, efficient data based modeling will enable the future era to exploit the huge database available; in newer dimensions and perspective; embraced with never expected possibilities.

1.2.1 Chemometric Techniques

Chemometrics is the application of mathematical and statistical methods for handling, interpreting, and predicting chemical data. Present work is based on successful application and implementation of *Chemometric* techniques like PCA & PLS in process identification, monitoring & Control. PCA is a multivariate statistical technique that can extract the essential features from a data set by reducing its dimensionality without compromising any valuable information of it. PLS finds the latent variables from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor (***X***) variables and response (***Y***) variables. First proposed by Wold [1]; PLS has been successfully applied in diverse fields including process monitoring; identification of process dynamics & control and it deals with noisy and highly correlated data, quite often, only with a limited number of observations available. A tutorial description along with some examples on the PLS model was provided by Geladi and Kowalaski (1986) [2].

1.3 STATISTICAL PROCESS MONITORING

Process safety and product quality are the goals, people pursuit unremittingly. Monitoring a chemical processes is preferably a data based approach because of their multivariate, highly correlated and non-linear nature. Data collection has become a mature technology over the years and the analysis of process historical database has become an active area of research. [3-5]In order to ensure product quality and process safety, extraction of meaningful information from the process database seems to be a natural and logical

choice. Statistical performance monitoring of a process detects process faults or abnormal situations, hidden danger in the process followed by the diagnosis of the fault. The diagnosis of abnormal plant operation can be greatly facilitated if periods of similar plant performance can be located in the historical database. New methodologies based on clustering time series data and moving window based pattern matching can be used for detection of faulty conditions as well as differentiating among various normal operating conditions.

1.4 MULTIVARIATE STATISTICAL AND NEURO STATISTICAL CONTROLLER

Since the last decade, design & development of data based model control has taken its momentum. This very trend owes an explanation. Identification and control of chemical process is a challenging task because of their multivariate, highly correlated and non-linear nature. Very often there are a large number of process variables are to be measured thus giving rise to a high dimensional data base characterizing the process of interest. To extract meaningful information from such a data base; meticulous preprocessing of data is mandatory. Otherwise those high dimensional dataset maybe seen through a smaller window by projecting the data along some selected fewer dimensions of maximum variability. Principal component decomposition of the data into the latent subspaces provides us the opportunity to compress the data without losing meaningful information of it. Originally process static data were used for the aforesaid PCA decomposition.

Stationary time series data consisting of process input and output may be correlated by models like ARX (Auto regressive model with exogenous inputs). ARMA (Auto regressive moving average), ARMAX (Auto regressive moving average with exogenous inputs) ARIMA (Auto regressive integrated moving average) etc... This kind of identification of process dynamics requires a priori selection of model order and handle a lot of parameters, hence, encourages empiricism. Prior to apply those models for time series identification, firstly non-stationarity, if any; has to be removed from the data and secondly, the cross-correlation coefficients. have to be determined in order to detect the maximum effect of inputs having lag ranging from 0 , 1, 2, 3…. on specific outputs. Partial least squares technique with inner PCA core offers an alternative for identification of process dynamics. PLS reduces the dimensionality of the measured data, finds the latent variables from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor X variables and response Y variables.

In PLS based process dynamics, the inner relationship between (X) and (Y) scores; hence the process dynamics in latent subspace could not be well identified by linear or quadratic relationships. In view of this, neural identification of process dynamics in latent subspace or NNPLS based process identification deserves attention.

.Any Physical model based on first principle is designed on the basis of certain heuristics. The broad generality of such model is at low ebb when the process is non-linear and complex. Development of transfer function using such a model resulted in a poor controller performance irrespective of the controllers chosen. To get rid of this situation, data based models is a logical choice. Discrete time domain (z) transfer functions developed using ARX, ARMA, ARMAX models are less desirable for their extreme dependency on a large numbers of model parameters. In view of this, data driven PLS and NNPLS identified processes (transfer function) are logical alternative to design multivariable controllers.

Development of PLS and NNPLS controllers and their performance evaluation (both in servo and regulator mode) for some of the selected benchmark processes is one of the scopes of the present work. Discrete input-output time series data (X-Y) were used for this purpose. By combining the PLS with inner ARX model structure, non-linear dynamic processes could be modeled apart from using neural networks, which logically built up the framework for PLS based process controllers. The inverse dynamics of the latent variable based process was identified as the direct inverse neural controller (DINN) using the historical database of latent variables. The disturbance process in lower subspace was also identified using NN.

For multivariable processes, the Partial least squares (PLS) controllers offer the opportunity to be designed as a series of SISO controllers (Qin and McAvoy (1992, 1993))[6-7]. Because of the diagonal structure of the dynamic part of the PLS model, input-output pairings are automatic. Series of SISO controllers designed on the basis of the dynamic models identified into latent subspaces and embedded in the PLS framework are used to control the process. Till date there is no reference on NNPLS controllers in the open literature though PLS & NNPLS based process identification, PLS controllers are well documented.

1.5 PLANTWIDE CONTROL

Most of the industrial processes are having multiple units, which interact with each other. The subject *plant wide control* deals with these inter unit interactions by the proper selection of manipulated and measured variables, selection of proper control strategies. The

goal of any process design is to minimize capital costs while operating with optimum utilization of material and energy. Recycle streams in process plants have been a source of major challenge for control system design.

SIMULINK, can be used as the platform for *plant wide* process simulation. Apart from using classical controllers, artificial neural network (ANN) based controllers can be used to control the *plant wide* process of interest. Over the last decade, a focused R&D activity on *plant wide* control has been taken up by several researchers [8-11].

1.6 OBJECTIVE

In the context of aforesaid discussion, it is worthy to mention that application of multivariate statistical process monitoring and control deserves extensive research before its widespread use and commercialization. For very complex and non-linear processes, data based process monitoring, identification & control seems to be unequivocally superior to physical model based approaches. Expert systems or knowledge based system can be developed by integration of process knowledge derived from the process data base. In view of this, the objectives of the present dissertation are as follows:

- Multivariable process monitoring using MVSPC techniques
- Process identification in latent subspaces and multivariate statistical process control for benchmark multivariable processes
- *Plant wide* process control.

1.7 ORGANIZATION OF THE THESIS

First chapter renders an overview of multivariate process, chemometric techniques and their use in process monitoring, identification of process dynamics, control as well as *plant wide* process control. This chapter also presents the objectives of the present work with the thesis outline. The second chapter emphasized on different chemometric techniques used with the mention of significant previous contributions in this field. In chapter three statistical process monitoring algorithms with case studies are presented. Application of statistical/ neuro statistical process identification and control and plant wide control are the excerpts of the fourth chapter. In an ending note, the fifth chapter concludes with recommendation of future research initiatives.

REFERENCE

1. Wold H., "Estimation of principal components and related models by iterative least squares." In MultiVariate Analysis II. Krishnaiah, P. R., Ed., Academic Press, New York, 1966: 391-420.
2. Geladi, P., Kowalski, B. R., "Partial least-squares regression: A tutorial." Anal. Chim. Acta.1986, 185: 1-17.
3. Agrawal, R., Stoloroz, P., Piatetsky, S. G., Eds. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining; AAAI Press: Menlo Park, CA, 1998.
4. Apte´, C., "Data mining: an industrial research perspective." IEEE Trans. Comput. Sci. Eng. 1997, 4 (2): 6-9.
5. Fayyad, U., Piatetsky, S. G., Smyth, P., "The KDD process for extracting useful knowledge from volumes of data." Commun. ACM. 1996, 39: 27-34.
6. Qin, S. J.,McAvoy, T. J., "Nonlinear PLS modeling using neural network." Comput. Chem. Engr.1992, 16(4): 379-391.
7. Qin, S. J.,"A statistical perspective of neural networks for process modelling and control," In *Proceedings of the 1993 International Symposium on Intelligent Control*, Chicago, IL, 1993; pp 559-604.
8. Luyben, W. L., Tyreus, B. D., Luyben, M. L., "Plant-wide process control." McGraw Hill, New York, 1999.
9. Stephanopoulos, G., "Chemical process control." Prentice Hall, Englewood Cliffs, NJ., 1984.
10. Douglas, J. M., "Conceptual design of chemical processes." McGraw Hill, New York, 1988.
11. Downs, J. J., Vogel, E. F., "A plant-wide industrial process control problem." Comput. Chem. Engr. 1995, 17: 245-255.

Chapter 2

DATA DRIVEN MULTIVARIATE STATISTICAL TECHNIQUES AND PREVIOUS WORK

Chapter 2

CHEMOMETRIC TECHNIQUES AND PREVIOUS WORK

2.1 INTRODUCTION

Chemometric is the science relating measurements made on a chemical system to the state of the system via application of mathematical or statistical methods. The data based approaches; supervised learning, unsupervised learning and multivariate statistical techniques are falling under this category. Data based modeling is one of the very recently added crafts in process identification, monitoring and control. PCA belongs to unsupervised and PLS belongs to supervised category of *chemometric* models, which have been the mathematical foundation of present work. A parsimonious presentation of the statistical rationale is included in the chapter. *Chemometric* techniques offer the new approaches for process monitoring & control, hence it might be appropriate to describe the state of art of those advancements.

2.2 THEORETICAL POSTULATION ON PCA

PCA is a MVSPC technique used for the purpose of data compression without losing any valuable information. Principal components (PCs) are transformed set of coordinates orthogonal to each other. The first PC is the direction of largest variation in the data set. The

projection of original data on the PCs produces the score data or transformed data as a linear combination of those fewer mutually orthogonal dimensions. PCA technique was applied on the auto-scaled data matrix to determine the principal eigenvectors, associated eigen values and scores or the transformed data along the principal components. The drawbacks are that the new latent variables often have no physical meaning and the user has a little control over the possible loss of information.

Generally, PCA is a mathematical transform used to find correlations and explain variance in a data set. The goal is to map the raw data vector *E* onto vectors *S*, where, the vector x can be represented as a linear combination of a set of *m* orthonormal vectors u_i

$$x = \sum_{i=1}^{m} z_i u_i \tag{2.1}$$

where the coefficients z_i can be found from the equation $z_i = {u_i}^T x$. This corresponds to a rotation of the coordinate system from the original x to a new set of coordinates given by z. To reduce the dimensions of the data set, only a subset ($k < m$) of the basic feature vectors are preserved. The remaining coefficients are replaced by constants b_i and each vector x is then approximated as

$$\check{x} = \sum_{i=1}^{m} z_i u_i + \sum_{i=1}^{d} b_i u_i \tag{2.2}$$

The basic vectors u_i are called principal components which are equal to the eigenvectors of the covariance matrix of the data set. The coefficients $b_i$and the principal components should be chosen such that the best approximation of the original vector on an average is obtained. However, the reduction of dimensionality from *m* to *k* causes an approximation error. The sum of squares of the errors over the whole data set is minimized if we select the vectors $u_i$that correspond to the largest Eigen values of the covariance matrix. As a result of the PCA transformation, the original data set is represented in fewer dimensions (typically 2-3) and the measurements can be plotted in the same coordinate system. These plots show the relation between different observations or experiments. Grouping of data points in those bi-plots suggest some common properties and those can be used for classification.

Considering the following matrix:

$$\mathbf{X}=\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{2.3}$$

where, each row in $\boldsymbol{X}$ represents one measurement and the number of columns *m* is equal to the length of the measurement sequence or features. Following the step described above, the covariance matrix$\boldsymbol{C} = cov(\boldsymbol{X})$ and its Eigen values λ were calculated. Its eigenvectors $u_i$form

an orthonormal basis $\boldsymbol{U} = [u_1, u_2, u_3, \ldots u_m]$; that is $\boldsymbol{U^T U} = 1$. The original data set can be represented in the new basis using the relation: $\boldsymbol{Z = U^T X}$**.** After this transformation, a new data matrix of reduced dimension can be constructed with the help of Eigen values of the matrix $\boldsymbol{C}$. This is done by selecting the highest values λ since they correspond to the principal components with highest significance. The number of PCs to be included should be high enough to ensure good separation between the classes. Principal components with low contribution (low values of λ) should be neglected. Let the first *k* PCs as new features be selected neglecting the remaining (*m-k*) principal components. In this way, a new data matrix *D* of dimension $n \times k$ was obtained.

$$\boldsymbol{D}=\begin{bmatrix} z_{11} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nk} \end{bmatrix} \tag{2.4}$$

The PCA score data sets are grouped into number of classes following the rule of nearest neighborhood clustering algorithm. In the present work PCA based similarity was used for process monitoring purposes.

2.3 SIMILARITY

Similarity measures among the datasets pertaining to various operating conditions of a process provide the opportunity to cluster them into various groups, hence their classification. Fault detection and diagnosis (FDD), is comfortably relying on this principle which is conducive of safe plant operation. Apart from *Euclidian* and *Mahalanobis* distance, recently some new criterions have been proposed to determine the similarity or dissimilarity among the process historical database**.**

2.3.1 PCA Based Similarity

PCA similarity factor was developed by choosing largest k principal components of each multivariate time series dataset that describe at least 95 % of variance in the each dataset. These principal components are the eigen vectors of the covariance matrix. The PCA similarity factor between two datasets is defined by equation (2.5)

$$S_{PCA} = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} Cos^2\theta_{ij} \tag{2.5}$$

Where k is the number of selected principal components in both datasets, θ_{ij} is the angle between the i^{th} principal component of X_1 and j^{th} principal component of X_2. When first two principal components explain 95% of variance in the datasets, S_{PCA} may not capture the degree of similarity between two datasets because it weights all PCs equally. Obviously S_{PCA} has to modify to weight each PC by its explained variance. The modified S_{PCA}^{λ} is defined as

$$S_{PCA}^{\lambda} = \frac{\sum\sum(\lambda_i^{(1)}\lambda_j^{(2)})Cos^2\theta_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i^{(1)}\lambda_j^{(2)}} \tag{2.6}$$

Where $\lambda_i^{(1)}, \lambda_i^{(2)}$ are the eigen values of the first and second datasets respectively.

2.3.2 Distance Based Similarity

In addition to above similarity measure, distance similarity factor can be used to cluster multivariate time series data. Distance similarity factor compares two datasets that may have similar spatial orientation. The distance similarity finds its worth when the process variables pertaining to different operating conditions may have similar principal components. The distance similarity factor is defined as

$$S_{dist} = 2\times\frac{1}{\sqrt{2\pi}}\int_{\phi}^{\infty} e^{\frac{-z^2}{2}}dz = 2\times[1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\phi} e^{\frac{-z^2}{2}}dz \tag{2.7}$$

Where $\Phi = \sqrt{(\bar{x}_2-\bar{x}_1)\Sigma_1^{*-1}(\bar{x}_2-\bar{x}_1)^T}$, $\bar{x}_2$ & $\bar{x}_1$ are sample means row vector, Σ_1 is the covariance matrix for dataset X_1 and Σ_1^{*-1} is pseudo inverse of X_1. Dataset X_1 is assumed to be a reference dataset. In equation (2.7), a one side Gaussian distribution is used because $\Phi \geq 0$. The error function can be calculated by using any software or standard error function tables. The integration in equation (2.7) normalizes S_{dist} between zero and one.

2.3.2 Combined Similarity Factor

The combined similarity factor (*SF)* combines S_{PCA}^{λ} and S_{dist} using weighted average of the two quantities and used for clustering of multivariate time series data. The combined similarity is defined as

$$SF = \alpha_1 S_{PCA}^{\lambda} + \alpha_2 S_{dist} \tag{2.8}$$

The selection of α_1 and α_2 is up to the user but ensure that the sum of them is equal to one. In this work we selected the values of α_1 & α_2 are 0.67 and 0.33.

2.4 THEORETICAL POSTULATION ON PLS

Projection to latent structures or partial least squares (PLS) is a multivariable statistical regression method based on projecting/viewing the information in a high-dimensional data space down onto a low dimensional one defined by some latent variables. It selects the latent variables so that variations in predictor data ***X*** which is most predictive of ***Y*** data. PLS already has been successfully applied in diverse fields including process monitoring and quality control; identification of process dynamics & control and deals with noisy and highly correlated data, quite often, only with a limited number of observations available. When dealing with nonlinear systems, this approach assumes that the underlying nonlinear relationship between predictor data (***X***) and response data (***Y***) can be approximated by quadratic PLS (QPLS) or neural network based PLS (NNPLS) while retaining the outer mapping framework of linear PLS algorithm.***X*** and ***Y*** matrices were auto-scaled before they were processed by PLS algorithm. PLS model consists of outer relations (***X&Y***data are expressed in terms of their respective scores) and inner relations that links ***X*** data to ***Y***data in the latent subspace.

2.4.1 Linear PLS

X and *Y* matrices are scaled in the following way before they are processed by PLS algorithm.

$$X = XS_X^{-1} \text{ and } Y = YS_Y^{-1} \tag{2.9}$$

Where $S_X = \begin{bmatrix} s_{x1} & 0 \\ 0 & s_{x2} \end{bmatrix}$ and $S_Y = \begin{bmatrix} s_{y1} & 0 \\ 0 & s_{y2} \end{bmatrix}$

The S_X and S_Y are scaled matrices.

The outer relationship for the input matrix and output matrix with predictor variables can be written as

$$X = t_1 p_1^T + t_2 p_2^T + \ldots\ldots\ldots\ldots + t_n p_n^T + E = TP^T + E \quad (2.10)$$

$$Y = u_1 q_1^T + u_2 q_2^T + \ldots\ldots\ldots\ldots + u_n q_n^T + F = UQ^T + F \quad (2.11)$$

Where $\boldsymbol{T}$and $\boldsymbol{U}$represents the matrices of scores of $\boldsymbol{X}$ and $\boldsymbol{Y}$ data while $\boldsymbol{P}$and $\boldsymbol{Q}$represent the loading matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$. If all the components of $\boldsymbol{X}$and $\boldsymbol{Y}$ are described, the errors $\boldsymbol{E}$&$\boldsymbol{F}$become zero. The inner model that relates $\boldsymbol{X}$ to$\boldsymbol{Y}$ is the relation between the scores $\boldsymbol{T}$&$\boldsymbol{U}$.

$$\boldsymbol{U} = \boldsymbol{TB} \quad (2.12)$$

Where B is the regression matrix. The response $\boldsymbol{Y}$ can now be expressed as:

$$\boldsymbol{Y} = \boldsymbol{TBQ^T} + \boldsymbol{F} \quad (2.13)$$

To determine the dominant direction of projection of $\boldsymbol{X}$and $\boldsymbol{Y}$ data, the maximization of covariance within $\boldsymbol{X}$ and $\boldsymbol{Y}$ is used as a criterion.The first set of loading vectors p_1 and q_1 represent the dominant direction obtained by maximization of covariance within $\boldsymbol{X}$and$\boldsymbol{Y}$. Projection of $\boldsymbol{X}$ data on p_1 and $\boldsymbol{Y}$ data on q_1 resulted in the first set of score vectors $t_1$and u_1 , hence the establishment of outer relation. The matrices $\boldsymbol{X}$and $\boldsymbol{Y}$ can now be related through their respective scores, which is called the inner model, representing a linear regression between $t_1$and u_1: $\hat{u}_1 = t_1 b_1$. The calculation of first two dimensions is shown in Fig. 2.1.

The residuals are calculated at this stage is given by the following equations.

$$E_1 = X - t_1 p_1^{'} \quad (2.14)$$

$$F_1 = Y - u_1 q_1^{'} = Y - t_1 b_1 q_1^{'} \quad (2.15)$$

The procedure for determining the scores and loading vectors is continued by using the newly computed residuals till they are small enough or the number of PLS dimensions required are exceeded. In practice, the number of PLS dimensions is calculated by percentage of variance explained and cross validation. For a given tolerance of residual, the number of principal components can be much smaller than the original variable dimension. To get $\boldsymbol{T}$ and

$\boldsymbol{P}$matrices iteratively, deduction of (tp^T) from $\boldsymbol{X}$ keeps continuing until the given tolerance gets satisfied.This is the so called NIPALS (non-linear iterative PLS) algorithm.$\boldsymbol{U}$and $\boldsymbol{Q}$ matrices are also iteratively found using NIPALS. The irrelevant directions originating from noise and redundancy are left as $\boldsymbol{E}$ and $\boldsymbol{F}.\boldsymbol{Y}$ and $\boldsymbol{X}$ matrices are related following equation (2.13). The multivariate regression problems are translated into series of univariate regression problems with the application of PLS.

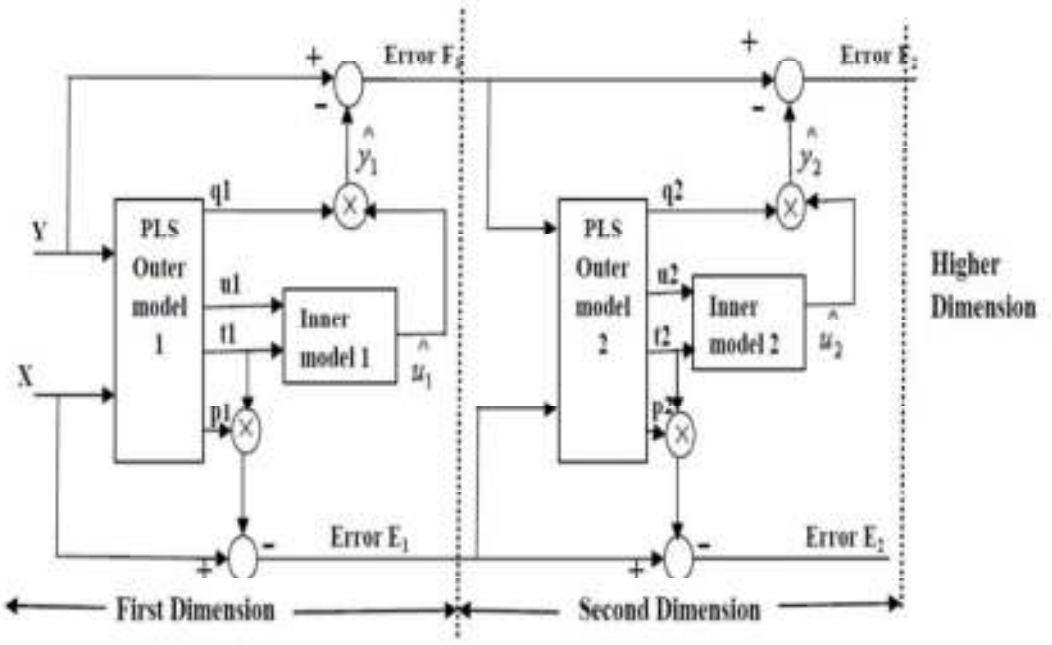


Fig. 2.1 Standard linear PLS algorithm.

2.4.2 Dynamic PLS

For incorporation of linear dynamic relationship in a time series data in the PLS framework, the decomposition of X block is given by equation (2.10), the dynamic analogue of equation (2.11) is as follows:

$$Y = G_1(t_1)q_1^T + G_2(t_2)q_2^T + \cdots G_n(t_n)q_n^T + F = Y_1^{exp} + Y_2^{exp} + .. + Y_n^{exp} + F \qquad (2.16)$$

where G_i s denote the linear dynamic model identified at each time instant by ARX model as well as FIR model and $G_i(t_i)q_i^T$ is a measure of Y space explained by the i th PLS dimension in latent subspaces.$Y_1^{exp}$is the experimental y data in 1st dimension. $G$is the diagonal matrix comprising the dynamic elements identified at each of the n th latent subspaces. Equation

(2.17) represents the linear second order ARX structure, which relates input and output signals of a time series data.

$$y(k) + a_1 y(k-1) + a_2 y(k-2) = b_1 x(k-1) + b_2 x(k-2) \quad (2.17)$$

where$y(k)$ =output at k^{th} instant, $x(k)$=input. The parameters of the ARX based inner dynamic models relating the scores $T\&U$ were estimated by least square regression. The ARX based input matrix used in regression analysis is as follows:

$$X_{ARX} = \{U_{k-1}, U_{k-2}, T_{k-1}, T_{k-2}\} \quad (2.18)$$

Finite Impulse Response Model or FIR model is also tested for inner model development. The FIR based input matrix is as follows:

$$X_{FIR} = \{T_{k-1}, T_{k-2}, T_{k-3}, T_{k-4}\} \quad (2.19)$$

$T$and Urepresent the matrices of scores of X and Y, respectively. The identified process transfer-function in discrete time domain:

$$G_p(z) = \frac{U(z)}{T(z)} = {b_1 z + b_2}/{z^2 + a_1 z + a_2} \quad (2.20)$$

The post compensation of U matrix (PLS inner dynamic model output) with loading matrix Q provided the PLS predicted output Y. The input matrix T to the PLS inner dynamic model was generated by post compensating the original X matrix with loading matrix$P$.Figure 2.2 represents the PLS based identification of process dynamics. Prior to dynamic modeling, order of the model should be selected. It is difficult to choose the order of the model. Autocorrelation signals renders a good indication about order that depends on how many past input and past output values taken in the input matrix for FIR and ARX models. The model parameters for both ARX and FIR models are estimated by linear least square technique.

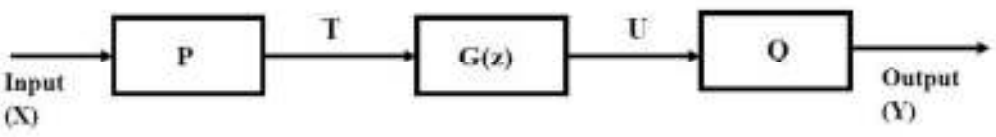


Fig. 2.2 Schematic of PLS based dynamics

2.5 PREVIOUS WORK

Data mining of process historical database has received attention in the computer science literature; however problems involving time-series identification & control have been addressed only recently. It is essential to extract required information regarding abnormality in the data to make a detection and diagnosis of abnormality in the process. Traditional clustering techniques were extensively used for grouping of the similar objects having analogous characteristics into the same cluster. The information acquired from the clusters is helpful in decision making, fault detection and process monitoring. Sudjianto & Wasserman (1996) and Trouve & Yu (2000).have successfully applied principal component analysis to extract features from large datasets **[1-2]. Kano et al. (2000)** implemented various PCA based statistical process monitoring methods using simulated data obtained from the Tennessee Eastman process **[3]**. Wise and Gallagher (1996) reviewed some of the *chemometric* techniques and their application to chemical process monitoring and dynamic process modeling [**4**].Numerous clustering techniques are available for clustering of univariate time series data [5-7]. Very few researchers have reported the clustering of dynamic and high dimensional multivariate time series data based on similarity measures. Wang and McGreavy (1998) used clustering methods to classify abnormal behavior of a refinery fluid catalytic cracking process **[8]**. Ng and Huang (1999) also used a clustering approach to classify different types of stars based on their light curves **[9]**. Johannesmeyer and Seborg (1999) developed an efficient technique to locating similar records in the historical database using PCA similarity factors **[10]**. Huang et al (2000) used PCA to cluster multivariate time series data by splitting large clusters into small clusters based on the percentage of variance explained by principal component analysis. This method is restrictive if the number of principal components is not known a priori and because predetermined principal components are inadequate for some operating conditions. Such feature extracting serves as dimensionality reduction of the datasets **[11]**. Singhal & Seborg (2005) used PCA and *Mahalonbis* distance similarity measures for location of similar operating condition in large multivariate database **[12]**. In juxtaposition, Kavitha & Punithavalli (2010) claimed that all traditional clustering or unsupervised algorithms are inappropriate for real time data **[13]**.

