Fault Diagnosis In Batch Process Monitoring

M.Venkata Ramana

A Report Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Technology

Department of Electrical Engineering

June, 2018

Declaration

'I declare that this written submission represents my ideas in my own words, and where ideas or words of others have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

mvramana

(Signature)

M.Venkata Ramana

(M.Venkata Ramana)

EE16mtech11035

(Roll No.)

Approval Sheet

This Thesis entitled Fault Diagnosis in Batch Process Monitoring by M.Venkata Ramana is approved for the degree of Master of Technology from IIT Hyderabad

M. Vidyasagar

(Prof. M Vidyasagar) Examiner
Dept. of Electrical Eng
IITH

(Dr. Siva Kumar K) Examiner
Dept. Electrical Eng
IITH

(Dr. Ketan Detroja) Adviser
Dept. of Electrial Eng
IITH

(Dr. Phanindra Jampana) Chairman
Dept. of Chemical Eng
IITH

Acknowledgements

I would like to thank my supervisor Dr. Ketan Detroja for his continuous encouragement and constructive comments throughout my project work.

Abstract

Every process plant nowadays highly complex to produce high-quality products and to satisfy demands in time. Other than that, plant safety is also crucial event had to be taken care to increase plant efficiency. Due to poor monitoring strategies leads to huge loss of income and valuable time to regain its normal behavior. So, when there is any fault occurs in the plant it should be detected and need to take supervisory action before propagating it to new locations and new equipment failure leads to plant halt. Therefore process monitoring is very crucial event had to be done effectively.

In Chapter 1 Importance of fault detection and diagnosis(FDD) in plant monitoring, what are the typical situations will leads to fault and their causes of fault is discussed. How data will be transformed in different stages in diagnostic system before certain action, desirable characteristics for good diagnostic systems are discussed briefly. And in final part of this chapter what are the basic classifications of FDD methods are discussed. Principle component analysis is multivariate statistical technique helps to extract major information with few dimensions. Dimensionality of reduced space is very low compared to original dimension of data set. Number of principle component(PC) selection depends on variability or information required in lower dimensional space. So PCA is effective dimensionality reduction technique. But for process monitoring both PC and residual space are important. In chapter 2 mainly discussed about PCA and its theory.

Batch Process Monitoring is relatively not easy to monitor compared to Continuous process because of their dynamic nature and non-linearity in the data. So there are methods like MPCA(multi-way Principle component analysis), MCA(multi-way correspondence analysis) and Kernal PCA, Dissimilarity Index based(DISSIM) etc., are there to monitor batch process. Kernal based methods need to choose right kernal based on the non-linearity in the data. Dissimilarity Index based methods well suits for continuous process monitoring since it can able to detect the changes in distribution of data. Extension of DISSIM method to batch process monitoring is EDISSIM, which is discussed in chapter 3. And also MPCA is very traditional method which can able to detect abnormal sample but these cannot be able to detect small mean deviations in measurements. Multi Way PCA is applied after unfolding the data. Batch data Unfolding discussed in section 3.2 and selection of control limits discussed in 3.2.3. Apart from these methods there is another strategy called Pattern matching method introduced by Johannesmeyer. This method will helps to quickly locate the similar patterns in historical database. In Process industries we frequently collect the data so that there will be lot of data available. But there will be less information containing in it, used PCA to extract main information. In pattern matching strategy to detect the similar patterns in historical data base we need to provide some quantitative measure of similarity between two data sets those are similarity factors. So by using PCA method we are extracting high informative data in lower dimensional space. So Using PCA method similarity factors are calculated. Different similarity factors and their calculation is shown in chapter 4. On-line monitoring of Acetone Butanol batch process discussed using pattern matching strategy. Acetone Butanol fermentation process mathematical model will be simulated to different nominal values with different operating conditions to develop historical database. In this case study there will be 500 batches with five operations conditions like one NOC and 4 different faulty operation batches. In each batch there will be 100 batches. After calculation of similarity factors instead of going for candidate pool selection directly we are trying to detect the batches which are similar to snapshot data. Performance of On-line monitoring using pattern matching strategy is discussed. On-line monitoring strategy will change the way we are anticipating

the unfilled data. Here we are trying to fill with reference batch data. Reference data will be average of NOC batches. The performance of this method verified in MATLAB as shown in section 4.3.