The diagnosis of abnormal plant operation can be greatly facilitated if periods of similar plant performance can be located in the historical database. Singhal and Seborg (2002) proposed a novel methodology based on PCA and distance similarity for pattern

matching problems. At a given time period of interest; for a multivariate time series data or template data, a similar pattern can be located in the historical database using the proposed pattern matching algorithm **[14]**.

Development of the multivariate statistical controller is one of the objectives of the present dissertation. Partial least square is one of the important data based multivariable statistical process control (MVSPC) techniques. It finds the latent variables (through principal component decomposition; PCA) from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor (X) variables and response (Y) variables.When dealing with nonlinear systems, the underlying nonlinear relationship between predictor variables (X) and response variables (Y) can be approximated by quadratic PLS (QPLS) or splines. Sometimes it may not function well when the non-linearities cannot be described by quadratic relationship. Qin and McAvoy (1992; 1993) suggested a new approach to replace the inner model by neural network model followed by the focused R& D activities taken up by several other researchers like Wilson et al. (1997); Holcomb & Morari (1992); Malthouse et al. (1997); Zhao et al. (2006); Lee et al. (2006)**[15-21]**. This approach of NNPLS employs the neural network as inner model keeping the outer mapping framework as linear PLS algorithm. Kaspar and Ray (1993) demonstrated their approach for identification and control problem using linear PLS models **[22]**. Lakshminarayanan et al., (1997) proposed the ARX/Hammerstein model as the modified PLS inner relation and used successfully in identifying dynamic models and proposition of PLS based feed forward and feedback controllers **[23]**.

Since the last 5 years or so, design & development of data based monitoring & control has been the focused area for applied research. It might not be inappropriate here to mention some of the efforts addressing diversified fields of applications. . Akamatsu et al. (2000) proposed data based control of an industrial tubular reactor **[24]**. Christian Rosen (2001) applied chemometric approach to process monitoring and control for wastewater treatment plant **[25]**. Cerrillo and MacGregor (2005) used latent variable models (LVM) for controlling batch processes **[26]**. Magda and Ordonez (2008) implemented multivariate statistical process control and case based reasoning for situation assessment of sequencing batch reactors [**27]**. Kano et al. (2008) implemented dynamic partial least squares regression in developing inferential control system of distillation compositions **[28]**. AlGhazzawi and Lennox (2009) developed MPC condition monitoring tool based on multivariate statistical process control (MSPC) techniques **[29]**. Laurí et al. (2010) implemented data driven latent variable model based predictive control (LV-MPC) for a 2×2 distillation process claiming their proposition

outperformed the conventional data driven MPC in terms of reliable multi-step ahead predictions with reduced computational complexity and reference tracking**[30]**. They have also chronicled that how the latent variable models (LVM) in recent years were implemented for various purposes.

Multiloop (decentralized) conventional control systems (especially PID controllers) with decouplers are often used to control interacting multiple input multiple output processes because of their ease in understandability apart from MPC (Model predictive), IMC based multivariable controllers.. In this context, Damala and Kundu (2010) decoupled a multivariable bioreactor process and used the inverse dynamics of the decoupled process to create a series of neural network based SISO (NN-SISO) controllers which were tuned independently without influencing the performance of other loops **[31].** For multivariable processes, the Partial least squares (PLS) controllers offer the opportunity to be designed as a series of SISO controllers instead of using multivariable controllers. Till date there is no reference on NNPLS controllers in the open literature though PLS controllers are well documented. In view of this, present work aims towards the development of NNPLS controllers.

Over the last decade, a focused R&D activity on *plant wide control* has been taken up by several researchers [32-35]. Present study is aiming one incremental lip in this regard; by incorporating neural controllers beside the classical controllers for *plant wide* process control.

REFERENCES

1. Sudjianto, A., Wasserman, G. S., "A nonlinear extension of principal component analysis for clustering and spatial differentiation." IIE Trans. 1996, 28: 1023–1028.
2. Trouve, A., Yu, Y., "Unsupervised clustering trees by nonlinear principal component analysis." In Proc. 5th Intl. Conf. on Pat. Rec. and Image Analysis: New Info. Tech. Samara, Russia, 2000: 110–114.
3. Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., Bakshi, B., "Comparison of statistical process monitoring methods: Application to the Eastman challenge problem." Comput. Chem. Engr. 2000, 24: 175-181.
4. Wise, B. M., Gallagher, N. B., "The process chemometrics approach to process monitoring and fault detection." J. Process Contr. 1996, 6: 329–348.
5. Agrawal, R., Gehrke, J., Gunopulos, D., Raghavan, P., "Automatic subspace clustering of high dimensional data for data mining applications." In Proc. ACM SIGMOD Intl. Conf. on Management of Data. Seattle, WA, 1998: 94–105.
6. Anderberg, M. R., "Cluster analysis for applications." New York: Academic Press, 1973.
7. Allgood, G. O., Upadhyaya, B. R., "A model-based high-frequency matched filter arcing diagnostic system based on principal component analysis." In Proc. of the SPIE—The Intl. Soc. for Optical Engr. Orlando, FL, 2000, 4055: 430–440.
8. Wang, X. Z., McGreavy, C., "Automatic classification for mining process operational data." Ind. Eng. Chem. Res. 1998, 37: 2215–2222.
9. Ng, M. K., Huang, Z., "Data-Mining massive time-series astronomical data: challenges and solutions". Inform.Software Tech. 1999, 41: 545–556.
10. Johannesmeyer, M. C., Seborg, D. E., "Abnormal situation analysis using pattern recognition techniques." AIChE.Annual Meeting, Dallas. TX, 1999.
11. Huang, Y., McAvoy, T. J., Gertler, J., "Fault isolation in nonlinear systems with structured partial principal component analysis and clustering analysis." Can. J. Chem. Eng. 2000, 78: 569–577.
12. Singhal, A., Seborg, D. E., "Clustering multivariate time series data." J. Chemometr. 2005, 19: 427-438.
13. Kavitha, V., Punithavalli, M., "Clustering time series data stream-A literature survey." IJCSIS. 2010, 8: 289-294.
14. Singhal, A, Seborg, D. E., "Pattern matching in historical batch data using PCA."

IEEE Contr. Syst. Mag. 2002, 22: 53–63.

15. Qin, S. J., McAvoy, T. J., "Nonlinear PLS modeling using neural network." Comput. Chem. Engr.1992, 16(4): 379-391.
16. Qin, S. J., "A statistical perspective of neural networks for process modelling and control." In Proceedings of the 1993 Internation Symposium on Intelligent Control, Chicago, IL, 1993: 559-604.
17. Wilson, D. J. H., Irwin, G. W., Lightbody, G., "Nonlinear PLS using radial basis functions." Trans. Inst. Meas. Control. 1997, 19(4): 211-220.
18. Holcomb, T. R., Morari, M., "PLS/neural networks." Comput. Chem. Engr.1992, 16(4): 393-411.
19. Malthouse, E. C., Tamhane, A. C., Mah, R. S. H., "Nonlinear partial least squares". Comput. Chem. Engr.1997, 21 (8): 875-890.
20. Zhao, S. J., Zhang, J., Xu, Y. M., Xiong, Z. H., "Nonlinear projection to latent structures method and its applications." Ind. Eng. Chem. Res. 2006, 45: 3843-3852.
21. Lee, D. S., Lee, M. W., Woo, S. H., Kim, Y., Park, J. M., "Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant." Process Biochem. 2006, 41: 2050-2057.
22. Kaspar, M. H., Ray, W. H., "Dynamic modeling for process control." Chem. Eng. Sci. 1993, 48 (20): 3447-3467.
23. Lakshminarayanan, S., Sirish, L., Nandakumar, K., "Modeling and control of multivariable processes: The dynamic projection to latent structures approach." AIChE J. 1997, 43: 2307-2323.
24. Akamatsu' K.,Lakshminarayanan, S.,Manako, H., Takada, H., Satou, T., Shah,S. L., **"**Data-based control of an industrial tubular reactor." Control Eng. Pract. 2000, 8 (7): 783-790
25. Rosen, C., "A chemometric approach to process monitoring and control with applications to wastewater treatment operation." Ph. D Thesis, Department of Industrial Electrical Engineering and Automation, Lund University, Sweden, 2001.
26. Cerrillo, F., MacGregor, J., "Latent variable MPC for trajectory tracking in batch processes." J. Process Contr. 2005, 15: 651–663.
27. Magda, L., Ordonez, R., "Multivariate statistical process control and case based reasoning for situation assessment of sequencing batch reactors." Doctoral Thesis, Girona, Spain, 2008.
28. Kano, M., Nakagawa, Y., "Data-based process monitoring, process control, and

quality improvement: recent developments and applications in steel industry." Comput. Chem. Engr. 2008, 32: 12–24.

29. AlGhazzawi, A., Lennox, B., "Model predictive control monitoring using multivariate statistics." J. Process Contr. 2009, 19: 314–327.
30. Laurí, D., Rossiter, J. A., Sanchis, J., Martínez, M.,"Data-driven latent-variable model-based predictive control for continuous processes." J. Process Contr. 2010, 20 (10): 1207-1219.
31. Damarla, S. K., Kundu, M., "Design of multivariable controllers using a classical approach."IJCEA. 2010, 1 (2): 165-172.
32. Luyben, W. L., Tyreus, B. D., Luyben, M. L., "Plant wide process control." McGraw Hill, New York, 1999.
33. Stephanopoulos, G., "Chemical process control." Prentice Hall, Englewood Cliffs, NJ., 1984.
34. Douglas, J. M., "Conceptual design of chemical processes." McGraw Hill, New York, 1988.
35. Downs, J. J., Vogel, E. F., "A plant-wide industrial process control problem." Comput. Chem. Engr. 1995, 17: 245-255.

Chapter 3

MULTIVARIATE STATISTICAL PROCESS MONITORING

Chapter 3

MULTIVARIATE STATISTICAL PROCESS MONITORING

3.1 INTRODUCTION

Quality and safety are the two important aspects of any production process. With the invention of multisensory array, mature data capture technology, advancement in data collection, compression and storage, data driven process monitoring including product quality monitoring, fault detection and diagnosis are getting due attention and wide spread acceptance. In view of this, a MVSPC method; based on clustering time series data and moving window based pattern matching technique was adapted for successful process monitoring. Biochemical reactor, Drum-boiler, continuous stirred tank with cooling jacket and Tennessee Eastman challenge processes were taken up to implement the proposed monitoring techniques. Instead of using first hand plat data, the above processes were modeled using first principles and processes were perturbed at industrially relevant operating conditions including faulty ones to create vector time series databases.

3.2 CLUSTERING TIME SERIES DATA: APPLICATION IN PROCESS MONITORING

Clustering time series data depends on measures of similarity. Similarity factor depends on PCA; especially the angles between the principal components in the latent subspaces. PCA was successfully applied by Kourti & MacGregor (1996) and Martin

&Morris (1996)to cluster multivariate time series data [1-2].A modified *K*-means clustering algorithm using similarity measures as a convergence criterion has been used for clustering datasets pertaining to different operating conditions including faulty one. Cluster purity and efficiency; the two indices were determined as a performance index of the proposed method. Different kinds of similarity including PCA similarity, weighted PCA similarity and distance based similarity have already been discussed in Chapter 2

Monitoring time series data pertaining to various operating conditions depends upon successful discrimination followed by classification. Discrimination is concerned with separating distinct sets of objects (or observations) on a one-time basis in order to investigate observed differences when casual relationships are not well understood. The operational objective of classification is to allocate new objects (observations) to predefined groups based on a few well defined rules evolved out of discrimination analysis of allied group of observations. Discrimination and classification among the time series data can be done using multivariate statistics. In these procedures, an underlying probability model must be assumed in order to calculate the posterior probability upon which the classification decision is made. One major limitation of the statistical methods is that they work well only when the underlying assumptions are satisfied.

3.2.1 K-means Clustering Using Similarity Factors

Clustering technique is more primitive in that; no priori assumptions are made regarding the group structures. Grouping of the data can be made on the basis of similarities or distances (dissimilarities). The number of clusters K can be pre-specified or can be determined iteratively as a part of the clustering procedure. In general, the *K*-means clustering proceeds in three steps, which are as follows:

1. Partition of the items in to K initial clusters.
2. Assigning an item to the cluster whose centroid is nearest (distance is usually Euclidian). Recalculation of the centroid for the cluster receiving the new item and for the cluster losing that item.
3. Repeating the step-2 until no more reassignment takes place or stable cluster tags are available for all the items.

The *K*-means clustering has a specific advantage of not requiring the distance matrix as required in hierarchical clustering, hence ensures a faster computation. The time series data

pertaining to various operating conditions were discriminated and classified using the following similarity based *K*-means algorithm.

Given: Q datasets, $\{x_1, x_2, ...x_q ....x_Q\}$ to be clustered into K clusters

1. Let j^{th} dataset in the i^{th} cluster be defined as x_i^j . Computation of the aggregate dataset $X_i (i = 1,2,....k)$, for each of the k clusters as,

$$X_i = [(x_1^{(i)})^T .....(x_j^{(i)})^T .....(x_{Qi}^{(i)})]^T \tag{3.1}$$

Where Q_i is the number of datasets in the database. Note that $\sum_{i=1}^{k} Q_i = Q$.

2. Calculation of the dissimilarity between dataset x_q and each of the k aggregate datasets $X_i, i = 1,2....k$ as,

$$d_{i,q} = 1 - SF_{i,q} \tag{3.2}$$

Where $SF_{i,q}$ is similarity between q^{th} dataset and i^{th} cluster described by equation (2.6). Let the aggregate dataset X_i in equation (3.2) be the reference dataset. Dataset x_q is assigned to the cluster to which it is least dissimilar. Repetition of the aforesaid steps for Q numbers of datasets.

3.2.2 Selection of Optimum Number of Cluster

Selection of the number of clusters is crucial in *K*-means clustering algorithm. Rissanen (1978) proposed model building and model order selection to calculate the optimum number of clusters based on the model complexity [3]. Large number of resulted clusters indicates the model complexity. This method penalized the more complex model more than the less complex model. Several methods have been developed to identify optimum number of clusters in multivariate time series data such as Akaike Information Criterion (AIC) and Schwartz Information Criterion (SIC). However, preliminary results obtained using these methods were not promising. Later Smyth (1996) introduced the cross validation of model to identify optimum number of clusters in the data. In this method, data splits into two or more parts. One part is used for clustering whereas other part is used for cross validation [4]. Singhal & Seborg (2005) proposed new methodologies for finding the optimum number of clusters [5]. In the present work efficiently designed modified *K*-means clustering algorithm ensures the evolution of optimum number of clusters.

3.2.3 Clustering Performance Evaluation

Krzanowski (1979)developed PCA similarity factor by choosing largest k principal components of each multivariate time series dataset that describe at least 95 % of variance in the each dataset[6].Some key definitions were introduced by Singhal &. Seborg (2005) to evaluate the performance of the clusters obtained using similarity factors [5].

Assuming the number of operating conditions is N_{op} and the number of datasets for operating condition j in the database is N_{DBj}. Cluster purity is defined to characterize each cluster in terms of how many numbers of datasets for a particular operating condition present in the i^{th} cluster.

Cluster purity is defined as,

$$P_j \overset{\Delta}{=} \left(\frac{\max_j N_{ij}}{N_{pi}}\right) \times 100\% \tag{3.3}$$

Where N_{ij} is the number of datasets of operating condition j in the i^{th} cluster and N_{pi} is the number of datasets in the i^{th} cluster.

Cluster efficiency measures the extent to which an operating condition is distributed in different clusters. This method is to penalize the large values of K when operating condition j distributed into different clusters. Clustering efficiency is defined as,

$$\eta = \left(\frac{\max_i N_{i,j}}{N_{DBj}}\right) \times 100\% \tag{3.4}$$

Where N_{DBj} is the number of datasets for operating condition j in the database. Large number of datasets of operating condition present in a cluster can be considered as dominant operating condition.

3.3 MOVING WINDOW BASED PATTERN MATCHING

In this approach, the snapshot or template data with unknown start and end time of operating condition (in sample wise manner) moves through historical data and the similarity between them is characterized by distance and PCA based combined similarity factor [7-9].The snapshot data can also approach in as dataset. In order to compare the snapshot data to historical data, the relevant historical data are divided into data windows that are the same size as the snapshot data. The historical data sets are then organized by placing windows side-

by-side along the time axis, which results in equal length, non-overlapping segments of data. The historical data windows with the largest values of similarity factors are collected in a candidate pool and are called records to be analyzed by the process Engineer. For the present work, the historical data window moved one observation at a time, with each old observation is getting replaced by new one. Pool accuracy, Pattern matching efficiency and Pattern matching algorithm efficiency are important metrics that quantify the performance of the proposed pattern matching algorithm. The proposed pattern matching technique as depicted in Fig. 3.1is consisting three steps which are follow as:

1. Specification of the snapshot (variables and time period)
2. Comparison between snapshot and periods of historical windows using moving window
3. Collection of historical windows with the largest values of similarity factors

.

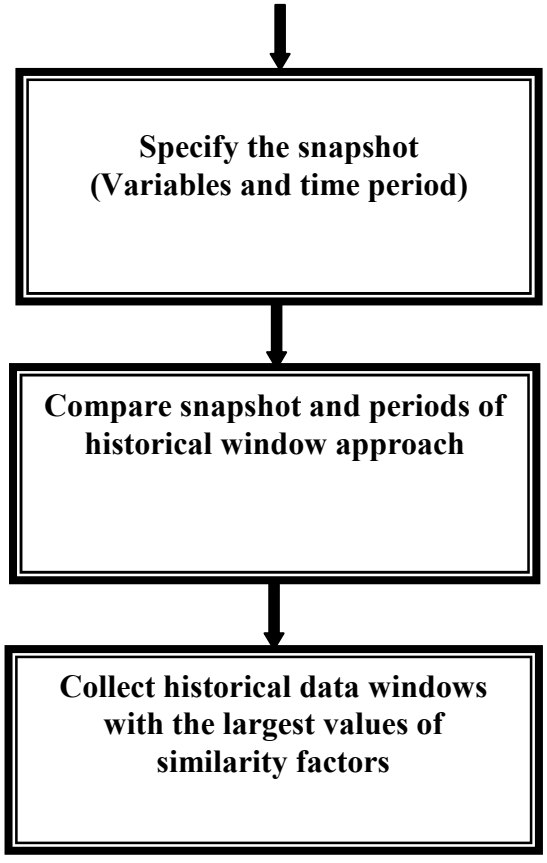


Fig. 3.1 Moving window based pattern matching

N_P: The size of the candidate pool, it is the number of historical data windows that have been labeled ‘‘similar’’ to the snapshot data by a pattern matching technique. The data windows collected in the candidate pool are called records.

N_1 = number of records in the candidate pool that are exactly similar to the current snapshot data, i.e. having a similarity of 1.0/or number of correctly identified record.

N_2 = number of records in the candidate pool that are not correctly identified.

$N_P = N_1 + N_2$

N_{DB}: The total number of historical data windows that are actually similar to the current snapshot. In general, $N_{DB} \neq N_P$

Pool accuracy, $P = (N_1 / N_P) \times 100\%$

Pattern matching efficiency, $H = [1 - ((N_P - N_1) / N_{DB})] \times 100\%$

Pattern matching algorithm efficiency, $\xi = (N_P / N_{DB}) \times 100\%$

A large value of Pool accuracy is important in case of detection of small number of specific previous situations from a small pool of records without evaluating incorrectly identified records. A large value of Pattern matching efficiency is required in case of detection of all of the specific previous situations from a large pool of records. The proposed method is completely data driven and unsupervised; no process models or training data are required. The user should specify only the relevant measured variables.

3.4 DRUM BOILER PROCESS

Drum boiler is crucial benchmark process in view of modeling and control system design. This process was addressed by Pellegrinetti & Bentsman (1996), Astrom & Bell (2000), Tan et al (2002), and Jawahar & Pappa (2005) [10-13]regarding modeling and control aspects. In the present work Drum-boiler process has been monitored using similarity measures. Drum-boiler model has been derived using first principles and is characterized by few physical parameters. The parameters used in the model are those from a Swedish Power Plant. The plant is an 80 MW unit in Sweden. 16 numbers of datasets belonging to four operating conditions including an abnormal operating condition were generated perturbing the process to evaluate the performance of the proposed techniques.

3.4.1 The Brief Introduction of Drum-boiler System

The utility boilers in the thermal/nuclear power plants are water tube drum boilers. This type of boiler usually comprises two separate systems. One of them is the steam–water system, which is also called the water side of the boiler. In this system preheated water from the economizer is fed into the steam drum, and then flows through the down-comers into the mud drum.

The diagram depicted in Fig. 3.2shows the drum that holds water at saturation or near saturation condition and denser water flows through the down-comer into lower header by force of gravity. After being heated up further, it returns to the drum through the riser. Between lower and upper header, are stretch of tubes that constitute water walls and receive radiant heat from furnace. Water walls permit use of high temperature of furnaces and combustion rates. Part of water in these tubes and risers evaporates, with the result that the fluid in the riser is composed of a mixer of water and steam. The difference in density of water in the riser and down-comer provides the necessary motive force to set up circulation of water in the boiler system.

Boiler is a reservoir of energy. Amount of energy stored in each part is a complicated function of temperature and pressure. The model can now be developed in detail, expressing stored energy, input power and output power as function of control variables and sate variables. Global mass and energy balances capture much of the behavior of the system.

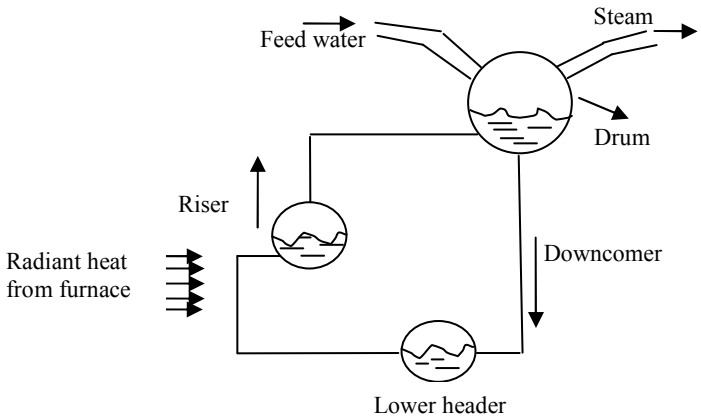


Fig. 3.2 Drum-boiler process

3.4.2 Modeling

Assumptions:

- The fundamental modeling simplification is that the two phases of the water inside the system are everywhere in saturated thermodynamic state.
- There is an instantaneous and uniform thermal equilibrium between water and metal everywhere.
- The energy stored in the steam and water is released or absorbed very rapidly when pressure changes. This is the key for understanding boiler dynamics. The rapid release of energy ensures that different parts of the boiler change their temperature in the same way.
- Steady state metal temperature is close to saturation temperature and the temperature differences are small dynamically.

Global mass and energy balance equations:

The inputs to the system are chosen to be

- Heat flow rate to the risers, Q
- Feedwater mass flow rate, q_f
- Steam flow rate, q_s

The outputs from the system are chosen to be:

- Drum level, L
- Drum pressure, P

A key feature of drum boiler is that there is an efficient energy and mass transfer between all parts that are in contact with steam. The mechanism responsible for heat transfer is boiling and condensation.

Global mass balance:

$$\frac{d[\rho_s V_{st}+\rho_w V_{wt}]}{dt} = q_f - q_s \tag{3.5}$$

Global Energy balance:

$$\frac{d[\rho_s V_{st} h_s+\rho_w V_{wt} h_w - PV_t + m_t c_p t_m]}{dt} = Q + q_f h_f - q_s h_s \tag{3.6}$$

Total Volume of Drum, Down-comer and Risers:

$$V_t = V_{st} + V_{wt} \tag{3.7}$$

Where

ρ_s and ρ_w represent the densities of steam and water respectively,

h_s and h_w represent the enthalpies of steam and water per unit mass,

V_{st} and V_{wt} represent the total steam and water volume in the system,

V_t, is the total volume of the drum,

m_t, is the total metal mass,

t_m, is the metal temperature

A simple drum boiler model is obtained by combining equations (3.5), (3.6)and(3.7)with saturated steam tables. Mathematically the model is a differential algebraic system.

3.4.2.1 Proposed Second Order Boiler Model

To have a better insight into the key physical mechanism that affect the dynamic behavior of the system the state variable approach is considered. Drum pressure, *P* is chosen as a key state variable, since it is the most globally uniform variable in the system and is also easily measurable. The variablesρ_s, ρ_w, h_s, h_w are expressed as function of steam pressure using the steam table. The second state variable is chosen to be the total volume of water in the system i.e. V_{wt}.Using equation (3.7), V_{st} is eliminated from equations (3.5) and (3.6).

The state equations then take the following form:

$$\left.\begin{aligned} e_{11}\frac{dV_{wt}}{dt} + e_{12}\frac{dP}{dt} &= q_f - q_s \\ e_{21}\frac{dV_{wt}}{dt} + e_{22}\frac{dP}{dt} &= Q + q_f h_f - q_s h_s \end{aligned}\right\} \quad (3.8)$$

Where

$$e_{11} = \rho_w - \rho_s$$

$$e_{12} = V_{st}\frac{\partial \rho_s}{\partial P} + V_{wt}\frac{\partial \rho_w}{\partial P}$$

$$e_{21} = \rho_w h_w - \rho_s h_s$$

$$e_{22} = V_{st}\left(h_s\frac{\partial \rho_s}{\partial P} + \rho_s\frac{\partial h_s}{\partial P}\right) + V_{wt}\left(h_w\frac{\partial \rho_w}{\partial P} + \rho_w\frac{\partial h_w}{\partial P}\right) - V_t + m_t c_p\frac{\partial t_s}{\partial P} \quad (3.9)$$

Equations (3.8) and (3.9) constitute a state model of the second order boiler system. This simplistic model captures the gross behavior of the boiler quite well. In particular, it describes the response of drum pressure to changes in input power, feed-water and steam flow rates reasonably well. But the model has a serious deficiency. It doesn't capture the behavior of the drum water level, as the distribution of steam and water are not taken into account. The drum level control is difficult due to shrink and swells effects. The drum level may be defined as the level at which water stands in the boiler. The steam level is the space above the water level.