In Chapter 5 described average PC's(Principle components) model. This method will helps to decrease the efforts in candidate pool selection and evaluation to find snapshot data in historical database. And also Incremental average model building and model updating will improves the quality of model ultimately.In incremental average model building If any of the snapshot data classified as any of the already existed operating condition data set it will be used in building average model. If not existed in any of the operating condition data set utilized to update average model. This method applied on Acetone Butanol fermentation process data and verified. Because of the fact that batch data highly non linear in nature So PCA not able to handle non-linear correlations. And pattern matching approach using PCA average model not give good discrimination. For better discrimination ability and self aggregation can be possible using Corresponding Analysis because of non-linear scaling. In chapter 6 pattern matching approach using corresponding analysis has been discussed briefly. Results obtained using CA based similarity factor displayed for Acetone Butanol fermentation process case study.

Contents

Chapter 1

Introduction

In last few decades process industries like chemical, petrochemical, cement, steel, automobile etc., are facing the problem of effective fault detection and diagnosis. Because the measurements obtained in these process plants are very large, if there is any fault occurrence in the process, detection of that fault and categorization of that fault and taking necessary action against the fault to keep our process plant in normal operation is major challenge. This is known as abnormal event management(AEM) [1]. We should have to detect the fault before going it into abnormal region, so that we can control from the productivity loss, major accidents which leads to loss of billions of dollars. Therefore, we have to detect the fault in time and we take necessary action against the fault i.e., isolation of that fault as early as possible. To improve abnormal event management, we should not rely on human involvement some times which will make the situation further worse. In big process industries there will be hundreds of process variables data will get for every second, by observing those data humans may not able to distinguish the normal and abnormal situation and their classification of fault may not be easy. Sometimes we will face the problem of measurement data may not be sufficient to take the proper action towards fault detection and diagnosis therefore human intervention in abnormal situations is not good idea. In process industries due to poor diagnostic activity we may not able take proper action towards fault in time therefore, apart from major accidents many minor accidents will takes place daily.

Before entering into brief explanation of different ways of fault detection and diagnosis, mentioned key definitions, different types of faults in a process plant, and desirable characteristics of diagnostic system, Need of transformations of the measured data from process variables, sensors, actuators etc., to detect the fault and their class of fault is discussed and Finally principle component analysis based fault detection and diagnosis explained.

1.1 Key definitions

These are the some of the definitions which are there in literature and these definitions are accepted by control societies across the world [2].

- Fault: An unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable/usual/standard condition.

- Failure: A permanent interruption of systems ability to perform a required function under specified operating conditions.
- Malfunction: An intermittent irregularity in the fulfillment of systems desired function.
- Residual: A fault indicator, based on a deviation between measurements and model equations-based computations.
- Fault detection: Determination of the faults present in a system and the time of detection.
- Fault isolation: Determination of kind, location and time of detection of a fault. Follows fault detection.
- Fault identification: Determination of size and time-variant behavior of a fault. Follows fault isolation
- Fault diagnosis: Determination of kind, size, location and time of detection of a fault. Follows fault detection,Includes fault isolation and identification.
- Diagnostic model: A set of static or a dynamic relation which link specific input variables-the symptoms-to specific output variables-the faults.
- Batch Process: The process which is consequence of discrete tasks that have to follow a predefined sequence from raw materials to final products is known as batch process [3].

1.2 Diagnostic Frame work

Below fig.1.2 is the simple diagnostic frame work with major components like sensors, actuators, plant and associated failures in it. We can broadly classify different failures as shown below.

1.2.1 Parameter changes in a model

Practically process will undergo different variations. It always has some deviation from modeling output since we may not consider all the parameters and limitations on the parameters. That's why there will be some mismatch will be there during normal and abnormal situation. Some processes like disturbances, uncertainties while we are modeling we just lumped it to account it into single parameter which will cause interactions.

1.2.2 Structural changes

Due to structural changes like valve stucking, leaking or damaged pipe leads to loss of information flow between different variables. If these type of malfunctions occur our designed model equation may not be sufficient to handle the situation therefore we have to restructured the model.

1.2.3 Malfunctioning sensors and actuators

Malfunctions in sensors and actuators very common problem leads to feedback failures. Because of feedback signal may not reach controller in time or may not reach ever and wrong readings of

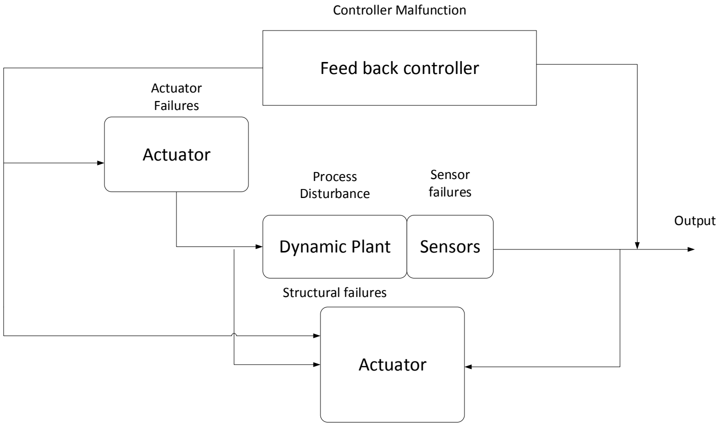


Figure 1.1: Simple diagnostic Model

instruments. In sensors and actuators there is a problem of biasing when the reading taken it will with some bias either positive or negative which will leads to misunderstanding of diagnostic activity.

1.3 Desirable characteristics of a fault diagnostic system

Below are the some of the desirable characteristics of diagnostic systems, to compare different strategies these we can consider as the benchmark in order to compare different diagnostic strategies in a process industry. Any diagnostic strategy may not satisfy all these characteristics because of their performance criteria.

- Quick detection and diagnosis: A good diagnostic system should respond quickly in detecting and diagnosing process malfunctions. A quick response to fault diagnosis and desired performance two conflicting things. A system that is designed to detect a failure (during abnormal changes) quickly will be sensitive to noise and can lead to regular false indications during Normal operation which should be avoided. A good daignostic will have

- Isolability: It is the ability of diagnostic system to distinguish different failures. In ideal (free from noise and uncertainties) diagnostic classifier should generate output that is orthogonal to faults that have not occurred. Sometimes these faults will overlap with modeling uncertainties.

- Robustness: Any Diagnostic system should be insensitive to noise and uncertainties of the system. To avoid frequently occurred false alarms threshold should be chosen reasonably. System performance should degrade reasonably with the time but it should not fail to operate abruptly.

- Novelty identifiability: Minimum requirement of diagnostic system is to be able to decide, given current process conditions, whether the process is functioning normally or abnormally, and if abnormal, whether the cause is a known malfunction or an unknown, novel, malfunction.

- Adaptability: Process should adapt general changes like production quantities increment or decrement, Environmental changes and disturbances existed in the process for future references. Structural changes should be adapted gradually to improve diagnostic range.
- Explanation facility: Diagnostic system should explain how fault is originated and how it is propagated to the current situation apart from detection of the fault.
- Modeling requirements: Modeling effort of diagnostic should as minimum as possible for the fast and easy development of real time classifier.
- Storage and computational requirements: For quick decision making require algorithms and implementation which is computationally less complex, but it requires high storage requirements. Therefore, we should have to take care while we are designing diagnostic system.
- Multiple fault identifiability: Multiple fault identifiability is crucial characteristic for any good diagnostic system. Diagnostic system should have able to detect multiple faults and it should able to distinguish the faults based on their origin of fault generation and it should take decision with high precision. Since process plant is highly non linear and huge amount of variables and their interdependency will leads to multiple faults.

1.4 Transformations of measurements in a diagnostic system

In measurement space we will have process data for different process variables like $X_1, X_2, ..., X_p$ will be input to the diagnostic system. And it will undergo different transformations or mappings while diagnosis. Measurement space is the input to diagnostic activity and it will have transformed to feature space is a space of points $y = (y_1, y_2, ..., y_i)$ in which we can have major trends in the data.-From measurement space to feature space can be obtained by either feature selection or feature extraction.

- Feature Selection: In feature selection we will choose some of the variables or attribute selection among available features.

  Suppose in a plant consider four sensor measurements X_1, X_2, X_3, X_4 are available, if two faults F_1, F_2 are occurred in the plant. If we know that F_1 will effect sensor measurement X_1, X_2 and F_2 will effect X_2, X_3 measurements for effective diagnostic activity we will remove X_4 measurement from diagnosis since faults no influence on that measurement.
- Feature Extraction: The Feature extraction can be done by transforming the data in higher dimensional space to lower dimensional space in this lower dimensional space obtained or derived data will be available is non-redundant and informative.