3.4.2.2 *Distribution of Steam in Risers and Drum*

The behavior of drum-level can best described by taking into account the distribution of water and steam in the system. The distribution of water and steam is considered separately for the riser section and the drum.

3.4.2.2.1 *Distribution of steam and water in risers*

The steam-water distribution varies along the risers. In the riser section water exists in two phases namely the liquid-phase i.e. water and the vapor-phase i.e. steam. The mass fraction or dryness fraction of a liquid-vapor mixture must be defined prior to further discussion. In a liquid-vapor mixture, x is known as the quality.

$x = \frac{m_v}{m_v + m_l}$ where , m_v and m_l are the masses of vapor and liquid respectively in the mixture. The value of x varies between 0 and 1. In order to determine the average density of steam-water mixture in the riser, it is necessary to define the void fraction. The void fraction α of a two phase mixture is a volumetric quantity and is given as: α =(volume of vapor)/(volume of vapor + volume of liquid).

α and x are related by:

$$\alpha = \frac{1}{1+\left(\frac{1-x}{x}\right)\varphi} \text{ or } x = \frac{1}{1+\left(\frac{1-x}{x}\right)\frac{1}{\varphi}} \tag{3.10}$$

Where, $\varphi = \frac{\vartheta_f}{\vartheta_g} s$. v_f and v_g are the specific volumes of saturated liquid and vapor respectively.s is the slip ratio of two-phase mixture. The two phases of the mixture do not travel at the same speed. Instead there is a slip between them, which causes the vapor to move faster than liquid. s is a dimensionless number, greater than 1. It is defined as the ratio of average vapor velocity to the average liquid velocity, at any cross-section of the riser. The slip ratio is neglected in the present work, as it doesn't have a major influence on the fit to experiment data.

The behavior of two-phase flow is complicated and is typically modeled by partial differential equations. Keeping a finite dimensional model it is assumed that that shape of the distribution is known. The assumed shape is based on solving the partial differential equations in the steady state. There exists a linear distribution of steam-water mass ratio along the risers. The ratio varies in the following form:

$$\alpha_m(\xi) = \alpha_r \xi,\ 0 \le \xi \le 1 \tag{3.11}$$

Where ξ is a normalized length coordinate along the risers and α_r is the mass ratio at the riser outlet. The volume and mass fractions of steam are related through

$$\alpha_v = f(\alpha_m) \tag{3.12}$$

$$\text{Where, } f(\alpha_m) = \frac{\rho_w \alpha_m}{\rho_s + (\rho_w - \rho_s)\alpha_m} \tag{3.13}$$

For modeling the drum-level the total amount of steam in the drum is to be obtained. The governing equation is the average steam volume fraction in the risers, which is obtained by integrating the equation (3.13) over the limits 0 to 1 can be given as:

$$\bar{\alpha}_v = \int_0^1 \alpha_v(\xi)\, d\xi = \frac{1}{\alpha_r}\int_0^1 f(\alpha_r \xi)\, d(\alpha_r \xi) = \frac{1}{\alpha_r}\int_0^{\alpha_r} f(\xi)\, d\xi = \frac{\rho_w}{(\rho_w - \rho_s)}\left(1 - \frac{\rho_s}{(\rho_w - \rho_s)\alpha_r}\ln\left(1 + \frac{\rho_w - \rho_s}{\rho_s}\alpha_r\right)\right) \tag{3.14}$$

The partial derivatives of $\bar{\alpha}_v$ with respect to drum pressure and steam mass fraction are obtained as:

$$\left.\begin{aligned} \frac{\partial \bar{\alpha}_v}{\partial P} &= \frac{1}{(\rho_w - \rho_s)^2}\left(\rho_w \frac{\partial \rho_s}{\partial P} - \rho_s \frac{\partial \rho_w}{\partial P}\right)\left(1 + \frac{\rho_w}{\rho_s}\frac{1}{1+\eta} - \frac{\rho_s + \rho_w}{\eta \rho_s}\ln(1+\eta)\right) \\ \frac{\partial \bar{\alpha}_v}{\partial \alpha_r} &= \frac{\rho_w}{\rho_s \eta}\left(\frac{1}{\eta}\ln(1+\eta) - \frac{1}{1+\eta}\right) \end{aligned}\right\} \tag{3.15}$$

Where $\eta = \dfrac{\alpha_r(\rho_w - \rho_s)}{\rho_s}$

The transfer of mass and energy between steam and water by condensation and evaporation is a key element in the modeling. When modeling the phases separately the transfer must be accounted for explicitly, hence the joint balance equations for water and steam are written for the riser section.

3.4.2.2..2 *Lumped parameter model*

The global mass balance for the riser section is:

$$\frac{d[\rho_s \bar{\alpha}_v V_r + \rho_w(1-\bar{\alpha}_v)V_r]}{dt} = q_{dc} - q_r \tag{3.16}$$

Where, q_r, is the total mass flow rate out of the risers.,

q_{dc}, is the total mass flow rate into the risers.

The global energy balance for the riser section is:

$$\frac{d[\rho_s h_s \bar{\alpha}_v V_r + \rho_w h_w (1-\bar{\alpha}_v)V_r - PV_r + m_r c_p t_s]}{dt} = Q + q_{dc}h_w - (\alpha_r h_c + h_w)q_r \tag{3.17}$$

3.4.2.2.3 *Circulation flow:*

In the natural circulation boiler the flow rate is driven by density gradients in the risers and down-comers. The flow through the down-comer (q_{dc}) can be obtained from a momentum balance.

The equation consists of three terms namely the internal term, driving force that in this case is the density difference and frictional force in the flow through pipes.

(a) Inertia force: $(L_r + L_{dc})\frac{dq_{dc}}{dt}$

(b) Driving force: $(\rho_w - \rho_s)\bar{\alpha}_v V_r g$

(c) Friction force: $\frac{k{q_{dc}}^2}{2\rho_w A_{dc}}$

Combining above three terms momentum balance equation is written as following:

$$(L_r + L_{dc})\frac{dq_{dc}}{dt} = (\rho_w - \rho_s)\bar{\alpha}_v V_r g - \frac{k{q_{dc}}^2}{2\rho_w A_{dc}} \quad (3.18)$$

Equation (3.15) is a first order system that has a time constant as given below:

$$\tau = \frac{(L_r + L_{dc})A_{dc}\rho_w}{kq_{dc}} \quad (3.19)$$

The steady state relation for the system is given as:

$$\frac{1}{2}k{q_{dc}}^2 = \rho_w A_{dc}(\rho_w - \rho_s)\bar{\alpha}_v V_r g \quad (3.20)$$

3.4.2.3 *Distribution of Steam in the Drum*

The physical phenomenon in the drum is complicated. Steam enters from many riser tubes: feed water enters through a complex arrangement, water leaves through the down-comer tubes and steam through the steam valve. The geometry and flow patterns are complex. Basic mechanisms that occur in the drum are separation of water and steam and condensation.

The mass balance for the steam under the liquid level is given as:

$$\frac{d(\rho_s V_{sd})}{dt} = \alpha_r q_r - q_{sd} - q_{cd} \quad (3.21)$$

Where, q_{cd} is the condensation flow, which is given by

$$q_{cd} = \frac{h_w - h_f}{h_c}q_f + \frac{1}{h_c}\left(\rho_s V_{sd}\frac{dh_s}{dt} + \rho_w V_{wd}\frac{dh_w}{dt} - (V_{sd} + V_{wd})\frac{dP}{dt} + m_d c_p \frac{dt_s}{dt}\right) \quad (3.22)$$

The flow q_{sd} is driven by density difference of water and steam, and the momentum of the flow entering the drum. The expression for q_{sd} is an empirical model and is a good fit to the experimental data and is given as:

$$q_{sd} = \frac{\rho_s}{T_d}(V_{sd} - V_{sd}^0) + \alpha_r q_{dc} + \alpha_r \beta(q_{dc} - q_r) \quad (3.23)$$

Where, V_{sd}^{0}, is the volume of steam in the drum in hypothetical situation when there is no condensation of steam in the drum and T_d is the residence time of steam in the drum.

3.4.2.4 *Drum Level:*

Having accounted for distribution of steam below drum-level, now the drum-level is modeled. The drum level is composed of two terms,

- The total amount of water in the drum,
- The displacement due to changes of the steam-water ratio in the risers.

Derivation of the drum-level l measured from its normal operating level is given by:

$$l = \frac{V_{sd}+V_{wd}}{A_d} = l_w + l_s \tag{3.24}$$

Where $l_w = \frac{V_{wd}}{A_d}$ and $l_s = \frac{V_{sd}}{A_d}$

$$V_{wd} = V_{wt} - V_{dc} - (1-\bar{\alpha}_v)V_r$$

Where, V_{wd}, is the volume of water in the drum, l_w, is the level variation caused by changes of amount of water in the drum, l_s, is the level variation caused by the steam in the drum, A_d, is the wet surface of the drum at the operating level.

3.4.2.5 *The Model*

Model is given by the following equations:

$$\left.\begin{aligned}
&\frac{d[\rho_s V_{st}+\rho_w V_{wt}]}{dt} = q_f - q_s \\
&\frac{d\left[\rho_s V_{st} h_s + \rho_w V_{wt} h_w - PV_t + m_t c_p t_m\right]}{dt} = Q + q_f h_f - q_s h_s \\
&\frac{d[\rho_s \bar{\alpha}_v V_r + \rho_w (1-\bar{\alpha}_v)V_r]}{dt} = q_{dc} - q_r \\
&\frac{d\left[\rho_s h_s \bar{\alpha}_v V_r + \rho_w h_w (1-\bar{\alpha}_v)V_r - PV_r + m_r c_p t_s\right]}{dt} = Q + q_{dc} h_w - (\alpha_r h_c + h_w) q_r \\
&\frac{d(\rho_s V_{sd})}{dt} = \alpha_r q_r - q_{sd} - q_{cd} \\
&q_{cd} = \frac{h_w - h_f}{h_c} q_f \quad + \frac{1}{h_c}\left(\begin{array}{c}\rho_s V_{sd}\frac{dh_s}{dt} + \rho_w V_{wd}\frac{dh_w}{dt} - (V_{sd}+V_{wd})\frac{dP}{dt} + \\ m_d c_p \frac{dt_s}{dt}\end{array}\right) \\
&l = \frac{V_{sd}+V_{wd}}{A_d} \\
&V_t = V_{st} + V_{wt}
\end{aligned}\right\} \tag{3.25}$$

The model as can be seen is a differential algebraic system. Since most available simulation software requires state equations, the state model is also derived.

3.4.2.6 *State Variable Model*

Prior to generation of database, linear model is required to design controllers. This was accomplished by the use of state space methodology. The selection of state variables is done in many different ways. A convenient way is to choose those variables as states, which have a good physical interpretation that describe storage of mass, energy and momentum. The variables used in this procedure are as follows:

I. State variables:
   a. Drum pressure P
   b. Total water volume of the system V_{wt}
   c. Steam mass fraction α_r
   d. Volume of steam in the drum V_{sd}

II. Manipulated inputs:
   a. Heat flow rate to the risers Q
   b. Feed water flow rate to the drum q_f
   c. Steam flow rate from the drum q_s

III. Measured outputs:
   a. Total water volume of the system V_{wt}
   b. Drum pressure P

3.4.2.6.1 *Derivation of state equations*:

The pressure and water dynamics are obtained from the global mass and energy balances equations (3.5) and (3.6). Combining these equations the state variable form is obtained as given by the set of equations (3.8). The riser dynamics is given by the mass and energy balance equations (3.16) and (3.17) are further simplified by eliminating 'q_r' and multiplying equation (3.16) by '$(h_w + \alpha_r h_c)$' and adding it to equation (3.17).

The resulting expression is given as:

$$h_c(1-\alpha_r)\frac{d(\rho_s\bar{\alpha}_v V_r)}{dt} + \rho_w(1-\bar{\alpha}_v)V_r\frac{dh_w}{dt} - \alpha_r h_c\frac{d[\rho_w(1-\bar{\alpha}_v)V_r]}{dt} + \rho_s\bar{\alpha}_v V_r\frac{dh_s}{dt} - V_r\frac{dP}{dt} + m_r c_p\frac{dt_s}{dt} = Q - \alpha_r q_{dc} h_c \quad (3.26)$$

If the state variables 'P' and 'α_r' are known, the riser flow rate 'q_r' can be computed from equation (3.22). This can be given as:

$$q_r = q_{dc} - V_r \frac{\partial[(1-\bar{\alpha}_v)\rho_w + \rho_s\bar{\alpha}_v]}{\partial P}\frac{dP}{dt} + V_r(\rho_w - \rho_s)\frac{\partial\bar{\alpha}_v}{\partial\alpha_r}\frac{d\alpha_r}{dt} \quad (3.27)$$

The drum dynamics can be captured by the mass balance equation (3.21). The expression for 'q_r', 'q_{sd}' and 'q_{cd}' are substituted in equation (3.21). The resulting simplified expression is given as:

$$\rho_s\frac{dV_{sd}}{dt} + V_{sd}\frac{d\rho_s}{dt} + \frac{1}{h_c}\left[\rho_s V_{sd}\frac{dh_s}{dt} + \rho_w V_{wd}\frac{dh_w}{dt} - (V_{sd}+V_{wd})\frac{dP}{dt} + m_d c_p\frac{dt_s}{dt}\right] +$$

$$\alpha_r(1+\beta)V_r\frac{d[(1-\bar{\alpha}_v)\rho_w+\rho_s\bar{\alpha}_v]}{dt} = \frac{\rho_s}{T_d}\left({V_{sd}}^0 - V_{sd}\right) + \frac{h_f - h_w}{h_c}q_f \quad (3.28)$$

The state equations are written as:

$$\left.\begin{aligned}
&e_{11}\frac{dV_{wt}}{dt} + e_{12}\frac{dP}{dt} = q_f - q_s\\
&e_{21}\frac{dV_{wt}}{dt} + e_{22}\frac{dP}{dt} = Q + q_f h_f - q_s h_s\\
&e_{32}\frac{dP}{dt} + e_{33}\frac{d\alpha_r}{dt} = Q - \alpha_r h_c q_{dc}\\
&e_{42}\frac{dP}{dt} + e_{43}\frac{d\alpha_r}{dt} + e_{44}\frac{dV_{sd}}{dt} =\\
&\frac{\rho_s}{T_d}\left({V_{sd}}^0 - V_{sd}\right) + \frac{(h_f - h_w)}{h_c}q_f
\end{aligned}\right\} \quad (3.29)$$

Where

$$e_{11} = (\rho_w - \rho_s)$$

$$e_{12} = V_{st}\frac{\partial\rho_s}{\partial P} + V_{wt}\frac{\partial\rho_w}{\partial P}$$

$$e_{21} = \rho_w h_w - \rho_s h_s$$

$$e_{22} = V_{st}\left(h_s\frac{\partial\rho_s}{\partial P} + \rho_s\frac{\partial h_s}{\partial P}\right) + V_{wt}\left(h_w\frac{\partial\rho_w}{\partial P} + \rho_w\frac{\partial h_w}{\partial P}\right) - V_t + m_t c_p\frac{\partial t_s}{\partial P}$$

$$e_{32} = \left(\rho_w\frac{\partial h_w}{\partial P} - \alpha_r h_c\frac{\partial\rho_w}{\partial P}\right)(1-\alpha_r)V_r + \left((1-\alpha_r)h_c\frac{\partial\rho_s}{\partial P} + \rho_s\frac{\partial h_s}{\partial P}\right)\bar{\alpha}_v V_r$$
$$+ (\rho_s + (\rho_w - \rho_s)\alpha_r)h_c V_r\frac{\partial\bar{\alpha}_v}{\partial P} - V_r + m_r c_p\frac{\partial t_s}{\partial P}$$

$$e_{33} = ((1-\alpha_r)\rho_s + \alpha_r\rho_w)h_c V_r\frac{\partial\bar{\alpha}_v}{\partial\alpha_r}$$

$$e_{42} = V_{sd}\frac{\partial\rho_s}{\partial P} + \frac{1}{h_c}\left(\rho_s V_{sd}\frac{\partial h_s}{\partial P} + \rho_w V_{wd}\frac{\partial h_w}{\partial P} - V_{sd} - V_{wd} + m_d c_p\frac{\partial t_s}{\partial P}\right)$$
$$+ \alpha_r(1+\beta)V_r\left(\bar{\alpha}_v\frac{\partial\rho_s}{\partial P} + (1-\bar{\alpha}_v)\frac{\partial\rho_w}{\partial P} + (\rho_w - \rho_s)\frac{\partial\bar{\alpha}_v}{\partial P}\right)$$

$$e_{43} = \alpha_r(1+\beta)(\rho_s - \rho_w)V_r\frac{\partial\bar{\alpha}_v}{\partial\alpha_r},$$

$e_{44} = \rho_s$

It is noted that the state space model obtained has an interesting lower triangular structure where state variables can be grouped as: $\big(((V_{wt}, P), \alpha_r), V_{sd}\big)$. The variables inside each parenthesis can be computed independently. Model is thus a nest of second, third and fourth order model. The second order model describes drum pressure and total water volume in the system. The third order model captures the steam dynamics in the risers and the fourth order model also describes the accumulation of steam below the water surface in the drum dynamics.

3.4.2.6.2 *Equilibrium values*

The steady state solution of the state model of equation (3.29) is given by:

$$\left.\begin{aligned} q_f &= q_s \\ Q &= q_s h_s - q_f h_f \\ Q &= q_{dc}\alpha_r h_c \\ V_{sd} &= V_{sd}^0 - \frac{T_d(h_w - h_f)}{\rho_s h_s} q_f \end{aligned}\right\} \quad (3.30)$$

Where, q_{dc} is given by

$$q_{dc} = \sqrt{\frac{2\rho_w A_{dc}(\rho_w - \rho_s) g \bar{\alpha}_v V_r}{K}}$$

A convenient way to find the initial values is to first specify steam flow rate q_s and steam pressure *P*. the feed water flow rate q_f and input power Q are given by first two equations of equation (3.30) and the steam volume in the drum is given by the last equation of (3.30). The steam quality α_r is obtained by solving the nonlinear equations:

$$\left.\begin{aligned} Q &= \alpha_r h_c \sqrt{\frac{2\rho_w A_{dc}(\rho_w - \rho_s) g \bar{\alpha}_v V_r}{K}} \\ \bar{\alpha}_v &= \frac{\rho_w}{\rho_w - \rho_s}\left(1 - \frac{\rho_s}{(\rho_w - \rho_s)\alpha_r} \ln\left(1 + \frac{\rho_w - \rho_s}{\rho_s}\alpha_r\right)\right) \end{aligned}\right\} \quad (3.31)$$

The equilibrium values of the state variables:

1. Drum pressure *P*=8.5 MPa
2. Total water volume of the system V_{wt}=57.5 m^3

3. Steam mass fraction α_r=0.051
4. Volume of steam in the drum V_{sd}=4.8 m^3

The equilibrium values of the input variables are assumed as:

1. Heat flow rate to the risers Q=80.40437506e6 MW
2. Feed water flow rate to the drum q_f=32.00147798 kg/sec
3. Steam flow rate from the drum q_s=32.00147798 kg/sec

3.4.2.6.3 *Parameters*

An interesting feature of the model is that it requires only a few parameters to characterize the system. The parameter values that are considered are those from a Swedish plant. The plant is an 80 MW unit. The model is characterized by the parameters given in Table 3.1.

3.5 BIOREACTOR PROCESS

Bioreactor control has been an active area of research over a decade or so. For optimization of cell mass growth and product formation continuous mode of operation of bioreactors are desirable not the traditional fed batch bioreactors. Several researchers like Edwards et al. (1972), Agrawal and Lim (1984), Menawat & Balachander (1991) have studied the continuous bioreactor problem [14-16]. Kaushiaram et al. (2010) designed neural controllers for various configurations of continuous bioreactor process. [17].

3.5.1 Modeling

A (2×2) bioreactor process was taken up. The primary aim of a continuous bioreactor is to avoid wash out condition which ceases reaction that may be achieved either by controlling cell mass (X g/L) or substrate concentrations (S g/L) at the various operating points of the bio-process. In order to maintain the reaction rate and product quality, both of them may be controlled with dilution rate ($D=F/V$ (h^{-1})) and feed substrate concentration (S_f, g/L) as manipulated variables, thus two degrees of freedom are available for control. The parameters like specific growth rate ($\boldsymbol{\mu}$), yield constant ($\boldsymbol{Y}$) , & saturation rate constant ($\boldsymbol{k_1}, \boldsymbol{k_m}$) of the kinetic models are either inadequately determined or vary from time to time regarding the process operation, hence they are considered as disturbance of the process. The study is based

on single biomass-single substrate process. The following are the model equation based on first principle.

$$\frac{dx_1}{dt} = (\mu - D)x_1 \tag{3.32}$$

$$\frac{dx_2}{dt} = D(x_{2f} - x_2) - \frac{\mu x_1}{Y} \tag{3.33}$$

The reaction rate is given by $r_1 = \mu x_1$ (3.34)

Where x_{2f} is the substrate concentration in the feed. $x_1 \& x_2$ are the biomass and substrate composition, respectively. μ , the specific growth is a function of substrate concentration and given by the substrate inhibition growth rate expression:

$$\mu = \frac{\mu_{max} x_2}{k_m + x_2 + k_1 x_2^2} \tag{3.35}$$

The relation between the rate of generation of cells and consumption of nutrients is defined by the yield given in the following equation

$$Y = \frac{r_1}{r_2} \tag{3.36}$$

Introducing the dilution rate ($D = \frac{F}{V}$) and assuming there is no biomass in the feed, i.e., x_{1f} =0.

The inputs are dilution rate and feed substrate concentration and the outputs are the concentrations of substrate and biomass (All values in deviation form). The values of steady state dilution rate (D_s), feed substrate concentration (x_{2fs}), the steady state values of the states at the stable and unstable operating points and the various parameters are presented in Table 3.2. When both the concentrations (biomass & substrate) are large; process leads to unstable equilibrium. When there is substrate limiting condition, process is at stable equilibrium. Direct synthesis controllers were designed to control the biomass and substrate concentration in both stable and unstable situations.

3.6 CONTINUOUS STIRRED TANK REACTOR WITH COOLING JACKET

A non-isothermal continuous stirred tank reactor with cooling jacket dynamics was considered in order to generate historical database by using distinct operating conditions including stable & unstable operating points and faulty conditions. An exothermic first order reversible reaction $A \rightarrow B$ was used. An inlet fluid stream continuously fed to the reactor and an exit stream continuously removed from the reactor so that the exit stream having same temperature as temperature in the reactor. A cooling jacket surrounded by CSTR having inlet & outlet fluid streams assumed to be perfect mixing and at a temperature less than the reactor temperature. A dynamic model (equations 3.37-3.39) represents mass, component and energy balance with the assumptions of perfect mixing and constant volume. Assuming cooling jacket temperature can be directly manipulated. So no energy balance is required for cooling jacket. Parameters used for simulation of CSTR process are presented in Table 3.3.

$$\frac{dv\rho}{dt} = F_{in}\rho_{in} - F\rho \tag{3.37}$$

$$V\frac{dC_A}{dt} = F_{in}C_{AF} - FC_A - rV \tag{3.38}$$

$$V\rho C_P\frac{dT}{dt} = F\rho C_P(T_f - T) + (-\Delta H)Vr \tag{3.39}$$

Where $r = -k_o \exp(\frac{-E_a}{RT})C_A$ and $-\Delta H$ is heat of reaction.

Direct synthesis controllers were designed to control concentration and temperature using relation between the manipulated inputs and the controlled outputs obtained by converting non-linear model into linear model using state space method.

3.7 TENNESSEE EASTMAN CHALLENGE PROCESS

The Tennessee Eastman Challenge process (TE process) is a benchmark industrial process. Downs & Vogel (1993) [18] first identified the suitability of the T; a non-linear dynamic process as an active research area with little modifications in component kinetics and process operating conditions. The TE process has wide applications like fault detection, development of different control strategies and authentication of performance of supervised controllers. Several researchers like Vinson (1995), Luyben (1996), Lyman & Georgakist

(1994), Zheng (1998), Molina et al. (2010) and Zerkaoui et al. (2010) [19-24]. Ricker (1995; 1996) addressed the TE process to assess the plant-wide control strategies with conventional and unconventional controllers. He has studied the TE process to find optimal steady states which provide six modes of operating points and designed a decentralized control system using proportional integral as well as model predictive controllers to maintain the process at desired values according to product demand [25-26]. The basic process has been adapted from his personnel web page and studied to evaluate the proposed pattern matching techniques.

The TE process comprises an exothermic two phase reactor, condenser, phase separator and stripper column for conversion of reactants namely A, C, D and E to products; G and H. Figure 3.3illustrates complete diagram of TE process. The gaseous reactants are fed to reactor equipped with internal cooling water supply to remove the heat of reaction. A nonvolatile catalyst is used. The products along with unreacted feed are sent to the condenser which condenses all vapors into liquid and from there liquid is fed to separator to get the inert and by product from purge. The uncondensed vapors are recycled to the reactor. The bottom products of the separator are sent to stripper where the products are withdrawn. The heat & material balance data and physical properties of components are presented in Tables 3.4 and 3.5.

The non-linear TE process model was obtained from the modeling equations of all unit operations involved in the process. There are virtually 52 state variables, out of which 41 can be controlled using 12 manipulated variables. The process can be operated in six modes and mode 1 was considered as the base case. The steady state values of process measurements, manipulated variables, and disturbance variables are listed in Tables 3.6-3.9(using base case).

3.8 GENERATION OF DATABASE

For the Drum-boiler process, direct synthesis controller design procedure was used for closed loop control system. Both the open loop and closed loop simulations were performed in order to generate 16 numbers of databases. Four numbers of different operating conditions for Drum-boiler process are listed in Table 3.10. The drum-boiler process was simulated for 1000 seconds with a sampling time of 2 seconds. Four distinct operating conditions were created by giving impulse, step and sinusoidal changes in the manipulated inputs in open loop as well as closed loop.