Above techniques will helpful to reduce the dimension in feature space from measurement space. Here y_i is the i^{th} feature which is the function of our measured space. From feature space to decision space is mapped with major desirability like minimizing misclassification. Decision space contains decision variables, by keeping some threshold to detect which decision is desirable for particular fault we can transform featured data to decision space. From decision space to class space we can use Boolean algebra to map into particular class of failure. Finally, class space output will be delivered to user from diagnostic activity. Below fig.1.4 shows simple transformation flow.

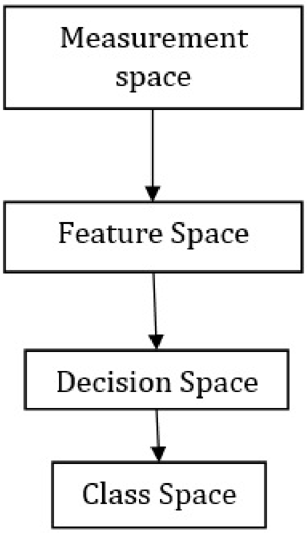


Figure 1.2: Transformations in diagnostic systems

1.5 Classification of diagnostic algorithms

Main components of diagnostic classifier are the type of knowledge and the type of diagnostic search strategy [1]. Basically diagnostic classification is mainly based on priori knowledge available.In priori knowledge there will be failures and symptoms to the particular failures.This knowledge can be developed from a basic understanding of process plant, which is model based knowledge.Other than model based approaches there is a process history based in which priori knowledge will be shallow and evidential coming from past experience with the process plant.The model based methods can be classified as Quantitative and Qualitative. Model based methods will be developed from fundamental understanding of process. In Quantitative methods this understanding can be expressed in-terms of mathematical functional relationships between input and outputs, in qualitative methods relationships are expressed in-terms qualitative functions expressed in terms of different units in a process [1].

1.5.1 Quantitative Model Based Approach

For model based approach we are modeling process variables from the input output relationships of process. Which will generate residuals during abnormal operation by comparing actual process output. This is known as quantitative approach to fault detection and diagnosis.

In model based approaches first we have to generate residual which are inconsistencies from the process that can be driven to residual generator. These approaches mainly rely on input output relations and state space models. These residuals will furtherly modified to get effective detection and decision making for potential faults [1].

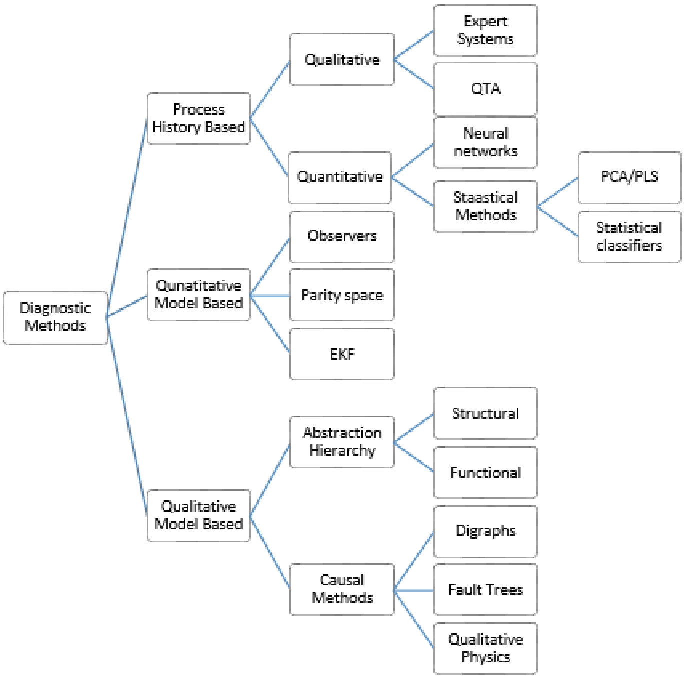


Figure 1.3: Classifications of diagnostic algorithms

Residuals are quantities that represent the inconsistency between the actual plant variables and the mathematical model[1]. In ideal plant residuals will be zero during normal operation but in practically residuals are not zero because of some disturbances, noise etc. To detect fault single residual is enough to detect the fault for isolation purpose diagnostic system requires a set of residuals to avoid mismatch or false operation. To make effective diagnostic activity we have to enhance residuals to fault specified direction. If residuals are white that means uncorrelated with time, will make statistical testing in noisy plant is easy. Below figure 1.4 shows residual generation block diagram.

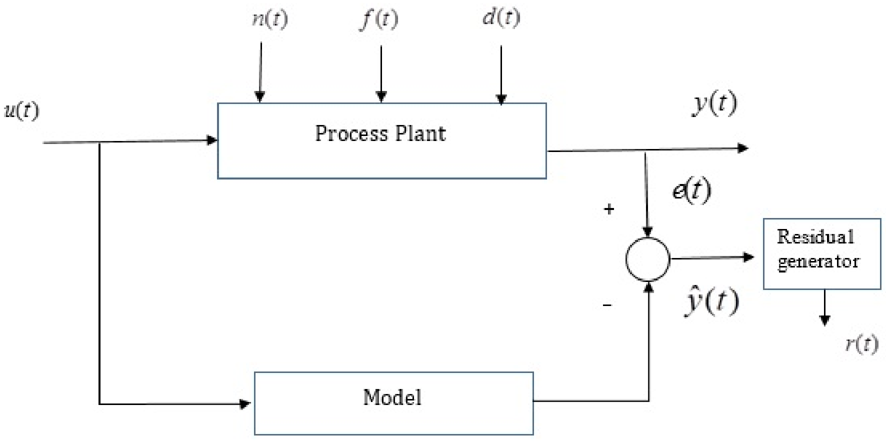


Figure 1.4: Residual generation Block diagram

where $n(t)$=Noise, $f(t)$=Faults, $d(t)$=disturbances, and primary residual generated is $e(t) = y(t) - \hat{y}(t)$, Processed residual is $r(t)$.

1.5.2 Qualitative Model Based Approach

In Qualitative model based , we are approaching with deep understanding of process system which will discriminate fault with high accuracy. In this approach we will use directed graphs, fault trees, Abstraction hierarchies (Analyzing each part in the process with their unique properties). These approaches mainly encounter the problem based on cause and effect strategy therefore we need good understanding which will affect the good diagnostic strategy.

Here Directed graphs, fault tree strategies will give fruitful results but due to complexities and model uncertainties will affect optimal diagnostic strategies. These method follows back propagation search strategies. In fault tree modeling of the process system involves with Boolean gates (OR, NOT, AND) which is very bulky for understanding but it has good diagnostic properties. To minimize the construction of fault trees with optimal number of gates we will go for minimal cut set which is very helpful in minimizing the fault tree modeling efforts.

1.5.3 Process History based Approach

And another classification is based on process history knowledge. In this we will try to reduce the dimension of measured data to featured data which will contain high variation in the data helpful for further decision making effectively in fault diagnosing system. Here we will use statistical methods to detect the abnormalities in the monitored data. These methods need not know the complete knowledge of process plant. One of the more popular method to diagnose the fault is using Principle component analysis(PCA) method. And another non-statistical method is using Neural networks.

Chapter 2

Principle Component Analysis Based Fault diagnosis

2.1 Introduction to Process history based Methods

Main per-requisite for Process history based method is need of large amount of historical process data. This priori data will be feed-ed to diagnostic system after transforming it into lower dimensional space. In the lower dimensional space major trends or variability will be carried as in the original space. This method is known as feature extraction. Based on feature extraction can be done either quantitative or qualitative in nature. Expert systems and trend modeling methods are the major methods uses extraction of qualitative history information. Methods that extract quantitative information can broadly classify as non-statistical and statistical methods [4]. Neural networks are belongs Non-statistical method whereas Principal component analysis is(PCA) or Principle least squares(PLS) methods belongs to statistical feature extraction type. we will mainly discuss about statistical extraction in this chapter.

In a process plant practically there is a effect of random disturbances. Therefore compared to deterministic system stochastic system can't be determined completely from past, present states and from future control actions. Because of that reason we will always keep probabilistic setting on control bounds. During normal operations each variable observation will have their own distributions with nominal mean and variance. When there is fault occurs then there will be deviations in their mean and variance if it crosses more than control bounds then fault should be detected. Earlier days for the efficient quality of products people are attempted to rely on statistical on-line monitoring techniques and change detection techniques are used. She-wart introduced control charts in 1931 called as she-wart control charts and another type of control charts are cumulative sum charts introduced by Page in 1954.These control charts are introduced on the basic assumption of that process will undergoes normal cause variation(known variation). By monitoring with these control charts we can detect the fault and by taking corrective action to driven back our process to normal operation. Normally in a process plant fault may not occurred because of single variable, since most of the variables are inter dependent on each other these univariate charts may not sufficient to take decision because of the fact inability to deal with correlation [4]. Therefore multivariate statistical process control(MSPC) techniques are needed to deal with correlations and to improve process efficiency .