For Bioreactor process, open loop and closed loop processes were considered in order to generate the database including faults using distinct operating conditions at stable & unstable operating points (by varying the controller tuning parameter, λ, the faulty operating condition was generated). Bioreactor was simulated for one hour and data was taken up with a sampling interval of 6 seconds using different operating conditions. Total database contains 26 numbers of datasets of 4 different operating conditions where each dataset contains 600 observations of two outputs and are presented in Table 3.11.

For the jacketed CSTR process, open loop and closed loop processes were simulated in order to generate the database using distinct operating conditions. Various operating conditions with their parameters are presented in Table 3.12. CSTR was simulated for one hour using different operating conditions. Total database contains 56 numbers of datasets of 3 numbers of operating conditions where each dataset contains 600 observations of two outputs.

For the TE process the distinctive operating modes of process are presented in Table 3.13 In this work, operating mode 1 and mode 3 have been considered in order to generate the historical database.13 numbers of datasets were created pertaining to 13 operating conditions including faulty one which could cause the plant to shut down. The actual values of controlled variables in operating conditions 10 to 13, which leads to plant shut down as specified by Downs & Vogel are bestowed in Table 3.14.Figures 3.4 to 3.7 illustrate the qualitative time responses of few process measurements corresponding to faulty operating conditions obtained from closed loop simulations of the decentralized TE process with proportional controller operated in mode 1 and mode 3. The plant was run for 72 hours with sampling time of 36 seconds for operating conditions FNM1 to FMD2 and FNM3 to FM32. However in the event of faulty situations, the process operation was completely terminated within one hour.

3.8 RESULTS & DISCUSSIONS

3.8.1 Drum Boiler process

Sixteen numbers of datasets were generated by perturbing the process with pseudo random binary signal (PRBS) generator. Signal to noise ratio were maintained as 10.0. Negative step change in feed water flow rate induced continuous decrease in total water volume and continuous increase in drum pressure that represents abnormal scenario. Figure.3.8 shows the response of abnormal process. The combined similarity measures are capable of detecting

faulty operating condition as cluster 3 and datasets pertaining to various other operating conditions were also identified correctly. The performance of clustering is presented in Table 3.15.

The proposed moving window based pattern matching technique perfectly identified all the snapshot data including normal and faulty data which find their similarity with the same process data present in the historical database. The technique was tested in sample wise as well as dataset wise manner. The similarity factors obtained in dataset wise pattern matching are shown in Table 3.16. The overall performance of moving window based pattern matching technique is same for both the cases and shown in Table 3.17. The computation time required in dataset wise manner was less than in comparison to sample wise manner. Pattern matching accuracy and efficiency and efficiency of the algorithm were determined to be 100 %.each

3.8.2 Bioreactor process

Table 3.11presents the 26 numbers of datasets, which were generated for various operating conditions by varying the parameters k_m , k_1 & λ . λ is the tuning parameter of direct synthesis controller, k_m - a parameter (both for Monod and substrate inhibition), k_1 -a parameter for substrate inhibition only. Four numbers of optimum clusters were obtained using similarity based; modified *K*-means algorithm. Faulty operating condition was well captured by cluster 4. The derived cluster purity and efficiency; both were 100 % as presented in Table 3.18.

Pattern matching was done using the moving window in a sample wise as well as dataset wise manner. 26 sets of data pertaining to four various operating conditions were considered as historical database and 4 snapshot data sets were considered. The similarity factors obtained for dataset wise pattern matching are shown in Table 3.19. The overall performance of moving window based pattern matching technique is same for both the cases as shown in Table 3.20. Pool accuracy, Pattern matching efficiency and efficiency of the algorithm were determined to be 100 %.each

3.8.3 Jacketed CSTR Process

By varying C_A, T, and λ, 56 numbers of datasets pertaining to three operating conditions were created. *K*-means clustering algorithm using combined similarity factors repeated for 56 times. Entire database were grouped into 3 clusters. Table 3.21 presents the clustering performance. Resulted clustering efficiency and purity are 100 % each. Figure 3.9presents the response for the CSTR process under faulty operating condition

Pattern matching was done using the moving window in a sample wise as well as dataset wise manner. 3 snapshot data sets were considered. The similarity factors obtained for dataset wise pattern matching are shown in Table 3.22. The overall performance of moving window based pattern matching technique is same for both the cases as shown in Table 3.23. Pool accuracy, Pattern matching efficiency and efficiency of the algorithm were determined to be 100 %.each

3.8.4 Tennessee Eastman Process

The proposed similarity based pattern matching technique was successfully applied to the TE process having 13 numbers of operating conditions. The combined similarity factors used were greater than 0.96. The similarity factors obtained for dataset wise pattern matching are shown in Table 3.24, where S represents snapshot dataset and R represents designed historic dataset. The overall performance of moving window based pattern matching technique is same for both the cases as shown in Table 3.25.

3.9 CONCLUSION

A Sample wise moving window based pattern matching technique was developed with a view to process monitoring. The proposed approach successfully located the arbitrarily chosen different operating conditions of current period of interest among the historical database of a Drum-boiler, continuous bioreactor, Jacketed CSTR and Tennessee Eastman process. The PCA and distance based combined similarity factors provided the effective way of pattern matching in a multivariate time series database of the processes considered. With the other proposed technique, the time series data pertaining to various operating conditions including abnormal ones of the considered processes were discriminated /classified

efficiently using a similarity based modified *K*-means clustering algorithm. Efficient modeling and simulation of the process taken up were the key factors behind the generation of design databases required for the successful implementation of the proposed monitoring techniques. The present work demanded an extensive modeling and simulation activity for the benchmark processes taken up including the prestigious Tennessee Eastman challenge process and industrial scale Drum-boiler process. The developed machine learning algorithms with their encouraging performances deserves immense significance in the current perspective of process monitoring.

Table 3.1 Values of the Drum-boiler model parameters

S. NO	Parameter name	Notation	Value with units
1	Riser volume	V_r	37 m^3
2	Downcomer volume	V_{dc}	11 m^3
3	Total volume	V_t	88 m^3
4	Drum area at normal operating level	A_d	20 m^2
5	Downcomer flow area	A_{dc}	0.355 m^2
6	Riser metal mass	m_r	160e3 kg
7	Total metal mass	m_t	300e3 kg
8	Drum metal mass	m_d	100e3 kg
9	Friction coefficient in downcomer-riser loop	k	25
10	Empirical q_{sd} coefficient	β	0.3
11	Residence time of steam in drum	T_d	12 sec
12	Bubble volume coefficient	V_{sd}^0	10 m^3
13	Acceleration due to gravity	g	9.81 m/sec^2
14	Drum volume	V_d	40 m^3

Table 3.2 Bioreactor process Parameters

Parameters	Value
μ_{max}	0.53 h^{-1}
k_m	0.12 g/L
k_1	0.4545 L/g
Y	0.4
x_{2fs}	4.0 g/L
D_s	0.3 h^{-1}
x_{1s} (at stable operating point)	1.5302 g/L
x_{2s}(at stable operating point)	0.1746 g/L
x_{1s} (at unstable operating point)	0.995103 g/L
x_{2s}(at unstable operating point)	1.512243 g/L

Table 3.3 CSTR parameters used in CSTR simulation

Parameter	Value
$F/V, hr^{-1}$	1
k_o, hr^{-1}	9703*3600
$(-\Delta H), kcal/kgmol$	5960
$E, kcal/kgmol$	11843
$\rho C_P, kcal/(m^3\,^oC)$	500
$T_f, ^oC$	25
$C_{AF}, kgmol/m^3$	10
$UA/V, kcal/(m^3\,^oChr)$	150
$T_j, ^oC$	25

Table 3.4 Heat and material balance data for TE process.

Process stream data						
Stream number	A feed	B feed	E feed	C feed	Strp. Ovhd.	Reactor feed
Stream number	1	2	3	4	5	6
Molar flow (kgmol h^{-1})	11.2	114.2	98.0	417.5	465.7	1890.8
Mass flow (kg h^{-1})	22.4	3664.0	4509.3	6419.4	8979.6	48015.4
Temperature (^{o}c)	45	45	45	45	65.7	86.1
Mole fractions A	0.99990	0	0	0.48500	0.43263	0.32188
B	0.00010	0.00010	0	0.00500	0.0044	0.08893
C	0	0	0	0.51000	0.45264	0.26383
D	0	0.99990	0	0	0.00116	0.06882
E	0	0	0.99990	0	0.07256	0.18776

	F	0	0	0.00010	0	0.00885	0.01657
	G	0	0	0	0	0.01964	0.03561
	H	0	0	0	0	0.00808	0.01659

Stream number		Reactor product	Recycle	Purge	Separation liquid	Product
Stream number		7	8	9	10	11
Molar flow (kgmol h^{-1})		1476.0	1201.5	15.1	259.5	211.3
Mass flow (kg h^{-1})		48015.4	30840	386.5	16788.9	14288.6
Temperature (^{o}c)		120.4	102.9	80.1	80.1	65.7
Mole fractions	A	0.27164	0.32958	0.32958	0	0.00479
	B	0.11393	0.13823	0.13823	0	0.00009
	C	0.19763	0.23978	0.23978	0	0.01008
	D	0.01075	0.01257	0.01257	0.00222	0.00018
	E	0.17722	0.18579	0.18579	0.13704	0.00836
	F	0.02159	0.02263	0.02263	0.01669	0.00099
	G	0.12302	0.04844	0.04844	0.47269	0.53724
	H	0.08423	0.02299	0.02299	0.37136	0.43828

Unit operation data	Reactor	Separator	Condenser	Stripper
Temperature (^{o}c)	120.4	80.1	-	65.7
Pressure (kPa gauge)	2705	2633.7	-	3102.2
Heat duty (kW)	-6468.7	-	-2140.6	1430
Liquid volume (m^3)	16.55	4.88	-	4.43
Utilities				
Reactor cooling water flow (m^3 h^{-1})	93.37			
Condenser cooling water flow (m^3 h^{-1})	49.37			
Stripper stream flow (kg h^{-1})	230.31			

Table 3.5 Component physical properties (at 100^{o}c) in TE process

Component	Molecular weight	Liquid density (kg m^{-3})	Liquid heat capacity (kJ kg^{-1} $^oc^{-1}$)	Vapour heat capacity (kJ kg^{-1} $^oc^{-1}$)	Heat of vaporization (kJ kg^{-1})
A	2	-	-	14.6	-
B	25.4	-	-	2.04	-
C	28	-	-	1.05	-
D	32	299	7.66	1.85	202
E	46	365	4.17	1.87	372
F	48	328	4.45	2.02	372
G	62	612	2.55	0.712	523
H	617	617	2.45	0.628	486

Vapor pressure (Antoine equation):
P= exp(A+B/(T+C))
p=pressure (Pa)
T= temperature (^{o}c)

Component	Constant A	Constant B	Constant C
D	20.81	-1444.0	259
E	21.24	-2114.0	266
F	21.24	-2144.0	266

G	21.32	-2748.0	233
H	22.10	-3318.0	250

#**Vapor pressure parameters are not listed for component A, B and C because they are effectively noncondensible.**

Table 3.6 Process manipulated variables in TE process

Variable name	Variable number	Base case value (%)	Units
D feed flow (stream 2)	XMV (1)	63.053	Kg h^{-1}
E feed flow (stream 3)	XMV (2)	53.980	Kg h^{-1}
A feed flow (stream 1)	XMV (3)	24.644	Kscmh
A and C feed flow (stream 4)	XMV (4)	61.302	Kscmh
Compressor recycle valve	XMV (5)	22.210	%
Purge valve (stream 9)	XMV (6)	40.064	%
Separator pot liquid flow (stream 10)	XMV (7)	38.100	m^3 h^{-1}
Stripper liquid product flow (stream 11)	XMV (8)	46.534	m^3 h^{-1}
Stripper steam valve	XMV (9)	47.446	%
Reactor cooling water flow	XMV (10)	41.106	m^3 h^{-1}
Condenser cooling water flow	XMV (11)	18.114	m^3 h^{-1}
Agitator speed	XMV (12)	50.000	rpm

Table 3.7 Process disturbances in TE process

Variable number	Process variable	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	step
IDV 12)	B composition. A/C ratio constant (stream 4)	step
IDV i3j	D feed temperature (stream 2)	step
IDV (4)	Reactor cooling water inlet temperature	step
IDV (5)	Condenser cooling water inlet temperature	step
IDV (6)	A feed loss (stream 1)	step
IDV (7)	C header pressure loss-reduced availability (stream 4)	step
IDV (8)	A, B, C feed composition (stream 4)	Random variation
IDV (9)	D feed temperature (stream 2)	Random variation
IDV (10)	C feed temperature (stream 4)	Random variation
IDV (I 1)	Reactor cooling water inlet temperature	Random variation
IDV (12)	Condenser cooling water inlet temperature	Random variation
IDV (13)	Reaction kinetics	Slow drift
IDV ii4j	Reactor cooling water valve	Sticking
IDV (15)	Condenser cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown

Table 3.8 Continuous process measurements in TE process

Variable name	Variable number	Base case value	Units
A feed (stream 1)	XMEAS (1)	0.25052	kscmh
D feed (stream 2)	XMEAS (2)	3664.0	kg h^{-1}
E feed (stream 3)	XMEAS (3)	4509.3	kg h^{-1}
A and C feed (stream 4)	XMEAS (4)	9.3477	kscmh
Recycle flow (stream 8)	XMEAS (5)	26.902	kscmh
Reactor feed rate (stream 6)	XMEAS (6)	42.339	kscmh
Reactor pressure	XMEAS (7)	2705.0	kPa gauge
Reactor level	XMEAS (8)	75.000	%
Reactor temperature	XMEAS (9)	120.40	^{0}C
Purge rate (stream 9)	XMEAS (10)	0.33712	kscmh
Product separator temperature	XMEAS (11)	80.109	^{0}C
Product separator level	XMEAS (12)	50.000	%
Product separator pressure	XMEAS (13)	2633.7	kPa gauge
Product separator under flow (stream 10)	XMEAS (14)	25.160	m^3 h^{-1}
Stripper level	XMEAS (15)	50.000	%
Stripper pressure	XMEAS (16)	3102.2	kPa gauge
Stripper under flow (stream 11)	XMEAS (17)	22.949	m^3 h^{-1}
Stripper temperature	XMEAS (18)	65.731	^{0}C
Stripper steam flow	XMEAS (19)	230.31	kg h^{-1}
Compressor work	XMEAS (20)	341.43	kW
Reactor cooling water outlet temperature	XMEAS (21)	94.599	^{0}C
Separator cooling water outlet temperature	XMEAS (22)	77.297	^{0}C

Table 3.9 Sampled process measurements in TE process

Reactor feed analysis (stream 6)				
Component	Variable number	Base case value	Units	Sampling frequency =0.1 h Dead time=0.1h
A	XMEAS (23)	32.188	mol%	
B	XMEAS (24)	8.8933	mol%	
C	XMEAS (25)	26.383	mol%	
D	XMEAS (26)	6.8820	mol%	
E	XMEAS (27)	18.776	mol%	
F	XMEAS (28)	1.6567	mol%	
Purge gas analysis (stream 9)				
Component	Variable number	Base case value	Units	Sampling frequency =0.1 h Dead time=0.1h
A	XMEAS (29)	32.958	mol%	
B	XMEAS (30)	13.823	mol%	
C	XMEAS (31)	23.978	mol%	
D	XMEAS (32)	1.2565	mol%	
E	XMEAS (33)	18.579	mol%	
F	XMEAS (34)	2.2633	mol%	
G	XMEAS (35)	4.8436	mol%	
H	XMEAS (36)	2.2986	mol%	
Product analysis (stream 11)				
Component	Variable number	Base case value	Units	Sampling frequency =0.25 h Dead time=0.25 h
D	XMEAS (37)	0.01787	mol%	
E	XMEAS (38)	0.83570	mol%	
F	XMEAS (39)	0.09858	mol%	

G	XMEAS (40)	53.724	mol%
H	XMEAS (41)	48.828	mol%

[#]The analyzer sample frequency is how often the analyzer takes a sample of the stream. The dead time is the time gap between sample collection and the completion of sample analysis. For an analyzer with a sampling frequency of 0.1 h and a dead time of 0.1 h, a new measurement is available every 0.1 h and the measurement is 0.1 h old.

Table 3.10 Various operating conditions for Drum-boiler process.

Operating condition	Parameter range	No. of datasets
1. Impulse change in q_f and q_s (open loop)	$4 \le q_f \ge 6$ $2 \le q_s \ge 3$	2
2. Simultaneous step changes in three manipulated inputs	$4 \le q_f \ge 10$ $2 \le q_s \ge 5$ $15 \le Q \ge 25$	7
3. Negative change in feed water flow (fault)	$4 \le q_f \ge 6$	4
4. Simultaneous sinusoidal change in three inputs	High, medium and low frequencies	3

Table 3.11 Database corresponding to various operating conditions for Bio-Chemical Reactor process

Op. Cond.	**Parameter range**	**No. of datasets**
1	$0.04545 \le K_1 \le 0.4545$	10
2	$0.015 \le K_m \le 0.12$	8
3	$0.2208 \le \lambda \le 1.1043$	5
4	$2.9446 \le \lambda \le 8.83838$	3

Table 3.12 Generated database for CSTR process at various operating modes

Operating Cond.	Description	Parameter range	No. of datasets
1	Stable operating point ($C_A = 2.359, T = 368.1$)	1. $C_{AF} = 0.5; 295 \le T \le 415$ 2. $C_{AF} = 9.5; 295 \le T \le 415$	24
2	Unstable operating point $C_A = 5.518, T = 339.1$	1. $C_{AF} = 0.5; 295 \le T \le 415$ 2. $C_{AF} = 9.5; 295 \le T \le 415$	22
3	Tuning parameters of controllers	1. $\lambda = \{-5, -10, -15, -20, -25, -30\}$ (Faulty operating cond.) 2. $\lambda = \{5, 10, 15, 20\}$	10

Table 3.13 Operating conditions for TE process

Operating condition ID		Number of datasets
	Mode 1	
FNM1	Normal operation	8
FM11	15% decrease in production set point	
FM12	15% decrease in reactor pressure	
FM13	10% decrease in %G set point	
FMD11	Step change in A/C feed ratio, B composition constant	
FMD12	Step change in B composition, A/C ratio constant	
FMD13	Simultaneous step changes in IDV (1) to IDV (7).	
FMD14	Simultaneous step changes in IDV (1) to IDV (7) and random variation in IDV (8) to IDV (12).	
	Mode 3	
FMN3	Normal operation	5
FM31	Simultaneous 10% decrease in reactor pressure and level	
FM32	Simultaneous 10% decrease in reactor pressure, reactor level, stripper level and %g in product	
FMD31	Simultaneous step changes in IDV (1) to IDV (8).	
FMD32	Simultaneous step changes in IDV (1) to IDV (7) and random variation in IDV (8) to IDV (12).	

Table 3.14 Constraints in TE process

	Normal operating limits		Shut down limits	
Process variable	Low limit	High limit	Low limit	High limit
Reactor pressure	none	2895 kPa	None	3000 kPa
Reactor level	50% ($11.8\ m^3$)	100% ($21.3\ m^3$)	$2.0\ m^3$	$24.0\ m^3$
Reactor temperature	none	150 ^oC	none	175 ^oC
Product separator level	30% ($3.3\ m^3$)	100% ($9.0\ m^3$)	$1.0\ m^3$	$12.0\ m^3$
Stripper base level	30% ($3.5\ m^3$)	100% ($6.6\ m^3$)	$1.0\ m^3$	$8.0\ m^3$

Table 3.15 Combined similarity factor based clustering performance for Drum-boiler process

Cluster no.	N_P	P	Op. cond. 1	Op. cond. 2	Op. cond. 3	Op. cond. 4
1	2	100	1	0	0	0
2	5	100	0	2	0	0
3	4	100	0	0	3	0
4	3	100	0	0	0	4
Avg.		$P = 100$	$\eta = 100$	$\eta = 100$	$\eta = 100$	$\eta = 100$

Table 3.16 Similarity factors in dataset wise moving window based pattern matching implementation for Drum-boiler process

	Op. Cond. 1	**Op. Cond. 2**	**Op. Cond. 3**	**Op. Cond. 4**
Op. Cond. 1	**1.00000000000000**	0.107756743367209	0.368641758177836	0.334190107766254
Op. Cond. 1	0.107756743367209	**1**	0.557074958601167	0.563432838627961
Op. Cond. 1	0.368641758177836	0.557074958601167	**1**	0.303751516080314
Op. Cond. 1	0.004231022400259 77	0.562974025107154	0.303751516079149	**1**

Table 3.17 Moving window based pattern matching performance for Drum-boiler process

Snapshot Op. Cond.	**Np, Total number of data windows similar to snapshot**	**N1, Correctly identified**	**N2, Incorrectly identified**	**P, Pattern accuracy**	**H, Pattern efficiency**	**ξ, Pattern matching algorithm efficiency**
Op. Cond. 1	1	1	0	100	100	100
Op. Cond. 2	1	1	0	100	100	100
Op. Cond. 3	1	1	0	100	100	100
Op. Cond. 4	1	1	0	100	100	100

Table 3.18 Combined similarity factor based clustering performance for Bio-Chemical Reactor process

Cluster No.	**Np**	**P**	**Op. Cond. 1**	**Op. Cond. 2**	**Op. Cond. 3**	**Op. Cond. 4**
1	10	100	10	0	0	0
2	8	100	0	8	0	0
3	5	100	0	0	5	0
4	3	100	0	0	0	3 (faulty cond.)
Avg.		$P = 100$	$\eta = 100$	$\eta = 100$	$\eta = 100$	$\eta = 100$

Table 3.19 Similarity factors in dataset wise moving window based pattern matching implementation for Bioreactor process

Snap.\Ref.	Op. Cond. 1	Op. Cond. 2	Op. Cond. 3	Op. Cond. 4
Op. Cond. 1	**1**	0.589302779704114	0.667232710866021	0.0548711784280415
Op. Cond. 2	0.589302779704114	**1.00000000000000**	0.993496099076534	0.629088229572527
Op. Cond. 3	0.667232710866021	0.993496099076534	**1.00000000000000**	0.550365271093825
Op. Cond. 4	0.054871178428041 5	0.629088229572527	0.550365271093825	**1**

Table 3.20 Moving window based pattern matching performance for Bioreactor process

Snapshot Op. Cond.	Np, Total number of data windows similar to snapshot	N1, Correctly identified	N2, Incorrectly identified	P, Pattern accuracy	H, Pattern efficiency	ξ, Pattern matching algorithm efficiency
Op. Cond. 1	**1**	**1**	**0**	**100**	**100**	**100**
Op. Cond. 2	**1**	**1**	**0**	**100**	**100**	**100**
Op. Cond. 3	**1**	**1**	**0**	**100**	**100**	**100**
Op. Cond. 4	**1**	**1**	**0**	**100**	**100**	**100**

Table 3.21 Combined similarity factor based clustering performance for Jacketed CSTR process.