- Motivation to Principle component analysis: Multivariate statistical techniques much needed tool for compressing the data and reducing the dimensionality so that useful information is retained in that lower dimensional space. In this lower dimensional space we can easily analyze the data compared to huge original data set. They can able to handle the noise and correlation to extract true information effectively [4]. This Dimensionality reduction can be effectively done by Principle component analysis(PCA).

We are trying to avoid spurious or correlated data in the measurement space i.e., some of the variables data will be having redundancy which will not helpful for feature extraction or it will not show variability in the original data need to be avoided. It is very helpful for further decision making in diagnostic system effectively if we avoid those data.

2.2 Principle component analysis(PCA)

PCA first proposed by Pearson in the year 1901, after that developed by Hotelling(1947). Basically PCA is a Multivariate statistical technique which will transform the original data set into a lower dimensional space. In this lower dimensional space derived variables are highly uncorrelated.

- PCA will decompose process variables data along orthogonal directions such that in these directions we can capture high variations in original data with few Principle components in this lower dimensional space [4].

Let us consider p-dimensional data set. For feature extraction purpose if we blindly eliminate the some of the variable data set in p-dimensional space then we will end up with mean square error(MSE) may be high MSE or low MSE based on our removed variable data. If these variable data containing high variance, then we are going to loss major trends in the original data (high MSE). Therefore, to reduce the dimension of the data set the main objective is to select the high variance data among available data (or) Eliminate the data which is having low variance

From PCA, we conclude that the variability in the data can be explained by Eigen values of covariance matrix. If higher the Eigen value corresponding principle component will have high variability in the data.

Let us consider process data matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p)$ which is having size $n \times p$, where n is the number of observations of each variable $\mathbf{x}_i$, where p is the no of variables in the process. Among this data points there will be correlated data or redundancy will exist. Therefore, instead of feeding this process data points to diagnostic system, we will feed featured data that means the data in lower dimensional space which is having high variability can be carried with very few derived variables. This featured data dimension is very less comparatively process data dimension. This dimensionality reduction technique can be effectively done by PCA. To approach to featured data we have to decompose covariance matrix orthogonally by using singular value decomposition(SVD) i.e. $\mathbf{\Sigma} = \mathbf{P\Lambda P}^T$ where $\mathbf{\Sigma}$ is the covariance matrix of mean centered standardized data of $\mathbf{X}$, and $\mathbf{P}$ is the orthonormal Eigen vectors corresponding to singular values (Eigen value(λ)=square(σ_i)) arranged in decreasing order. Also known as Principle component weighted matrix. Here $\mathbf{\Lambda}$ is diagonal matrix containing singular values. Now the transformed data i.e. score matrix can be obtained as $\mathbf{T} = \mathbf{XP}$

, $\mathbf{P}$ is the matrix containing weighted or principle component matrix with p principle component weighted vectors. Arranged such way that Which will explaining more variability corresponding to high to least Eigen values. Here each weight vector in matrix $\mathbf{P}$ is mutually orthogonal to each other and length of each weight Vector is unity(orthonormal). i.e. $\mathbf{p}_i\mathbf{p}_j{}^T = 0,\ i \neq j$ and $\mathbf{p}_i\mathbf{p}_j{}^T = 0,\ i = j$. And in Score matrix $\mathbf{T}$ each score vectors satisfies $\mathbf{t}_i\mathbf{t}_j{}^T = 0$, that means mutually orthogonal to each score vector. Now the Score Vector can be written as $\mathbf{t}_i = \mathbf{X}\mathbf{p}_i$. Therefore the original data matrix can be written as $\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}}$, since $\mathbf{P}$ satisfies orthogonal and orthonormal property i.e.$\mathbf{P}\mathbf{P}^{\mathrm{T}} = \mathbf{I}$.

Now, Approximation of $\mathbf{X}$ will be written, with $\mathbf{P}$ containing very few weighted vectors corresponding to selected m number of Principle components is $\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}_m{}^{\mathbf{T}} = \sum_{i=1}^{m} \mathbf{t}_i\mathbf{p}_i{}^T$. Where $m << p$, $\mathbf{t}_i$ is score vector and $\mathbf{p}_i$ is the weighted vector. Original data matrix $\mathbf{X} = \mathbf{T}\mathbf{P}_m{}^{\mathbf{T}} + \mathbf{E} = \sum_{i=1}^{m} \mathbf{t}_i\mathbf{p}_i{}^T + \mathbf{E}$ [5], where $\mathbf{E}$ is the residual matrix containing noise components.