Cluster no.	Np	P	Op. Cond. 1	Op. Cond. 2	Op. Cond. 3
1	24	100	24		
2	22	100		22	
3	10	100			10
Avg.		$P = 100$	$\eta = 100$	$\eta = 100$	$\eta = 100$

Table 3.22 Similarity factors in dataset wise moving window based pattern matching implementation for Jacketed CSTR process

Snap.\Ref.	**Op. Cond. 1**	**Op. Cond. 2**	**Op. Cond. 3**
Op. Cond. 1	**1**	**0.670000000000000**	**0.644230769230766**
Op. Cond. 2	**0.670000000000000**	**1**	**0.644230769230766**
Op. Cond. 3	**0.644230769230766**	**0.644230769230766**	**1.00000000000000**

Table 3.23 Moving window based pattern matching performance in Jacketed CSTR process

Snapshot Op. Cond.	**Np, Total number of data windows similar to snapshot**	**N1, Correctly identified**	**N2, Incorrectly identified**	**P, Pattern accuracy**	**H, Pattern efficiency**	**ξ, Pattern matching algorithm efficiency**
Op. Cond. 1	**1**	**1**	**0**	**100**	**100**	**100**
Op. Cond. 2	**1**	**1**	**0**	**100**	**100**	**100**
Op. Cond. 3	**1**	**1**	**0**	**100**	**100**	**100**

Table 3.24 Combined similarity factors in dataset wise moving window based pattern matching implementation for TE process

S/R	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**	**S13**
R1	**1.000000000000000**	0.3789922971841109	0.6521527723592 87	0.3346585486276 90	0.5065983672007 29	0.4637603741880 10	0.2243478292321 35	0.3596272714320 08	0.2079677801616 29	0.4741872069798 76	0.4672263011923 05	0.3367831484436 45	0.3360232639708 18
R2	0.3789922971841109	**1.000000000000000**	0.4526104976289 38	0.1017012847464 15	0.4978959693724 41	0.4695319963473 04	0.1582259869169 59	0.01585786628333 76	0.04085499079655 39	0.3452606953594 38	0.3827142012776 19	0.1617618161504 16	0.1624642968158 47
R3	0.6521527723592 87	0.4526104976289 38	**1**	0.3381554626916 87	0.4626830590759 48	0.4233600405815 53	0.2331969401183 20	0.3744488298714 38	0.2105628543468 80	0.3691508555899 41	0.3863849282016 04	0.2696635295402 64	0.2763070778319 96
R4	0.3346585486276 90	0.1017012847464 15	0.3381554626916 87	**1.000000000000000**	0.5054410978425 60	0.5030361082349 47	0.6298884285264 20	0.1139718733640 90	0.6630853118660 82	0.5610219623650 65	0.5597517215176 37	0.6237551207062 09	0.6176272492410 48
R5	0.5065983672 0	0.4978959693 7	0.4626830590 7	0.5054410978 4	**0.99999999999**	0.6602230891 0	0.5045130398 6	0.2628847661 0	0.3624462442 2	0.6409731194 0	0.6492249721 9	0.4094827526 0	0.4033984918 9

	0729	2441	5948	2560	**9999**	3110	0232	3766	9002	9822	5507	2808	3944
R6	0.463760374188010	0.469531996347304	0.423360040581553	0.503036108234947	0.660223089103110	**0.999999999999999**	0.546812401763452	0.274436407368823	0.391627719852375	0.658944693895653	0.664036506108927	0.414238790119853	0.407666186664437
R7	0.224347829232135	0.158225986916959	0.233196940118320	0.629888428526420	0.504513039860232	0.546812401763452	**0.999999999999999**	0.0143448574415571	0.606631569812647	0.642879963320915	0.624088541593125	0.588579568034468	0.575861525314424
R8	0.359627271432008	0.0158578662833376	0.374448829871438	0.113971873364090	0.262884766103766	0.274436407368823	0.0143448574415571	**1.00000000000000**	0.0406606229595791	0.451446306350285	0.386997490407397	0.257431157073034	0.253759857420549
R9	0.207967780161629	0.0408549907965539	0.210562854346880	0.663085311866082	0.362446244229002	0.391627719852375	0.606631569812647	0.0406606229595791	**1**	0.523487731334821	0.495413836490478	0.653167674702088	0.645930246292880
R10	0.474187206979876	0.345260695359438	0.369150855589941	0.561021962365065	0.640973119409822	0.658944693895653	0.642879963320915	0.451446306350285	0.523487731334821	**1.00000000000000**	0.667535101018091	0.448200438796640	0.449227233754385
R11	0.467226301192305	0.382714201277619	0.386384928201604	0.559751721517637	0.649224972195507	0.664036506108927	0.624088541593125	0.386997490407397	0.495413836490478	0.667535101018091	**1**	0.447093065514180	0.446093895213688
R12	0.336783148443645	0.161761816150416	0.269663529540264	0.623755120706209	0.409482752602808	0.414238790119853	0.588579568034468	0.257431157073034	0.653167674702088	0.448200438796640	0.447093065514180	**0.999999999999999**	0.669358486821244
R13	0.336023263970818	0.162464296815847	0.276307077831996	0.617627249241048	0.403398491893944	0.407666186664437	0.575861525314424	0.253759857420549	0.645930246292880	0.449227233754385	0.446093895213688	0.669358486821244	**1.00000000000000**

Table 3.25 Pattern matching performance for TE process

Snapshot Op. Cond.	**Size of the candidate pool ,** N_P	N_1	N_2	**Pattern matching efficiency**	**Pool Accuracy**
FNM1	1	1	0	100	100
FM11	1	1	0	100	100
FM12	1	1	0	100	100
FM13	1	1	0	100	100
FMD11	1	1	0	100	100
FMD12	1	1	0	100	100
FMD13	1	1	0	100	100
FMD14	1	1	0	100	100
FMN3	1	1	0	100	100
FM31	1	1	0	100	100
FM32	1	1	0	100	100
FMD31	1	1	0	100	100
FMD32	1	1	0	100	100

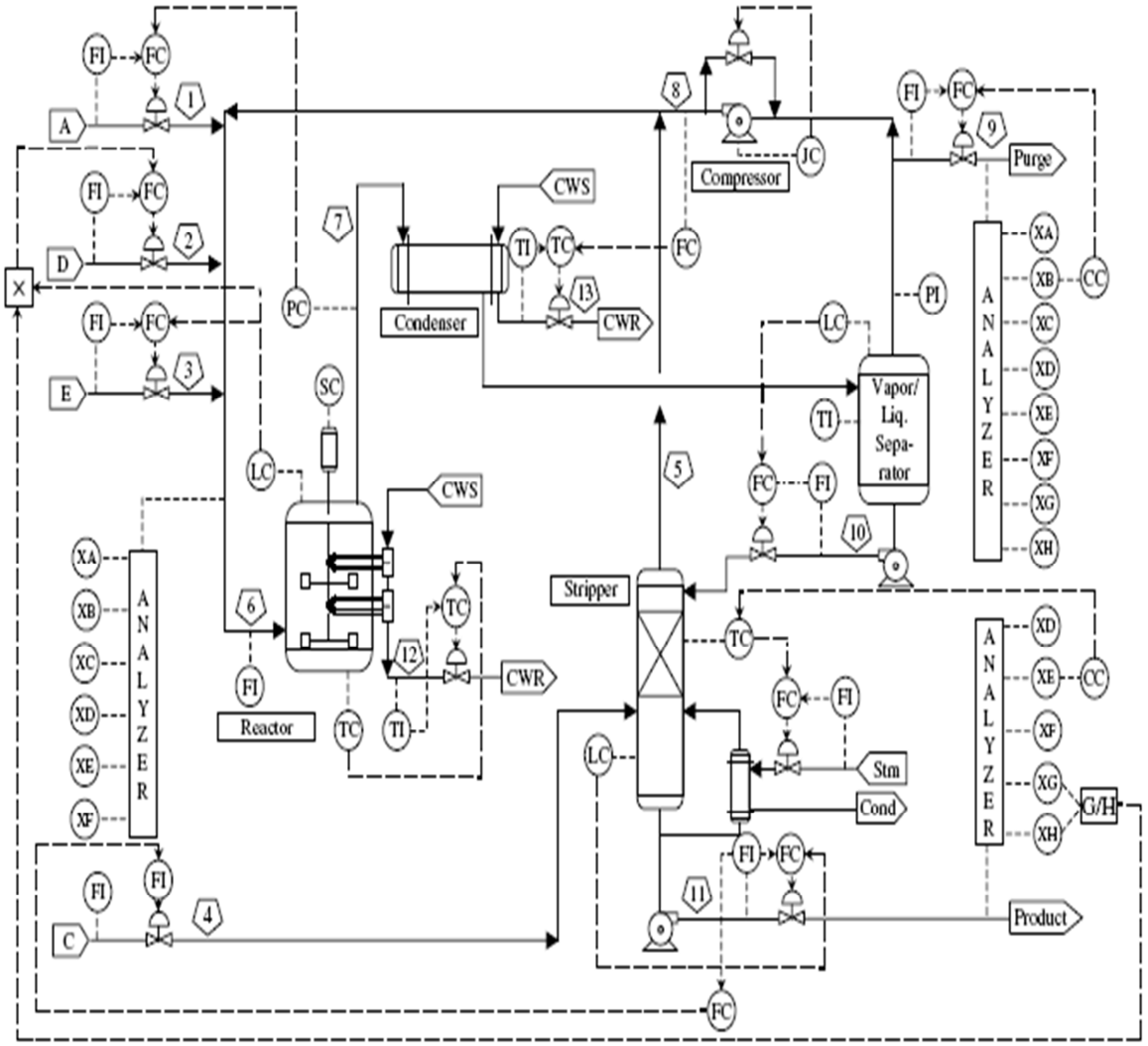


Fig. 3.3 Tennessee Eastman Plant

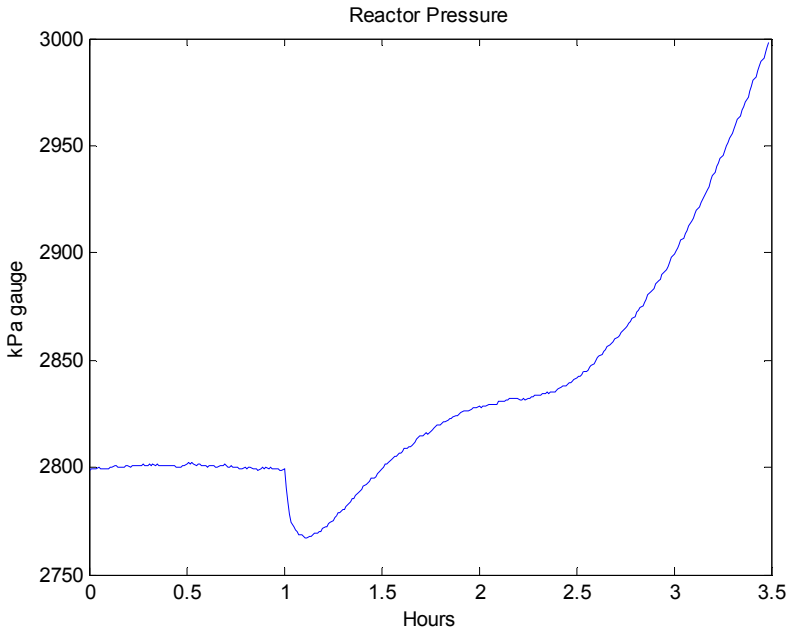


Fig. 3.4 (a) Response of reactor pressure at FMD13

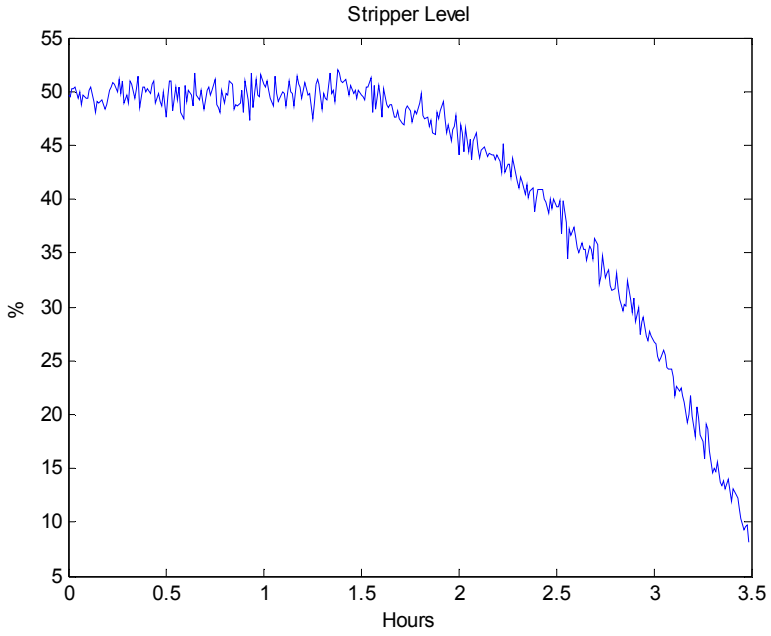


Fig. 3.4 (b) Response of stripper level at FMD13.

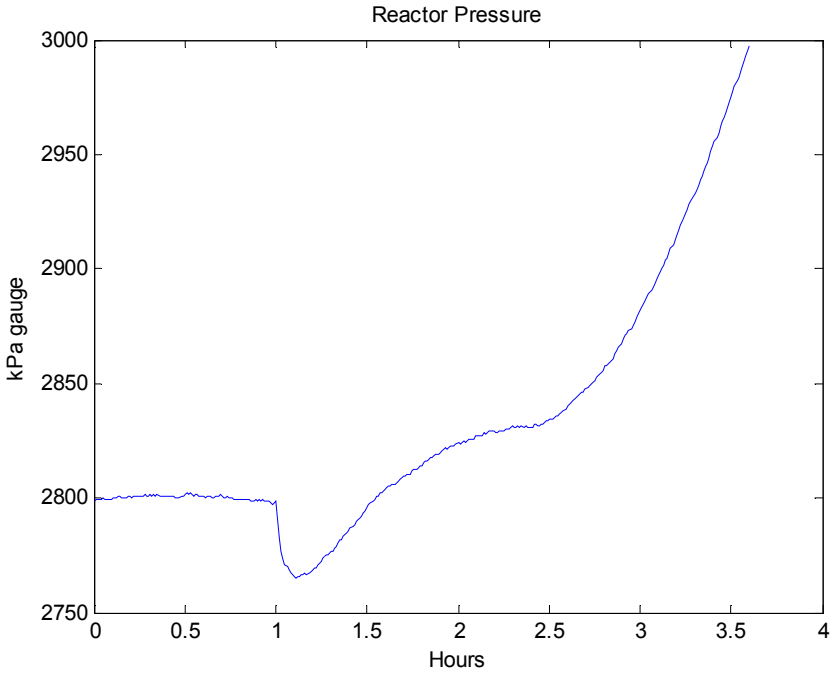


Fig. 3.5 (a) Response of reactor pressure at FMD14.

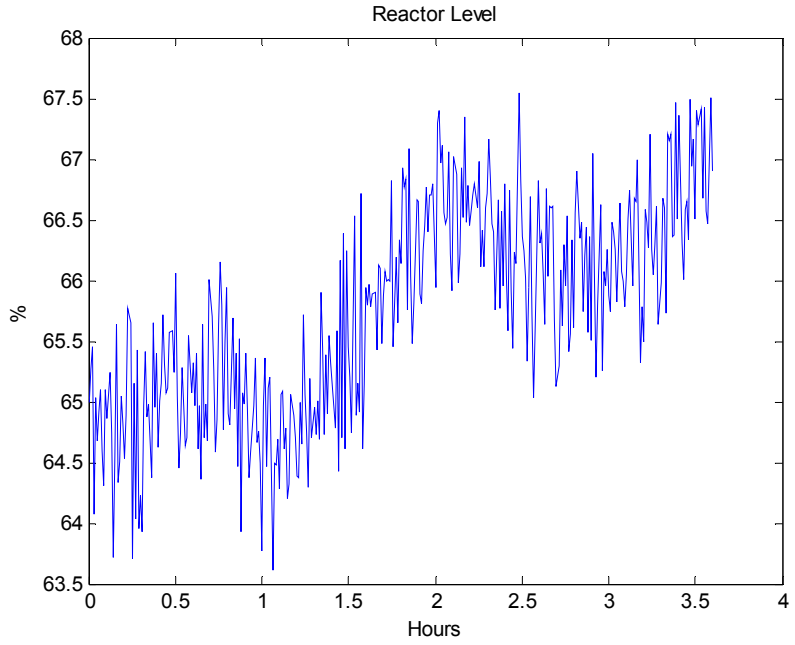


Fig. 3.5 (b) Response of reactor level at FMD14.

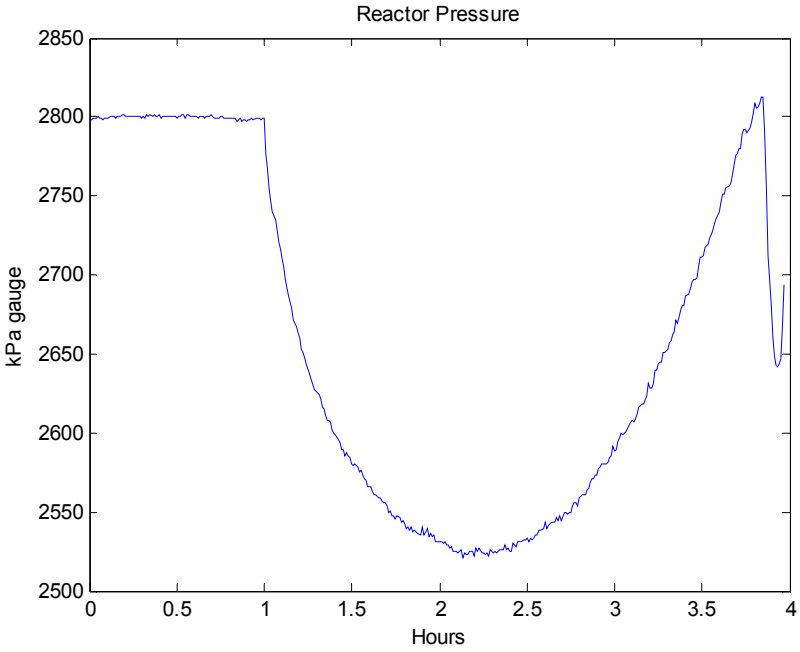


Fig. 3.6 (a) Response of reactor pressure at FMD31.

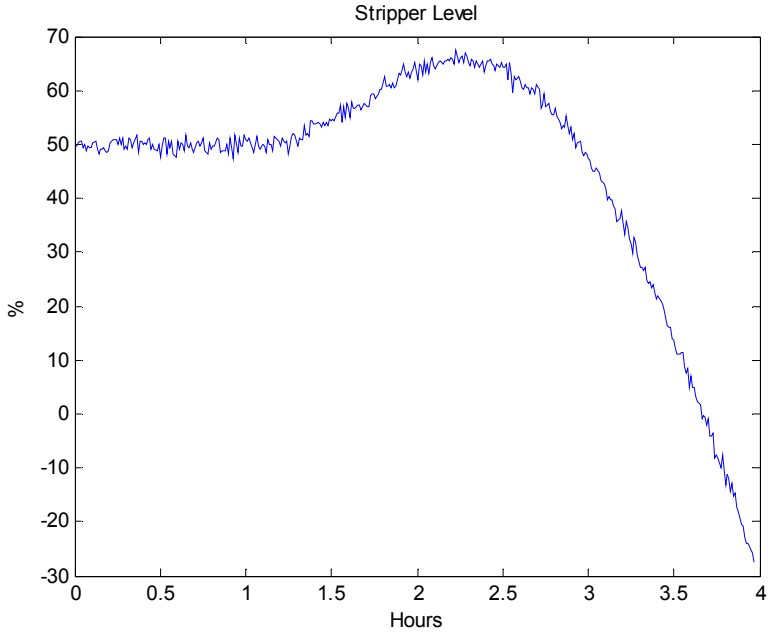


Fig. 3.6 (b) Response of stripper level of FMD31.

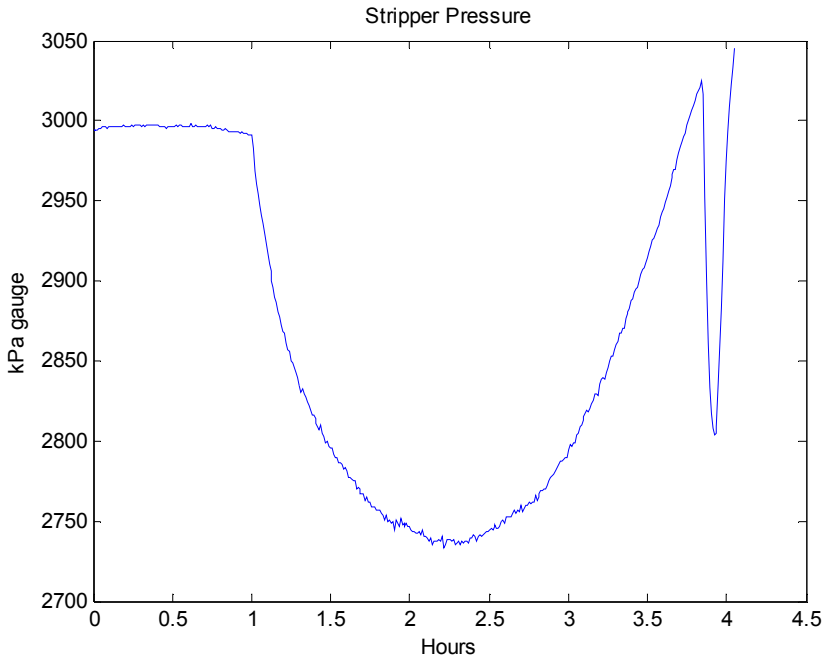


Fig. 3.7 (a) Response of reactor pressure of FMD32.

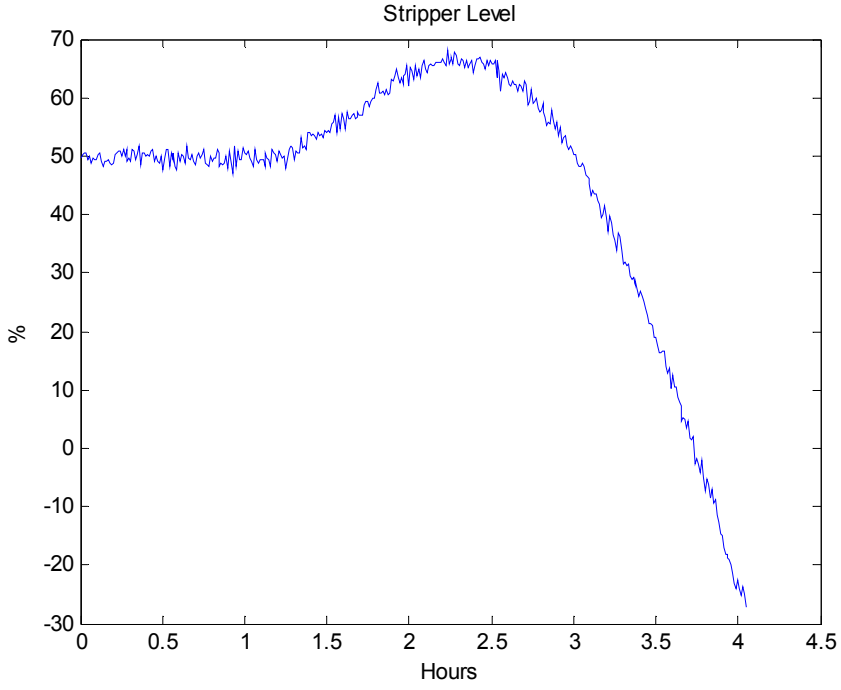


Fig. 3.7 (b) Response of stripper level of FMD32.

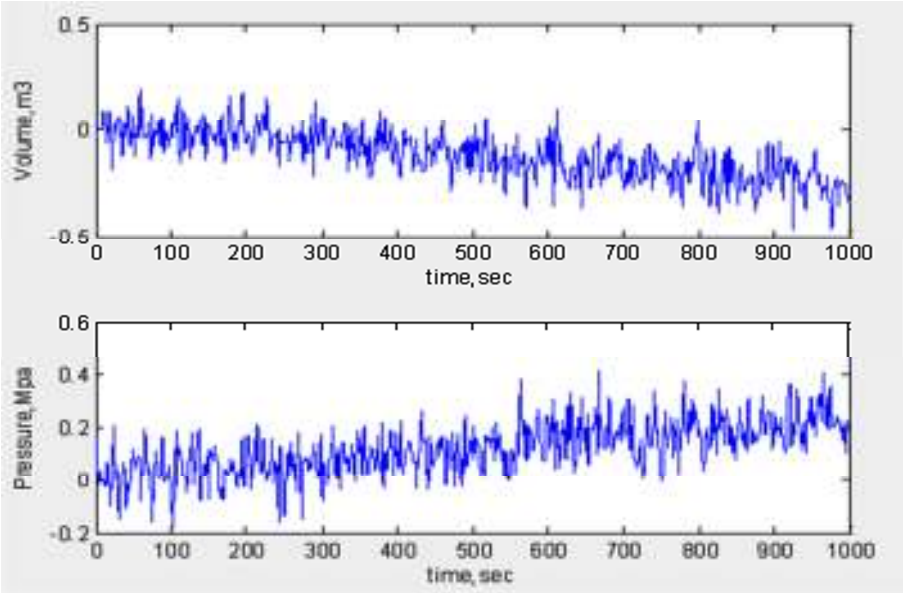


Fig. 3.8 Response of Drum-boiler process for a faulty operating condition

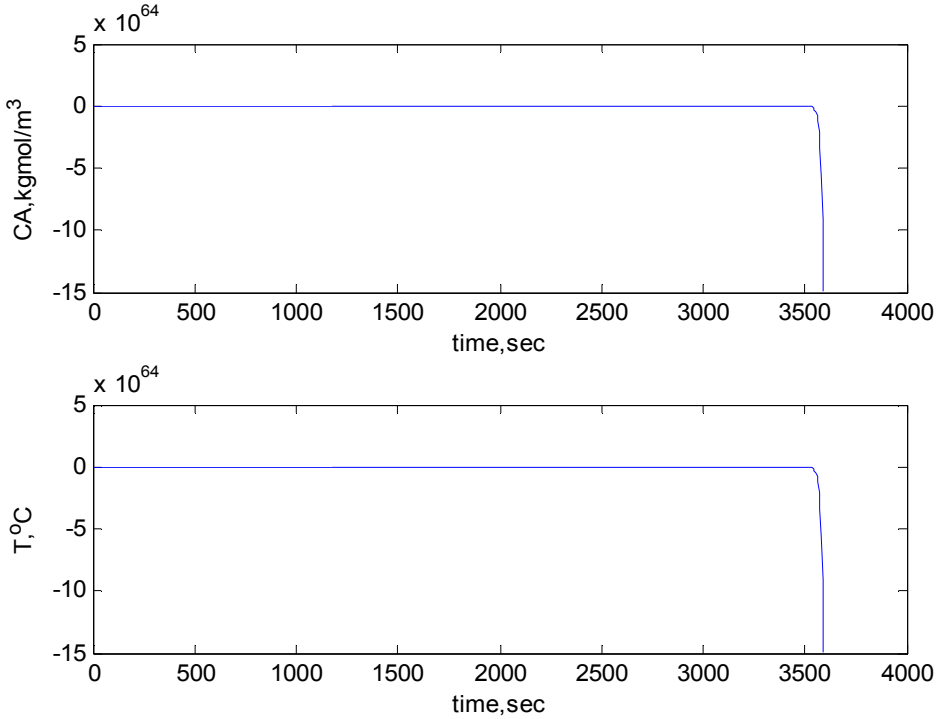


Fig.3. 9 Response for the Jacketed CSTR process under faulty operating condition

REFERENCE

1. Kourti, T., MacGregor, J. F.,"Multivariate SPC methods for process and product monitoring."J. Quality Tech.1996, 28: 409–428.
2. Martin, E. B., Morris, A. J.,"An overview of multivariate statistical process control in continuous and batch performance monitoring."T. I. Meas. Control. 1996, 18: 51–60.
3. Rissanen, J.,"Modeling by shortest data description." Automatica.1978, 14: 465–471.
4. Smyth, P.,"Clustering using Monte Carlo cross-validation." In Proc. 2nd Intl. Conf. Knowl. Discovery & Data Mining (KDD-96), Portland, OR, 1996: 126–133.
5. Singhal, A., Seborg, D. E., "Clustering multivariate time series data." J. Chemometr.2005, 19: 427-438.
6. Krzanowski, W. J.,"Between-groups comparison of principal components." J. Amer. Stat. Assoc.1979, 74: 703–707.
7. Johannesmeyer, M. C., Singhal, A., Seborg, D. E., "Pattern matching in historical data." AIChE J.September 2002, 48:2022-2038.
8. Singhal, A., Seborg, D. E., "Pattern matching in multivariate time series databases using a moving window approach." Ind. Eng. Chem. Res. 2002, 41: 3822-3838.
9. Singhal, A., Seborg, D. E.,"Matching patterns from historical data using PCA and distance similarity factors." Proceedingsof the 2001 American Control Conference; IEEE: Piscataway, NJ., 2001: 1759-1764.
10. Pellegrinetti, G., Bentsman, J., "Nonlinear control oriented boiler modeling-Abenchmark problem for controller design." IEEE T. Contr. Syst. T. January 1996, 4 (1).
11. Astrom, K. J., Bell, R. D., "Drum-boiler dynamics." Automatica. 2000, 36: 363-378.
12. Tan, W., Marquez, H. J., Chen, T., "Multivariable robust controller design for a boiler system." IEEE T. Contr. Syst. T. vol. September 2002, 10 (5).
13. Jawahar, K., Pappa, N., "Selftuning nonlinear controller." Proceedings of the 6th WSEAS Int. Conf. on Fuzzy Systems, Lisbon, Portugal, 2005: 118-126.
14. Edwards, V. H., Ko, R. C., Balogh, S. A., "Dynamics and control of continuous microbial propagators to subject substrate inhibition."Biotechnol. Bioeng.1972,14: 939-974.

15. Agrawal, P., Lim, H. C., "Analysis of various control schemes for continuous bioreactors."Adv. Biochem. Eng. Biot. 1984,30: 61-90.

16. Menawat, A. S., Balachander, J., 1991," Alternate control structures for chemostat."AIChE J. 1991,37 (2): 302-306.

17. Kaushikram, K. S., Damarla, S. K., Kundu, M., "Design of neural controllers for various configurations of continuous bioreactor."International Conference on System Dynamics and Control-ICSDC. 2010.

18. Downs, J. J., Vogel, E. F., "A plant-wide industrial process control problem." Comput. Chem. Engr. 1993, 17 (3): 245-255.

19. Vinson, D. R., "Studies in plant-wide controllability using the Tennessee Eastman Challenge problem, the case for multivariable control." Proceedings of the American Control Conference, Seattle, Washington, 1995.

20. Luyben, W. L., "Simple regulatory control of the Eastman process." Ind. Eng. Chem. Res. 1996, 35: 3280-3289.

21. Lyman, P. R., Georgakis, C., "Plant-wide control of the Tennessee Eastman problem." Comput. Chem. Engr. 1995, 19 (3): 321-331.

22. Zheng, A., "Non-linear model predictive control of the Tennessee Eastman process." Proceedings of the American Control Conference, Philadelphia, Pennsylvania, June 1998.

23. Molina, G. D., Zumoffen, D. A. R., Basualdo, M. S., "Plant-wide control strategy applied to the Tennessee Eastman process at two operating points." Comput. Chem. Engr. 2010.

24. Zerkaoui, S., Druaux, F., Leclercq, E.,Lefebvre, D., "Indirect neural control for plant-wide systems: Application to the Tennessee Eastman Challenge process." Comput. Chem. Engr. 2010, 30: 232-243.

25. Ricker, N. L., "Optimal steady-state operation of the Tennessee Eastman Challenge process." Comput. Chem. Engr. 1995, 19 (9): 949-959.

26. Ricker, N. L., " Decentralized control of the Tennessee Eastman Challenge process." J. Proc. Cont. 1996, 6 (4): 205-221.

Chapter 4

APPLICATION OF MULTIVARIATE STATISTICAL AND NEURAL CONTROL STRATEGIES

Chapter 4

APPLICATION OF MULTIVARIATE STATISTICAL AND NEURAL PROCESS CONTROL STRATEGIES

4.1 INTRODUCTION

Multiloop (decentralized) conventional control systems (especially PID controllers) are often used to control interacting multiple input, multiple output processes because of their ease in understandability and requirement of fewer parameters than more general multivariable controllers. Different types of tuning methods like BLT detuning method, sequential loop tuning method, independent loop tuning method, and relay auto-tuning method have been proposed by the researchers for tuning of multiloop conventional controllers. One of the earlier approaches of multivariable control had been the decoupling of process to reduce the loop interactions. For multivariable processes, the Partial least squares based (PLS) controllers offer the opportunity to be designed as a series of SISO controllers(Qin and McAvoy (1992, 1993)[1-2]. The PLS based process identification &control have been chronicled by researchers like Kaspar and Ray (1993), Lakshminarayanan et al., (1997) [3-4].Till date, there is no literature reference for NNPLS controller except the contribution by Damarla and Kundu (2011) [5]. In view of this, for controlling non-linear complex processes neural network based PLS (NNPLS) controllers have been proposed. (2×2), (3×3), and (4×4)Distillation processes were taken up to

implement the proposed control strategy followed by the evaluation of their closed loop performances. Control strategies using classical and neural controllers have been incorporated in a *plant wide* multistage Juice evaporation process plant.

4.2 DEVELOPMENT OF PLS AND NNPLS CONTROLLERS

4.2.1 PLS Controller

Discrete input-output time series data (X-Y) either collected from plant or generated by perturbing the benchmark processes with pseudo random binary signals can be used for the development of PLS controllers. For modeling dynamic process, the input data matrix (X) was augmented either with lagged input variables (called finite impulse response (FIR) model) or including lagged input and output variables (called auto regressive model with exogenous input, ARX). By combining the PLS with inner ARX model structure, dynamic processes can be modeled, which logically builds up the framework for PLS based process controllers.

Identification of process dynamics in latent subspace has already been discussed in Chapter 2 (section 2.6), which is supposed to be the key component for the successful development of a PLS based controller. The very fortes about the PLS identification of a multidimensional time series data is as follows:

- While deriving the outer relation among the scores ($(T \,\&\, U)$of multidimensional input-output data, the regression occurs at each of its individual dimension. The residuals are passed to the successive steps and then the whole procedure is repeated assuming the residuals ($E \,\&\, F$) from the previous steps as the input-output data. in the current step, Hence, multidimensional regression problems gets transformed in to a series of one-dimensional regression problems.
- It selects the most realistic part of the predictor variable which is outmost predictive of response variable at each dimension
- The predictor and the response variables are pre and post compensated by loading matrices $P \,\&\, Q.$, respectively

In view of this, the PLS identified process dynamics deserves to be more precise and realistic than other time series data based models like ARX, ARMA. ARMAX, etc.

Series of direct synthesis SISO controllers designed on the basis of the dynamic models identified into latent subspaces and embedded in the PLS frame work were used to control the process. This approach is having the advantage that instead of using a typical multivariable controller, independent SISO controllers can be designed based on the inner dynamic model $G_p(z)$ identified at each dimension along with the error and control signals being pre and post compensated by the loading matrices namely P and Q, respectively. The PLS-based control strategy is presented in Figure 4.1.

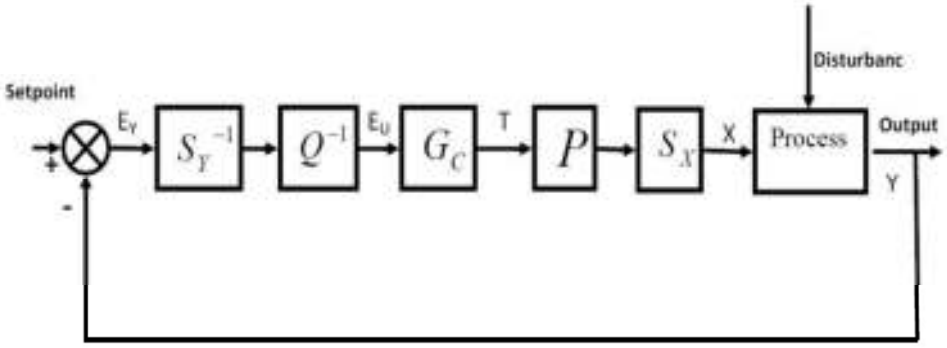


Fig. 4.1 Schematic of PLS based Control

$G(s) = G_p(z)$ represents the identified process transfer function in the latent subspaces. The controller acts on projected error E_u i.e., actually measured output error post compensated by matrices ${S_y}^{-1}\&Q^{-1}$, respectively. T, the score is computed actually by the controller in closed loop PLS framework. The T score then gets projected on loading matrix P and transformed in to real physical inputs through $S_x$which drive the processes.

4.2.2 Development of Neural Network PLS Controller (NNPLS)

In PLS based process dynamics, the inner relationship between (X) and (Y) scores; hence the process dynamics in latent subspace could not be well identified by linear or quadratic relationships; especially for complex and non-linear processes. In view of this, present work proposed neural identification of process dynamics using the latent variables. MIMO processes without any decouplers could be controlled by a series of NN based SISO controllers with the PLS loading matrices being employed as pre and post compensators of the error and control signals, respectively. The inverse dynamics of the latent variable based process was identified as the direct inverse neural network controller (DINN) using the

historical database of latent variables. In the regulator mode, the latent variable based disturbance process was identified using NN.

Several fortes about the NNPLS controllers are as follows:

- A multivariable controller can be used as a series of SISO NN controllers within a PLS framework.
- Because of the diagonal structure of the dynamic part of the PLS model, input-output pairings are automatic ($y_1 - u_1, y_2 - u_2, y_3 - u_3$….etc.).
- The set point or the desirable state passes to the controller as a projected variable down to the latent subspace. It might have been easier to deal for the controllers in tracking the set point because the infeasible part of the set point is not allowed to pass directly to the controllers.
- The PLS identified process is somewhat decoupled owing to the orthogonality of the input scores and the rotation of the input scores to be highly correlated with the output scores.
- The use of neural network in identifying the inner relation among the input and output scores might capture the process non-linearity.
- For every input-output pairs participated in training the neural networks, the use of historical database eliminates the necessity of determining cross correlation coefficient. The cross correlation coefficient would have guided the selection of most effective inputs and their corresponding targets to be used in training the neural networks.

The process transfer functions were simulated over a stipulated period of time to generate output-input data using signal to noise ratio as 10.0. The scores corresponding to all the time series data were generated using the principal component decomposition. Instead of the process being identified by linear ARX based model coupled with least squares regression, in the NNPLS design; the relationship between the $T\&U$ scores are estimated by feed forward back propagation neural network. The network input was arranged in ARX structure using the process historical database. In the closed loop simulation of the process, the inverse NN models were used as a controller typically in a feed forward fashion with the error, if any; adjusted as a bias. The inputs ($N1$) and outputs ($N3$) to the multilayer (3 layers) feed forward neural network (FFNN) representing forward dynamics of the process regarding its training & simulation phase were as follows:

Training phase

$$N1 = \{U(t-3), U(t-2), T(t-3), T(t-2)\} \quad (4.1)$$

$$N3 = U(t-1) \quad (4.2)$$

Simulation Phase

$$N1 = \{U(t-2), U(t-1),\ T(t-1), T(t)\} \quad (4.3)$$

$$N3 = U(t) \quad (4.4)$$

For DINN the simulation phase is synonymous to control phase. The number of hidden layer neurons varies from process to process. Figure 4.2 represents the NNPLS scheme; both in servo and regulator mode. In regulator mode, the effective disturbance transfer function was simulated to produce the Y data and D data; hence their corresponding scores. The FFNN representing disturbance dynamics acts as effective disturbance process. The disturbance rejection was done along with the existing servo mode.

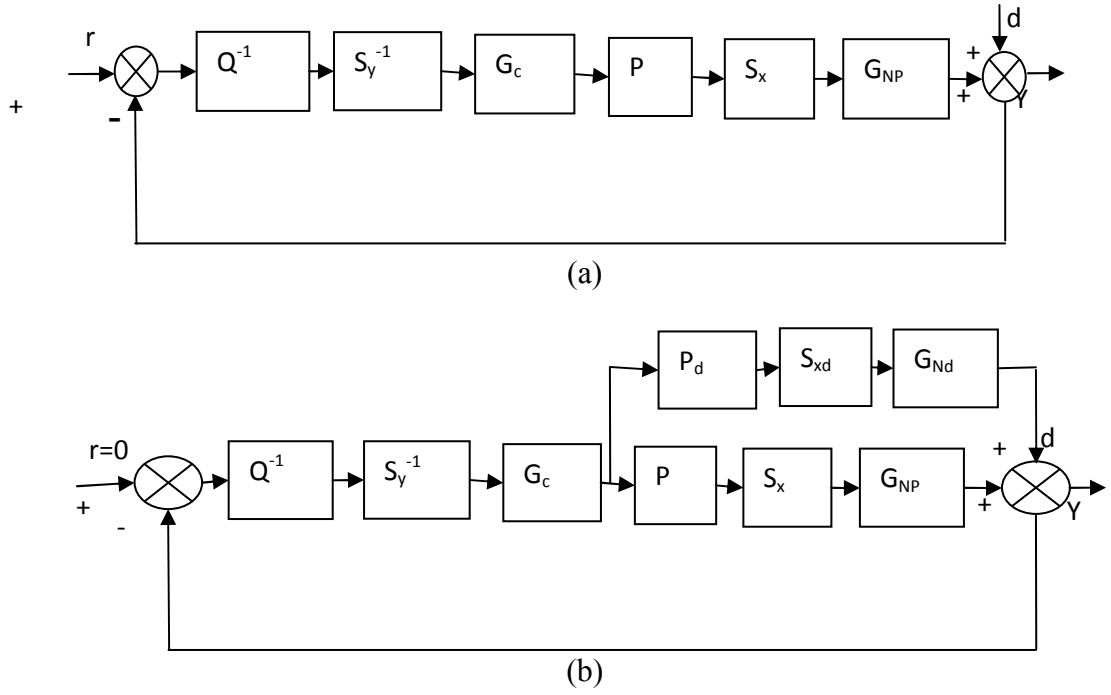


Fig. 4.2 NNPLS a) Servo & b) Regulatory mode of control

The inputs ($N1$) and outputs ($N3$) of the multilayer (3 layers) disturbance FFNN regarding the training & simulation phase were as follows,

Training phase

$$N1 = \{U(t-3), U(t-2), d(t-3), d(t-2)\} \quad (4.5)$$

$$N3 = U(t-1) \quad (4.6)$$

Simulation Phase

$$N1 = \{U(t-2), U(t-1), d(t-1), d(t)\} \quad (4.7)$$

$$N3 = U(t) \quad (4.8)$$

The number of hidden layer neurons varies for various disturbance processes. In all the multivariable processes considered, n_i , $(i = 1..4, no\ of\ variables\ to\ be\ controlled)$ numbers of input-output time series relations were identified using neural networks in a latent variable sub-space. Inverse neural network acted as controller. To control any of the multivariable processes, n_i , $(i = 1..4, no\ of\ variables\ to\ be\ controlled)$ numbers of SISO controllers were designed. During the training phase of the NNs', the input scores to the network were arranged as per the ARX structure following eq.(4.1) and output or target score was set as per eq.(4.2). In simulation mode, the inputs and outputs to the trained networks were arranged as per equations. (4.3) & (4.4). Neural networks were also used to mimic the disturbance dynamics for all the cases considered. During the training phase, the input scores to the network were arranged as per the ARX structure following eq. (4.5) and output or target score was set as per eq.
equation (4.6). In simulation mode, the inputs and outputs to the trained networks were as per equation. (4.7) & (4.8). The training algorithm used was gradient based. The convergence criterion was MSE or mean squared error. The performances of the networks representing various processes are presented in Table 4.1.

4.3 IDENTIFICATION & CONTROL OF (2×2) DISTILLATION PROCESS

4.3.1 PLS Controller

A 2×2 distillation process was chosen to identify the process dynamics in latent subspaces and compare the PLS predicted dynamics with the actual one. Top product composition X_D and bottom product composition X_B were controlled by reflux rate and

vapor boil-up using PLS based direct synthesis controllers as well as neural controllers. The following transfer function equation (4.9) was used to simulate the process when perturbed by pseudo random binary signals (1000 samples).

Inputs:

u_1=Reflux flow rate; u_2 = vapour boil up

Disturbances:

d_1 = Feed flow rate; d_2 = Feed light component mole fraction

$$\begin{bmatrix} \frac{0.878(1-0.6s)}{(1+0.6s)(75s+1)} & \frac{-0.864(1-0.6s)}{(1+0.6s)(75s+1)} \\ \frac{1.082(1-0.1s)}{(1+0.1s)(75s+1)} & \frac{-1.096(1-0.1s)}{(1+0.1s)(75s+1)} \end{bmatrix} \quad (4.9)$$

Both FIR based and ARX based inner models were used to identify the process dynamics in projected subspaces. Equations (4.10-4.11) represent the identified ARX based dynamic models for outputs 1& 2, respectively. Equations (4.12-4.13) represent the identified FIR based dynamic models for outputs 1& 2, respectively.

$$G_1(z) = \frac{-0.001791z-0.004398}{z^2+1.547z-0.5563} \quad (4.10)$$

$$G_2(z) = \frac{-0.1367z-0.3776}{z^2+1.045z-0.05656} \quad (4.11)$$

$$G_1(z) = \frac{0.01951z+0.02326}{z^2+0.0152z+0.0151} \quad (4.12)$$

$$G_2(z) = \frac{0.5637z+0.5447}{z^2+0.2816z+0.5448} \quad (4.13)$$

FIR and ARX based inner correlation between the scores T and U were established keeping the outer linear structure of the PLS intact. The predicted outputs corresponding to the inputs within a PLS framework were obtained by post compensating the U scores with Q matrix. The T score were generated by post compensating the original X matrix with P matrix. Figures 4.3 and 4.4 present the comparison of actual plant dynamics involving top product composition X_D and bottom product composition X_B with ARX based PLS predicted dynamics.

The desired transfer function for closed loop simulation was selected as second order system. The controller were designed as direct synthesis controller in the following way,

$$G_{CL}(s) = \begin{bmatrix} \frac{1}{(\lambda s+1)^2} & 0 \\ 0 & \frac{1}{(\lambda s+1)^2} \end{bmatrix} \tag{4.14}$$

Where λ is tuning parameter. Direct Synthesis controller resulted is as follows:

$$G_C(s) = \frac{G_{CL}(s)}{G(s)(1-G_{CL}(s))} \tag{4.15}$$

The performances of proposed direct synthesis controllers designed on the basis of equations (4.9-4.13& 4.15) and embedded in PLS framework were examined. PLS controller perfectly could track the set point (step change in top product composition from 0.99 to 0.996 and step change in bottom product composition from 0.01 to 0.005). Figures 4.5 and 4.6 illustrate and compare the performance of PLS controllers (FIR based/ARX based inner dynamic model) in servo mode.

4.3.2 NNPLS Controller

In the same (2×2) distillation process output 1 ($Topproductcomposition,\ x_D$) – input1 (reflux flow rate) & the output 2 ($Bottomproductcomposition,\ x_B$) – input 2 (vapour boil-up) time series relations were identified using neural networks in a latent variable sub-space. Inverse neural network acted as controller. To control the process, two numbers of SISO controllers were designed. Neural networks were also used to mimic the disturbance dynamics for output 1 ($Topproductcomposition,\ x_D$) – input disturbance 1 (d_1, feed flow rate) & the output 2 ($Bottomproductcomposition,\ x_B$) – input disturbance 2 (d_2, feed light component mole fraction). All the designed networks were three layered networks. The design procedure has already been discussed in the subsection 4.2.2.

Figure 4.7 presents the comparison between actual process outputs and NN identified process outputs namely the top and bottom product compositions. Figure 4.8 shows the closed loop performance of 2 numbers of SISO controllers. In servo mode, the controller proved to be a very reliable in set point tracking and reached the steady state value in less than 15 s, when the top product composition changes from 0.99 to 0.996. The 2 nd controller could track the set point changing in bottom product composition from 0.01 to 0.005 by

reaching the steady state value within 10 s. Figure 4.9 presents the closed loop simulation in regulatory mode in conjunction with the existing servo; showing the disturbance rejection performance in $x_D \& x_B$.

4.4 IDENTIFICATION & CONTROL OF (3×3) DISTILLATION PROCESS

The following equations (4.16) & (4.17) are the LTI process and disturbance transfer functions of a (3×3) distillation process, respectively which were identified by Ogunnaike et al. (1983). NNPLS controllers were proposed to control the process.

$$Gp(s) = \begin{bmatrix} \frac{0.66\exp(-2.6s)}{6.7s+1} & \frac{-0.61\exp(-3.5s)}{8.64s+1} & \frac{-0.0049\exp(-s)}{9.06s+1} \\ \frac{1.11\exp(-6.5s)}{3.25s+1} & \frac{-2.36\exp(-3s)}{5s+1} & \frac{-0.012\exp(-1.2s)}{7.09s+1} \\ \frac{-34.68\exp(-9.2s)}{8.15s+1} & \frac{46.2\exp(-9.4s)}{10.9s+1} & \frac{0.87(11.61s+1)\exp(-s)}{(3.89s+1)(18.8s+1)} \end{bmatrix} \quad (4.16)$$

$$G_D(s) = \begin{bmatrix} \frac{0.14\exp(-12s)}{6.2s+1} & \frac{-0.0011(26.32s+1)\exp(-2.66s)}{(7.85s+1)(14.63s+1)} \\ \frac{0.53\exp(-10.5s)}{6.9s+1} & \frac{-0.0032(19.62s+1)\exp(-3.44s)}{(7.29s+1)(8.94s+1)} \\ \frac{-11.54\exp(-0.6s)}{7.01s+1} & \frac{0.32\exp(-2.6s)}{7.765s+1} \end{bmatrix} \quad (4.17)$$

The following are the outputs, inputs and disturbances of the system Outputs:

y_1 = Overhead ethanol mole fraction,

y_2 = Side stream ethanol mole fraction

y_3 =19 th tray temperature.

Inputs:

u_1 =Reflux flow rate

u_2 =Side stream product flow rate,

u_3 = Reboiler steam pressure

Disturbances:

d_1 = Feed flow rate

d_2 = Feed temperature

The output 1 (Overhead ethanol mole fraction, y_1) – input1 (Reflux flow rate, u_1), output 2 (Side stream ethanol mole fraction , y_2) – input 2 (Side stream product flow rate, u_2), &

output 3 (19 th tray temperature, y_3) – input 3 (Reboiler steam pressure, u_3) time series pairings are by default in PLS structure and it was also supported by RGA (relative gain array) analysis. Because of the orthogonal structure of PLS, the aforesaid pairings are consequential. Three numbers of SISO NNPLS controllers were designed to control top, side product compositions and 19 th tray temperature of the (3×3)distillation process. The disturbance dynamics of the distillation process were identified in the form of three numbers of feed forward back propagation neural networks. The output 1 (Overhead ethanol mole fraction, y_1)– input disturbance 1 (d_1, feed flow rate), output 2 (Side stream ethanol mole fraction , y_2) – input disturbance 1 (d_1, feed flow rate) & output 3 (Side stream ethanol mole fraction , y_3) – input disturbance 1 (d_1, feed flow rate) time series relations were considered to manifest the disturbance rejection performance of the designed controllers.The design of NNPLS controllers was done as per the procedure discussed in subsection 4.2.2.

Figure 4.10 presents the comparison between the NNPLS predicted dynamics and the simulated plant output. Output 3 was comparatively poorly fitted than in comparison to the other 2 outputs as evidenced by the regression coefficient of the concerned network model. Figure 4.11 presents the NNPLS controller based closed loop responses of the process for all of its outputs in servo mode. The figure shows the response of the system when there is a set-point change of -0.05 iny_1, -0.08 in y_2 and no change in y_3. A perfectly decoupled set point tracking performance of the NNPLS control system design is also revealed. Figure 4.12shows the closed loop responses in regulator mode as a result of step change in disturbance 1 by + 0.2. The designed NNPLS controllers emancipated encouraging performance in servo as well as in regulator mode.

4.5 IDENTIFICATION & CONTROL OF(4×4)DISTILLATION PROCESS

The following transfer functions equations. (4.18) & (4.19) are the process and disturbance transfer functions of a (4×4) distillation process model adapted from Luyben (1990). NNPLS controllers were proposed to control the process.

$$Gp(s)=\begin{bmatrix} \frac{4.09\exp(-1.3s)}{(33s+1)(8.3s+1)} & \frac{-6.36\exp(-1.2s)}{(31.6s+1)(20s+1)} & \frac{-0.25\exp(-1.4s)}{21s+1} & \frac{-0.49\exp(-6s)}{(22s+1)^2} \\ \frac{-4.17\exp(-5s)}{45s+1} & \frac{6.93\exp(-1.02s)}{44.6s+1} & \frac{-0.05\exp(-6s)}{(34.5s+1)^2} & \frac{1.53\exp(-3.8s)}{48s+1} \\ \frac{1.73\exp(-18s)}{(13s+1)^2} & \frac{5.11\exp(-12s)}{(13.3s+1)^2} & \frac{4.61\exp(-1.01s)}{18.5s+1} & \frac{-5.49\exp(-1.5s)}{15s+1} \\ \frac{-11.2\exp(-2.6s)}{(43s+1)(6.5s+1)} & \frac{14(10s+1)\exp(-0.02s)}{(45s+1)(17.4s^2+3s+1)} & \frac{0.1\exp(-0.05s)}{(31.6s+1)(5s+1)} & \frac{4.49\exp(-0.6s)}{(48s+1)(6.3s+1)} \end{bmatrix}$$

(4.18)

$$G_D(s)=\begin{bmatrix} \frac{4.09\exp(-1.3s)}{(33s+1)(8.3s+1)} & \frac{-6.36\exp(-1.2s)}{(31.6s+1)(20s+1)} \\ \frac{-4.17\exp(-5s)}{45s+1} & \frac{6.93\exp(-1.02s)}{44.6s+1} \\ \frac{1.73\exp(-18s)}{(13s+1)^2} & \frac{5.11\exp(-12s)}{(13.3s+1)^2} \\ \frac{-11.2\exp(-2.6s)}{(43s+1)(6.5s+1)} & \frac{14(10s+1)\exp(-0.02s)}{(45s+1)(17.4s^2+3s+1)} \end{bmatrix}$$

(4.19)

The following are the outputs, inputs and disturbances of the system

Outputs:

y_1 = Top product composition (x_{D1})

y_2 = Side stream composition (x_{S2})

y_3 = Bottom product composition x_{B3})

y_4 =Temperature difference to minimize the energy consumption $\Delta T4$)

Inputs:

u_1 =Reflux rate

u_2 = Heat input to the reboiler

u_3 = Heat input to the stripper

u_4 = Feed flow rate to the stripper

Disturbances:

d_1 = Feed flow rate

d_2 = Feed temperature

The dynamics of the distillation process were identified in the form of four numbers of feed forward back propagation neural networks. The output 1 (Top product composition , y_1) – input1 (Reflux flow rate, u_1), the output 2 (Side stream composition , y_2) – input 2 (Heat input to the reboiler, u_2), the output 3 (Bottom product composition , y_3) – input 3 (Heat input to the stripper, u_3) & output 4 (Temperature difference , y_4) – input 4 (Feed flow rate to the stripper, u_4) were the four numbers of SISO processes considered for analysis. The disturbance dynamics of the distillation process were identified in the form of four numbers of feed forward back propagation neural networks. The output 1 (Overhead ethanol mole fraction, y_1)– input disturbance 1 (d_1, feed flow rate), output 2 (Side stream ethanol mole fraction , y_2) – input disturbance 1 (d_1, feed flow rate), output 3 (Side stream ethanol mole fraction , y_3) – input disturbance 1 (d_1, feed flow rate), and output 4 (Side stream ethanol mole fraction , y_4) – input disturbance 1 (d_1, feed flow rate) time series relations were considered. Four numbers of NNPLS controllers were designed as DINN of the process as per the procedure discussed in the subsection 4.2.2.

Figure 4.13 presents the comparison between the NNPLS model and the simulated plant output for the (4×4) Distillation process.Figure 4.14 presents the closed loop responses for all the process outputs in servo mode of the designed NNPLS controller. This figure shows the response of the system when there is a step change of +0.85 iny_1, -0.1 in y_2 , -0.05 in y_3 and no change in y_4 , and it is a perfect set point tracking performance by the NNPLS control system designed in servo mode. Figure 4.15 shows the closed loop responses for all the process outputs in regulator mode of the designed NNPLS controller and as a result of step change in disturbance 1 by + 0.2. The designed NNPLS controllers emancipated encouraging performance in servo as well as in regulator mode.

5 PLANT WIDE CONTROL

One of the important tasks for the sugar producing process is to ensure optimum working regime for the multiple effect evaporator; hence perfect control system design. A number of feedback and feed forward strategies based on linear and non-linear models were investigated in this regard. Generic model based control (GMC), generalized predictive control (GPC), linear quadratic Gaussian (LQG), and internal model control (IMC) were applied in the past for controlling multiple effect evaporation process (Anderson & Moore,1989; Mohtadi et. al., 1987) [6-7].

The multiple effect evaporation process in sugar industry is considered as the plant as shown in Figure 4.16. The number of effects of the process plant is five and fruit juice having nominal 15 % sucrose concentration. (brix) is concentrated up to 72 %. Juice to be concentrated enters the first effect with concentration C_0, flowrate F_0, enthalpy H_0 and temperature T_0. Steam of rate S is injected to the first effect to vaporize water from first effect producing vapor O_1, which is directed to the next effect with a vapor deduction VP_1. The first effect liquid flow rate F_1 at a concentration C_1 goes to the tube side of the second effect. The vapor from the last effect goes to the condenser and the syrupy product from there goes to the crystallizer. The liquid holdup of i th effect is W_i. The most important measurable disturbance of the plant is the demand of steam by the crystallizer, which is deducted from the third effect. Hence, it is necessary to control the brix C_3 . The other control objective isC_5.

The following model equations were considered to get the LTI transfer function for every effect..

Material balance:

$$\frac{dw_i}{dt} = F_{i-1} - F_i - O_i \qquad (4.20)$$

Where $O_1 = K_1 S$, $O_i = K_i (O_{i-1} - VP_{i-1})_{i=2-5}$

K_i = static gain

Component balance:

$$\frac{dC_i}{dt} = \frac{1}{W_i}[F_{i-1}C_{i-1} - F_i C_i] \qquad (4.21)$$

The dynamics of each portion of the plant (each effect here) were identified in state space domain. The variable with their steady state values are presented in Table 4.2. Two numbers of direct synthesis controllers were designed to controlC_3 and C_5. Direct inverse neural network controllers (DINN) were also designed to fulfill two numbers of control objectives. The performance of the DINN controllers were compared with direct synthesis controllers both in servo as well as regulator mode. It is evident from Figures 4.18 and 4.19 that ANN controllers served well in the servo mode whereas direct synthesis or model based controllers

did well in their disturbance rejection performance. Plant wide simulation was done in SIMULINK platform.

6 CONCLUSION

The present chapter is revealing the successful implementation of NNPLS controllers in controlling multivariable distillation processes and successful implementation of neural controller in a *plant wide* process. But the excerpts of the present chapter are far reaching. The physics of data based model identification through PLS, and NNPLS, their advantages over other time series models like ARX, ARMAX, ARMA, the merits of NNPLS based control over PLS based control were addressed in this chapter. The revelation of this chapter has got a sheer intellectual merit.

The NN controllers worked satisfactorily in regulator mode for the chosen plant wide process. In fine, it can be concluded that the processes need to be simulated rigorously under ASPEN PLUS environment before making any further comments on the applicability of the NN controllers in plant wide process, in general.

Table 4.1 Designed networks and their performances in $(2 \times 2, 3 \times 3 \& 4 \times 4$) Distillation processes.

system	Mode of Operation	Identified NNPLS	MSE	R^2
2×2	Servo	G1	0.00068067	0.994
		G2	0.0002064	0.999
	Regulatory	Gd1	0.00060809	0.995
		Gd2	0.00016509	0.998
3×3	Servo	G1	0.0038066	0.964
		G2	0.00702021	0.965
		G3	0.001353	0.962
	Regulatory	Gd1	0.0006922	0.99
		Gd2	0.0098331	0.962

		Gd3	**0.0002519**	**0.996**
4 × 4	**Servo**	**G1**	**0.000666**	**0.9882**
		G2	**0.0046903**	**0.9662**
		G3	**0.0016072**	**0.9773**
		G4	**0.00030866**	**0.9891**
	Regulatory	**Gd1**	**0.00040426**	**0.9918**
		Gd2	**0.0004304**	**0.9931**
		Gd3	**0.00012274**	**0.9965**
		Gd4	**0.0014292**	**0.9640**

Table 4.2 Steady state values for variables of Juice concentration plant

Input	First effect	Second effect	Third effect	Fourth effect	Fifth effect
$C_0 = 14.74$ kg/m^3 F_0 $= 187.97$ kg/s h_0=3.87 kJ/kg^0C SS=56.78 kg/s	W_1 $= 49.5\ m^3$ F_1 $= 131.19$ kg/s $C_1 = 21.12$ kg/m^3 $O_1 = 30.33$ kg/s h_1=3.66 kJ/kg^0C	$W_2 = 25\ m^3$ $F_2 = 00.86$ kg/s $C_2 = 27.47$ kg/m^3 $O_2 = 25$ kg/s h_2=3.5 kJ/kg^0C	W_3 $= 22\ m^3$ F_3 $= 75.864$ kg/s $C_3 = 36.52$ kg/m^3 $O_3 = 25.92$ kg/s h_3=3.27 kJ/kg^0C	$W_4 = 18\ m^3$ $F_4 = 49.94$ kg/s C_4 $= 55.482$ kg/m^3 $O_4 = 12.79$ kg/s h_4=2.8 kJ/kg^0C	W_5 $= 13.54\ m^3$ $F_5 = 37.15$ kg/s $C_5 = 72$ kg/m^3 $O_5 = 12.79$ kg/s h_5=2.314 kJ/kg^0C

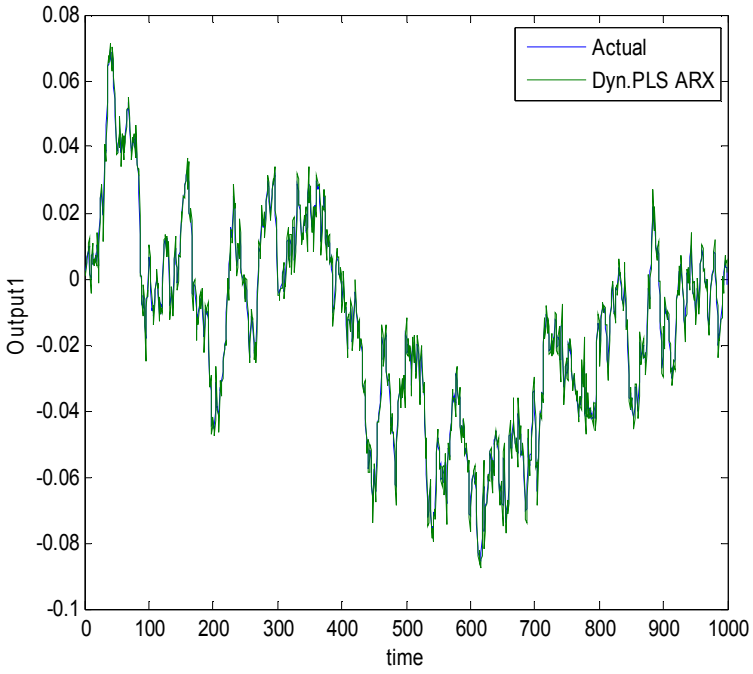


Fig. 4. 3 Comparison between actual and ARX based PLS predicted dynamics for output1 (top product compositionX_D) in distillation process.

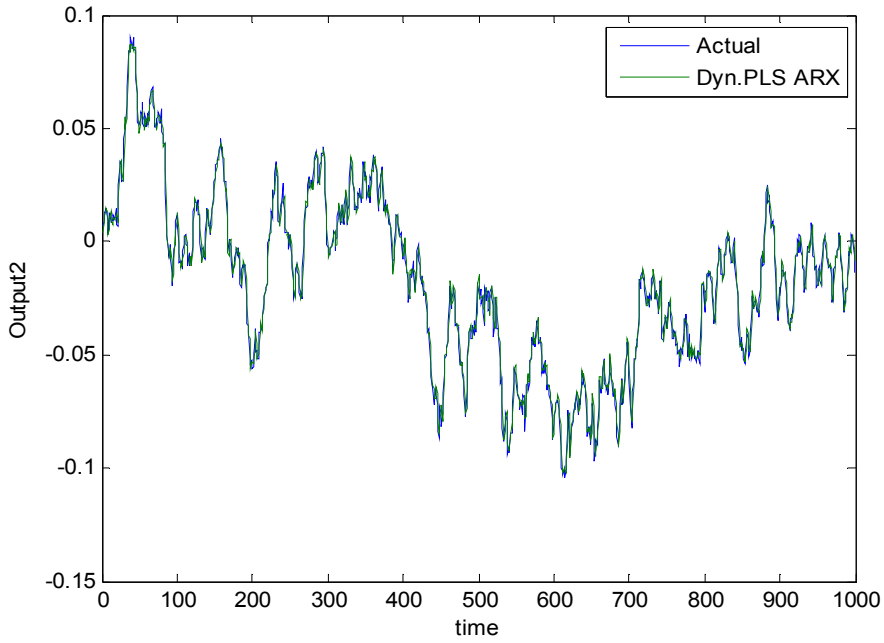


Fig.4.4 Comparison between actual and ARX based PLS predicted dynamics for output2 (bottom product compositionX_B) in distillation process.

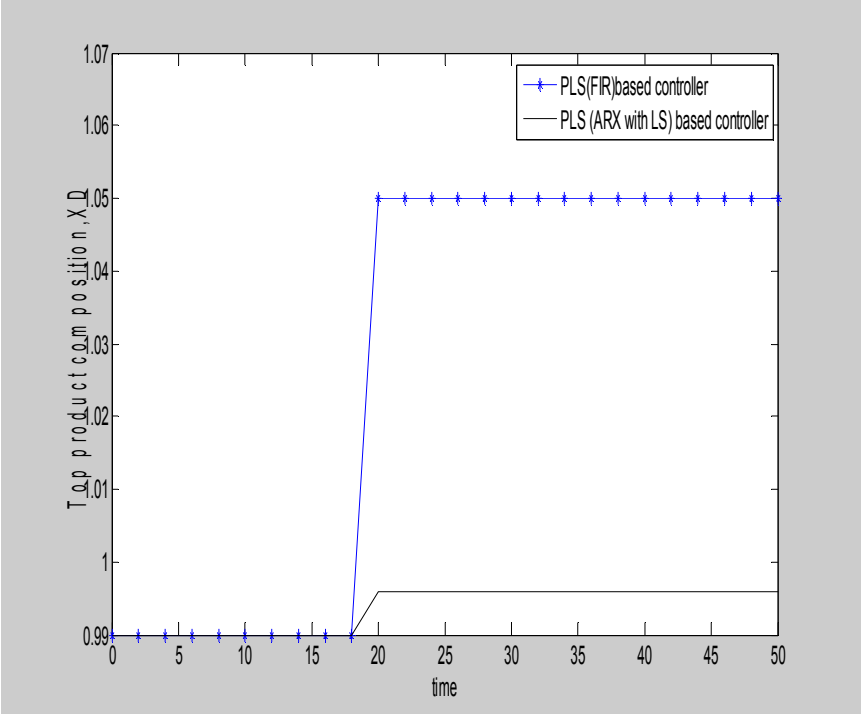


Fig.4.5 Comparison of the closed loop performances of ARX based and FIR PLS controllers for a set point change in X_D from 0.99 to 0.996.

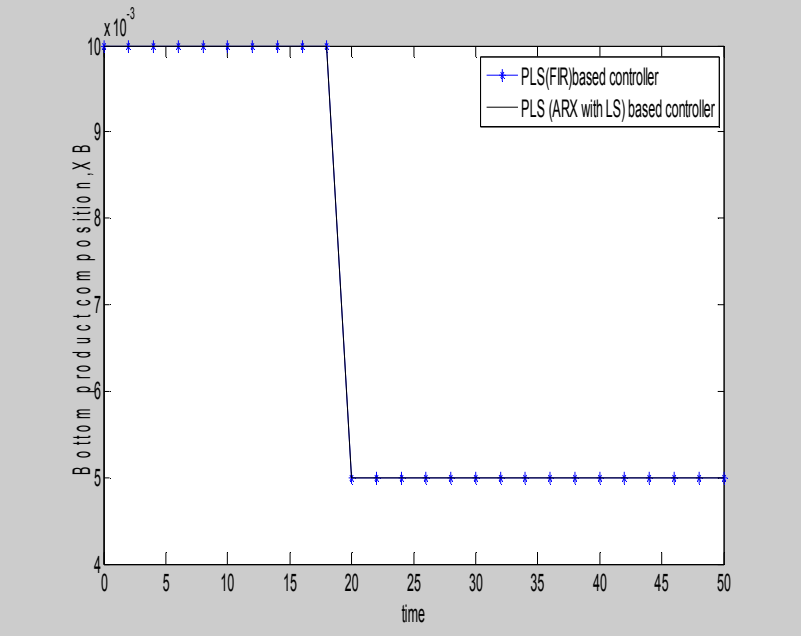


Fig. 4.6 Comparison of the closed loop performances of ARX based and FIR PLS controllers for a set point change in X_B from 0.01 to 0.005

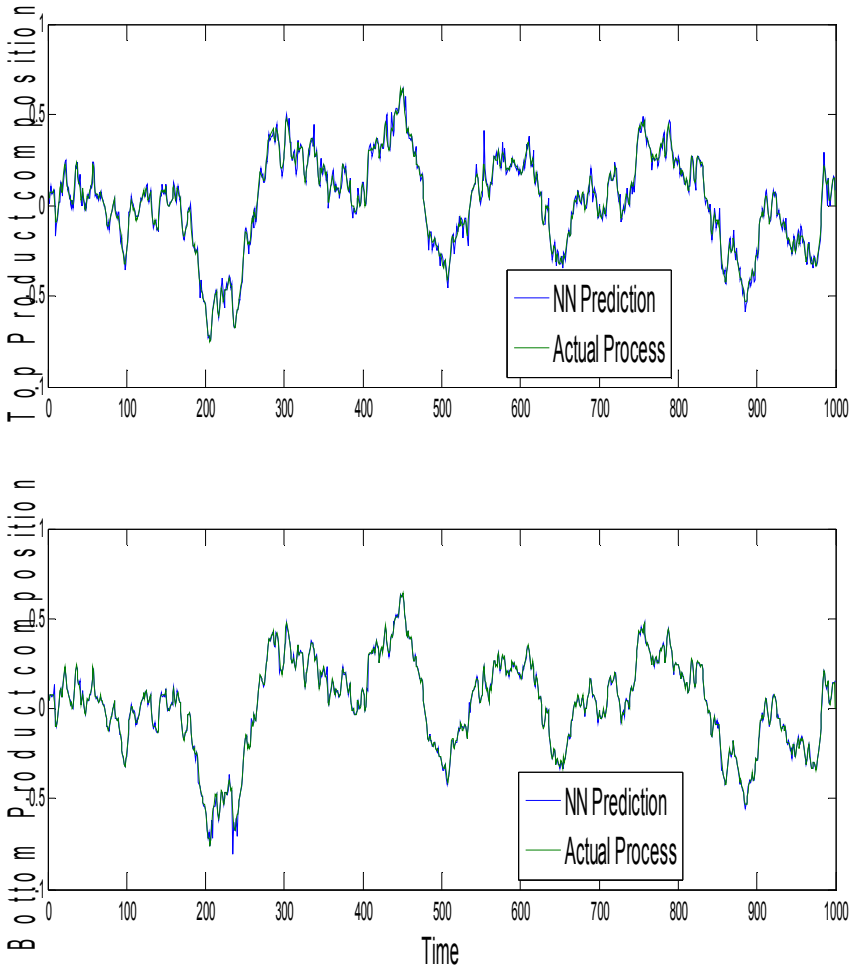


Fig. 4.7 Comparison between actual and neutrally identified outputs of a (2×2) Distillation process using projected variables in latent space.

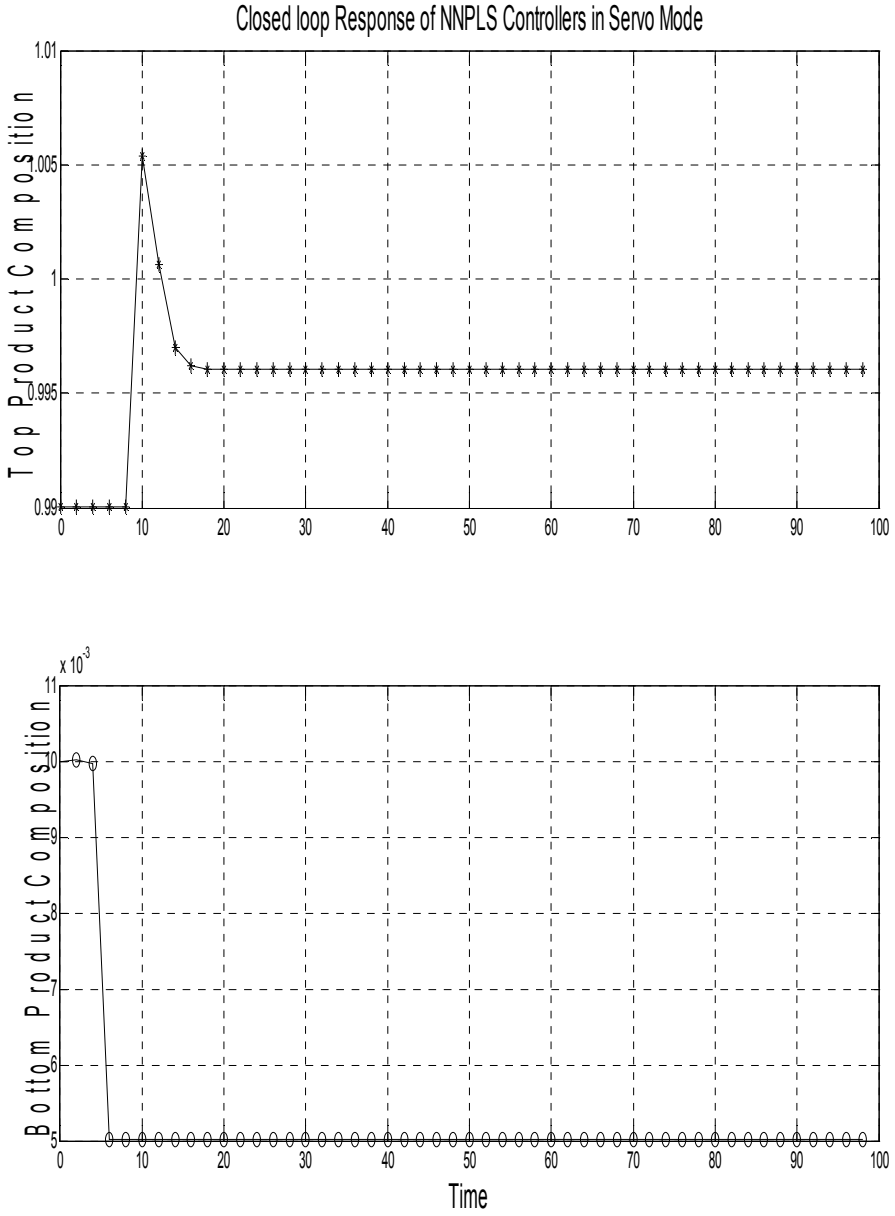


Fig. 4.8 Closed loop response of top and bottom product composition using NNPLS control in servo mode.

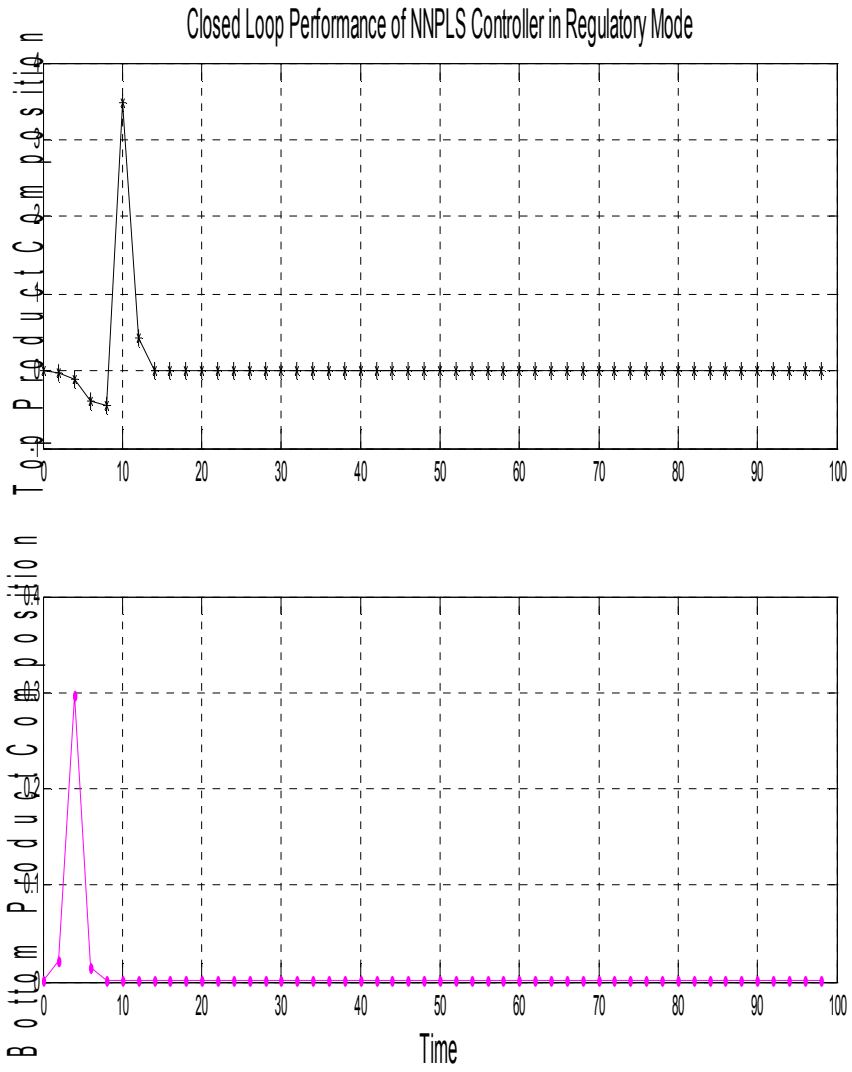


Fig. 4.9 Closed loop response of top and bottom product composition using NNPLS control in regulatory mode.

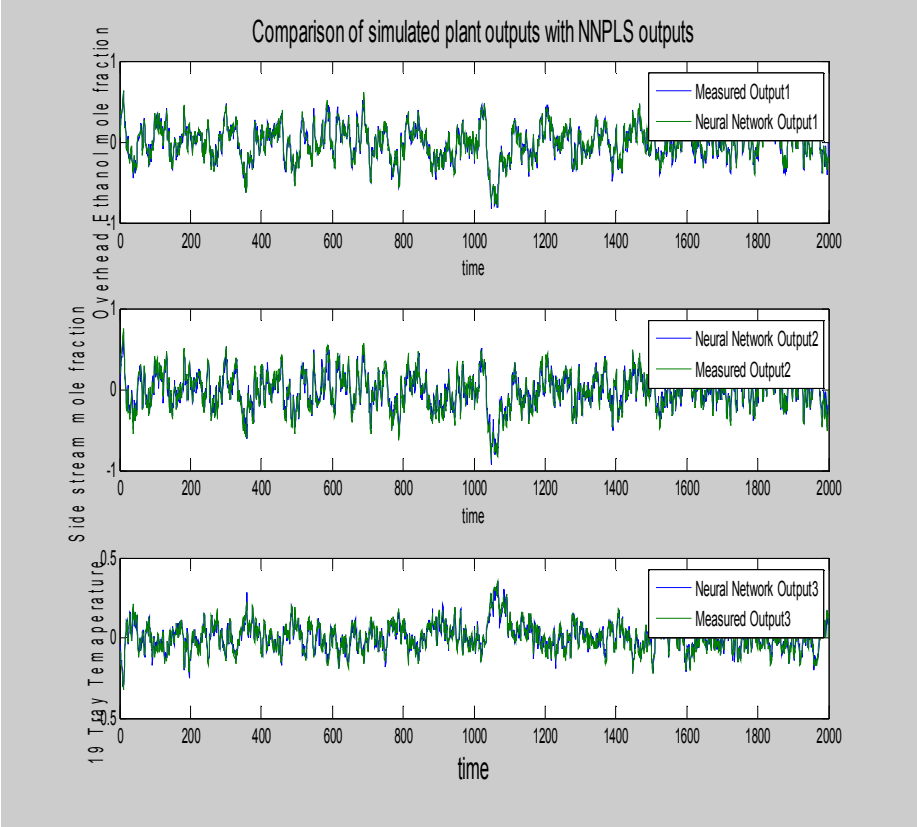


Fig. 4.10 Comparison between actual and NNPLS identified outputs of a (3×3) Distillation process using projected variables in latent space.

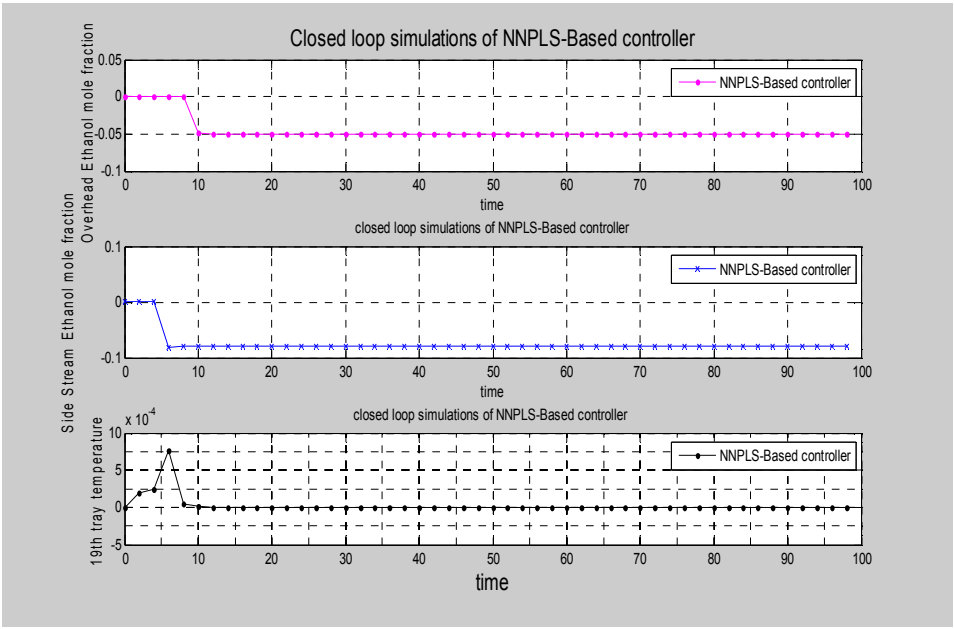


Fig. 4.11 Closed loop response of the three outputs of a (3×3) distillation process using NNPLS control in servo mode.

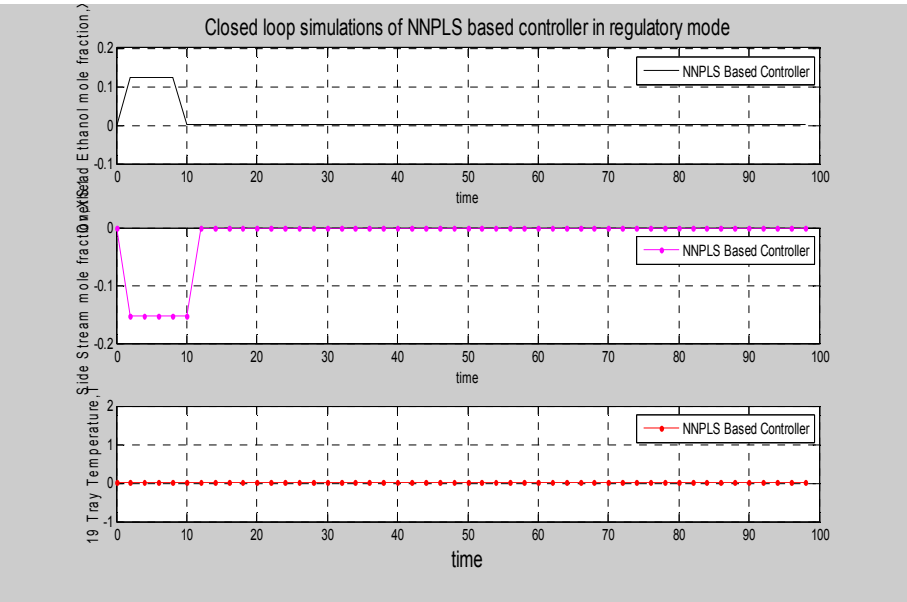


Fig. 4.12 Closed loop response of the three outputs of a (3×3) distillation process using NNPLS control in regulator mode.

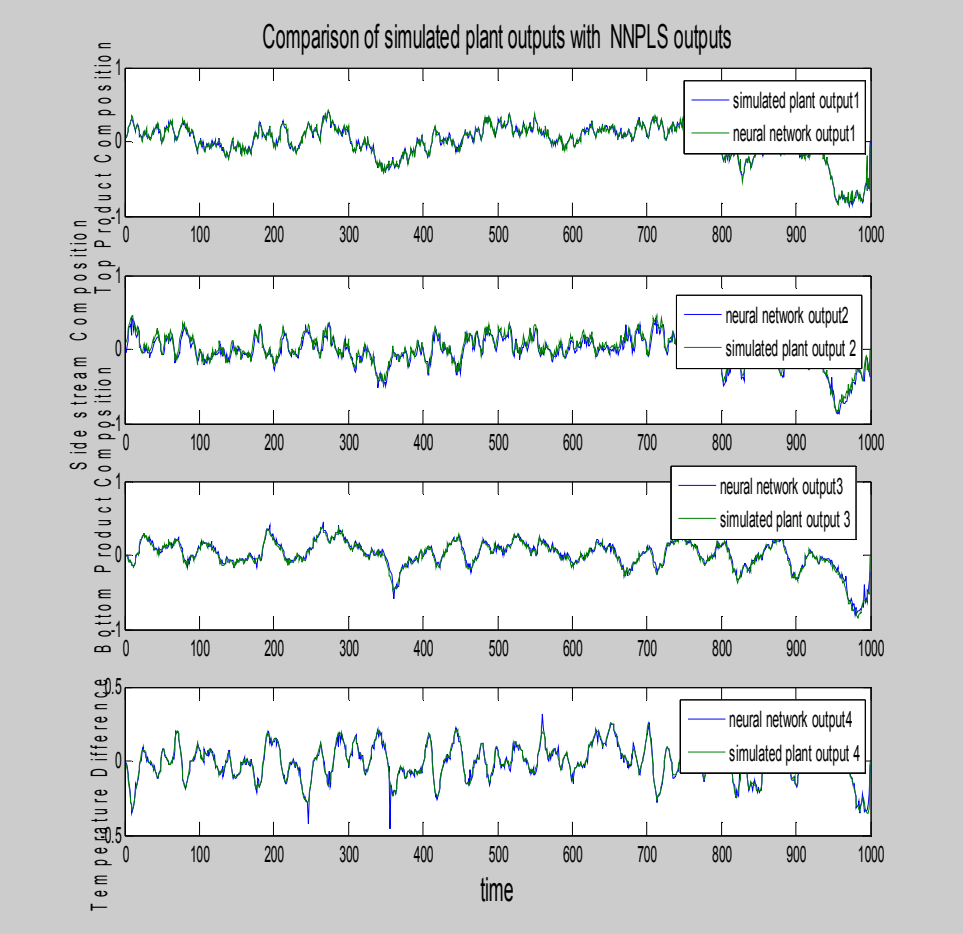


Fig. 4.13 Comparison between actual and NNPLS identified outputs of a (4×4) Distillation process.

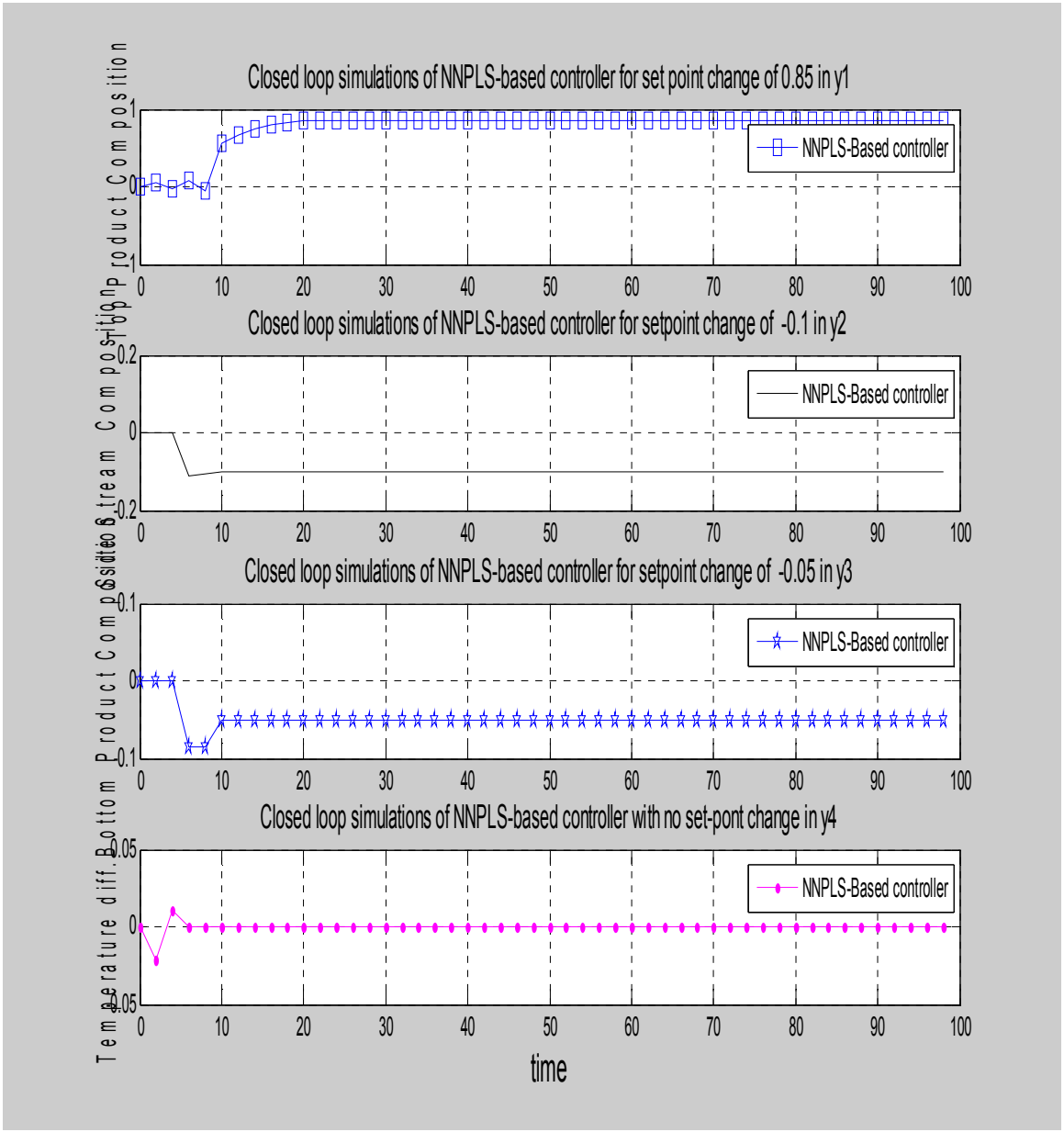


Fig. 4.14 Closed loop response of the four outputs of a (4×4) distillation process using NNPLS control in servo mode.

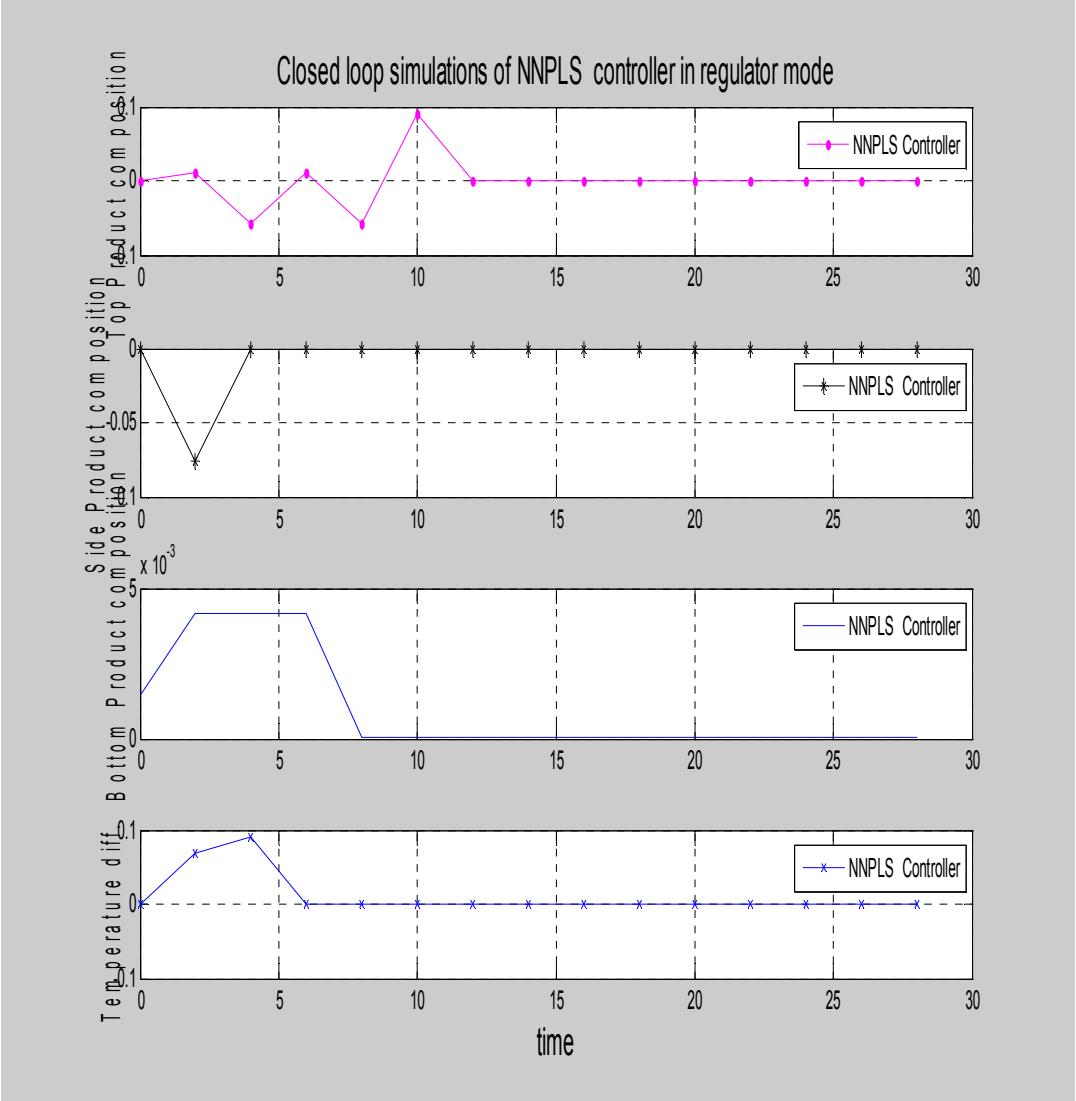


Fig. 4.15 Closed loop response of the four outputs of a (4×4) distillation process using NNPLS control in regulator mode.

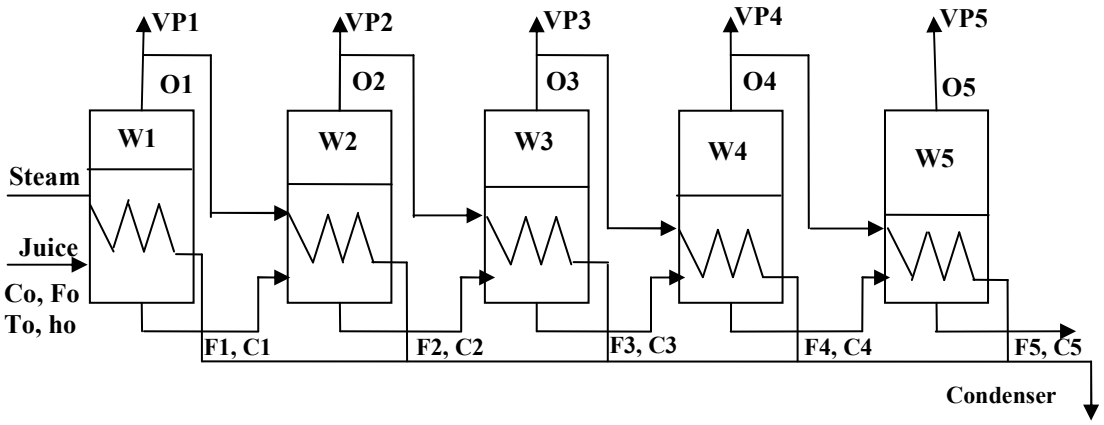


Fig. 4.16 Juice concentration plant

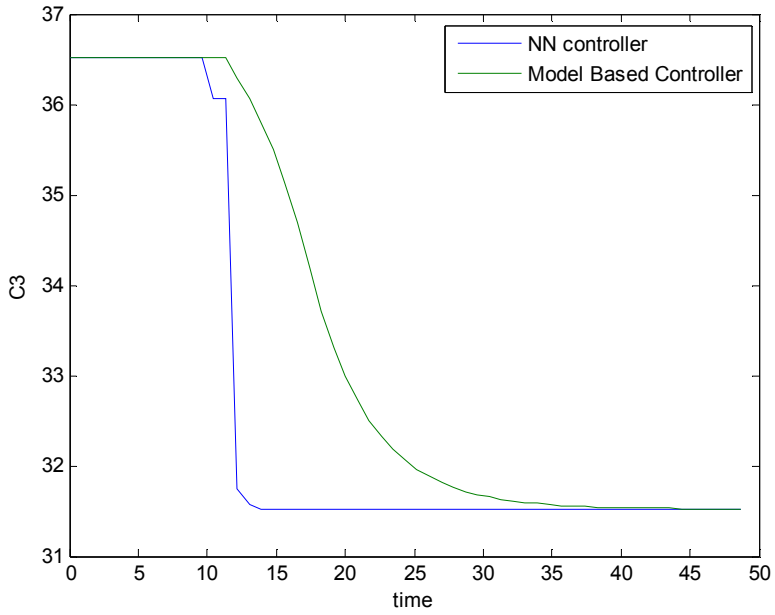


Fig. 4.17 (a) Comparison between NN and model based control in servo mode for maintaining brix from 3 [rd] effect of a sugar evaporation process plant using *plant wide* control strategy.

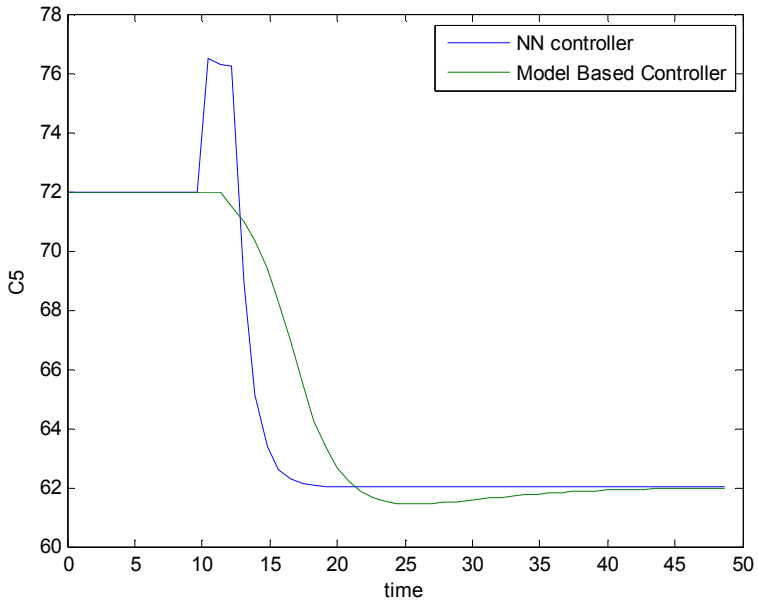


Fig. 4.17 (b) Comparison between NN and model based control in servo mode for maintaining brix from 5 [th] effect of a sugar evaporation process plant using *plant wide* control strategy.

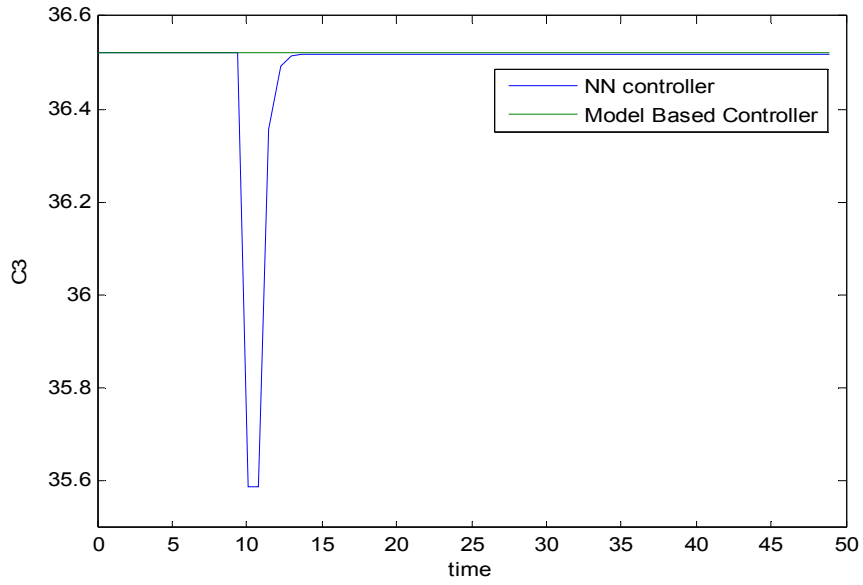


Fig. 4.18 (a) Comparison between NN and model based control in regulator mode for maintaining brix (at its base value) from 3 [rd] effect of a sugar evaporation process plant using *plant wide* control strategy.

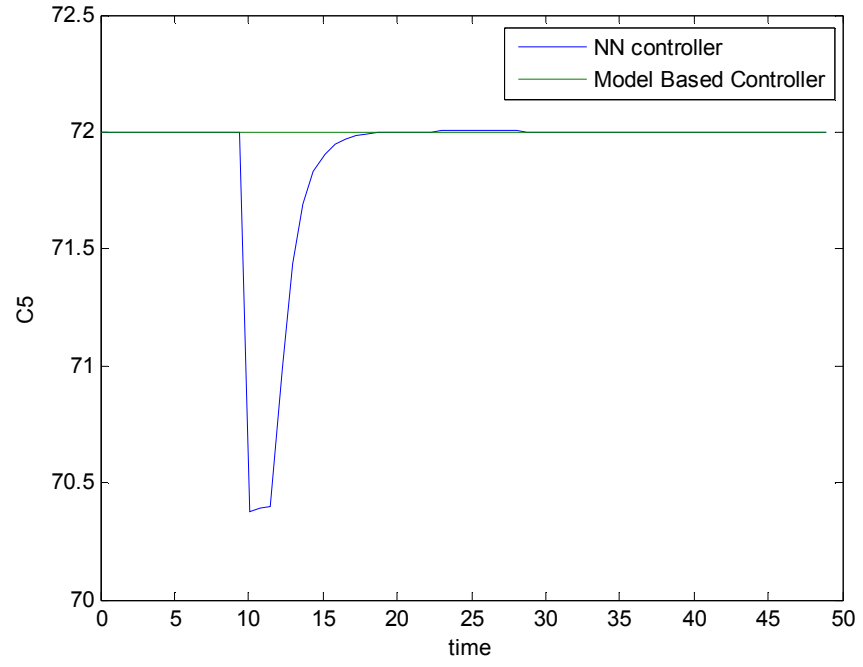


Fig. 4.18 (b) Comparison between NN and model based control in regulator mode for maintaining brix (at its base value) from 5 [th] effect of a sugar evaporation process plant using *plant wide* control strategy.

REFERENCES

1. Qin, S. J., McAvoy, T. J., "Nonlinear PLS modeling using neural network." Comput. Chem. Engr.1992, 16(4): 379-391.
2. Qin, S. J., "A statistical perspective of neural networks for process modelling and control," In Proceedings of the 1993 Internation Symposium on Intelligent Control. Chicago, IL, 1993: 559-604.
3. Kaspar, M. H., Ray,W. H., "Dynamic Modeling For Process Control." Chemical Eng. Science. 1993, 48 (20): 3447-3467.
4. Lakshminarayanan, S., Sirish, L., Nandakumar, K., "Modeling and control of multivariable processes: The dynamic projection to latent structures approach." AIChE Journal. 1997,43: 2307-2323, September 1997.
5. Damarla, S., Kundu, M., "Design of Multivariable NNPLS controller: An alternative to classical controller." (communicated to Chemical product & process modeling)
6. Anderson, B. D. O., Moore, J. B., "Optimal control –linear quadratic methods." Prentice Hall, 1989.
7. Mohtadi, C., Shah, S. L., Clarke, D. W., "Multivariable adaptive control without a prior knowledge of the delay matrix." Syst. Control Lett. 1987, 9: 295-306.

Chapter 5

CONCLUSIONS AND RECOIMMENDATION FOR FUTURE WORK

Chapter 5

CONCLUSIONS AND RECOMMENDATION FOR FUTURE WORK

5.1 CONCLUSIONS

Present work could successfully implement various MSPC techniques in process identification, monitoring and control of industrially significant processes. The issues regarding the implementation were also addressed. With the advancement of data capture, storage, compression and analysis techniques, the multivariate statistical process monitoring (MSPM) and control has become a potential and focused area of R&D activities. From this perspective, the current project was taken up. Every physical/mathematical model encourages a certain degree of empiricism, purely empirical models some time are based on heuristics, hence process monitoring & control on their basis may find inadequacies; especially for non-linear & high dimensional processes. Data driven models may be a better option than models based on first principles and truly empirical models in this regard. Efficient data pre-processing, precise analysis and judicious transformation are the key steps to success of data based approach for process identification, monitoring & control. The deliverables of the present dissertation are summarized as follows:

- Implementation of clustering time series data and moving window based pattern matching for detection of faulty conditions as well as differentiating among various normal operating conditions of Bioreactor, Drum-boiler, continuous stirred tank with cooling jacket and the prestigious Tennessee Eastman challenge processes. Both the techniques emancipated encouraging efficiencies in their performances.

- Identification of time series data/process dynamics & disturbance process dynamics with supervised Artificial neural network (ANN) model

- Identification of time series data/process dynamics in latent subspace using partial least squares (PLS).

- Design of multivariable controllers in latent subspace using PLS framework. For multivariable processes, the PLS based controllers offered the opportunity to be designed as a series of decoupled SISO controllers.

- Neural nets trained with inverse dynamics of the process or direct inverse neural networks (DINN) acted as controllers. Latent variable based DINNS' embedded in PLS framework was the termed coined as NNPLS controllers. The closed loop performance of NNPLS controllers in (2×2), (3×3), and (4×4) Distillation processes were encouraging both in servo and regulator modes.

- Model based *Direct synthesis* and DINN controllers were incorporated for controlling brix concentrations in a multiple effect evaporation process plant and their performances were comparable in servo mode but model based *Direct synthesis* controllers were better in disturbance rejection.

In fine, it can be acclaimed that the present work not only did address monitoring & control of some of the modern day industrial problems, it has also enhanced the understanding about data based models, their scopes, their issues, and their implementation in the current perspective. The subject matter of the present dissertation seems to be suitable to cater the need of the present time.

5.2 RECOMMENDATION FOR FUTURE WORK

- *Plant wide control* using non-traditional control system design requires to be implemented in complex processes.
- Integration of statistical process monitoring and control for industrial automation.
- Building of Expert system or Knowledge-based systems (KBS), which is emulating the decision-making capabilities and knowledge of a human expert in a specific field.