2.2.1 Basic Formulas

- Mean $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
- Standard Deviation $(S) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}}$
- Covariance$(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n-1}$

2.2.2 Steps to find PCA

- Start with measured data $\mathbf{X}$ for n observations for each p variables.
- Find Mean centered(zero mean) standardized data Matrix of size $n \times p$
- Calculate the covariance matrix $(\mathbf{\Sigma})$ for mean centered standardized data having the size $p \times p$.
- Find the Eigen values and Eigen vectors of the covariance matrix $(\mathbf{\Sigma})$ and arrange in decreasing order Eigen vectors corresponding to each Eigen value.
- Choose number of principle components(eigen vectors) and form a featured vector Matrix.
- Derive a new transformation data set.

2.3 PCA based Modeling

Basic assumption of PCA based modeling for multivariate process is that in the measured space the data will be correlated. Therefore for effective detection and diagnosis, transform this measured data into lower dimensional space as explained in the section 2.2.2.

2.3.1 Data Modeling using PCA

Let the measured data $\mathbf{X}$ having p variables with n observations of each variable. From PCA theory we can express $\mathbf{X}$ into a set of new directions as below:

$$\begin{aligned}\mathbf{X} =& \mathbf{T}\mathbf{P}^T \\ &= \mathbf{t}_1\mathbf{p}_1{}^T + \mathbf{t}_2\mathbf{p}_2{}^T + ... + \mathbf{t}_p\mathbf{p}_p{}^T \\ &= \sum_{i=1}^{p} \mathbf{t}_i\mathbf{p}_i{}^T\end{aligned} \tag{2.1}$$

Where $\mathbf{p}_i$ is an Eigen vector matrix of the covariance matrix $\mathbf{X}$ is defined as Principle component(PC) loading matrix to identify the which of the variables contribute most to individual PC's, and $\mathbf{t}_i$ is the score matrix of PC's, provides information on sample clustering [6].

2.3.2 Selection of Number of PC's

From the fact that Most of variance can be seen through first few PC's, Selection of number of PC's can be done by the below following methods

- Cumulative Percent Variance(CPV) Method: CPV indicates the number of PC's to be selected such that required amount of variance can be captured in the PCA model.

$$CPV(m) = \frac{\sum\limits_{i=1}^{m} \lambda_i}{\sum\limits_{i=1}^{p} \lambda_i} \tag{2.2}$$

Let $CPV(m) \geq 90$ Percent , indicates variance is 90 percent for m number of eigen values, Therefore select number of eigen vectors in $\mathbf{P}$ corresponding to m eigen values.

Consider a process with size of 13 × 4 represent 4 variables with each variable having 13 observations. After performing PCA, Eigen value Matrix($\mathbf{\Lambda}$) and their corresponding eigen vector matrix($\mathbf{P}$) are shown.

$$\mathbf{\Lambda} = diag([\ 517.7969, 67.4964, 12.4054, 0.2372]) \tag{2.3}$$

$$\mathbf{P} = \begin{bmatrix} 0.0678 & -0.6460 & 0.5673 & 0.5062 \\ -0.6785 & -0.0200 & -0.5440 & 0.4933 \\ 0.0290 & 0.7553 & 0.4036 & 0.5156 \\ 0.7309 & -0.1085 & -0.4684 & 0.4844 \end{bmatrix} \tag{2.4}$$

$$CPV = \begin{bmatrix} 86.5974 \\ 97.8856 \\ 99.9603 \\ 100.0000 \end{bmatrix} \tag{2.5}$$

From equation 2.2, first two eigen values are contributing more than 90 percent of variance therefore we can select first two eigen vectors corresponding to first two eigen values. Now the Principle component matrix $\mathbf{P}_m$ for $m = 2$ no. of PC's or eigen vectors will be modified equation 2.4 to 2.6 is shown below.

$$\mathbf{P}_m = \begin{bmatrix} -0.0678 & -0.6460 \\ -0.6785 & -0.0200 \\ 0.0290 & 0.7553 \\ 0.7309 & -0.1085 \end{bmatrix} \tag{2.6}$$

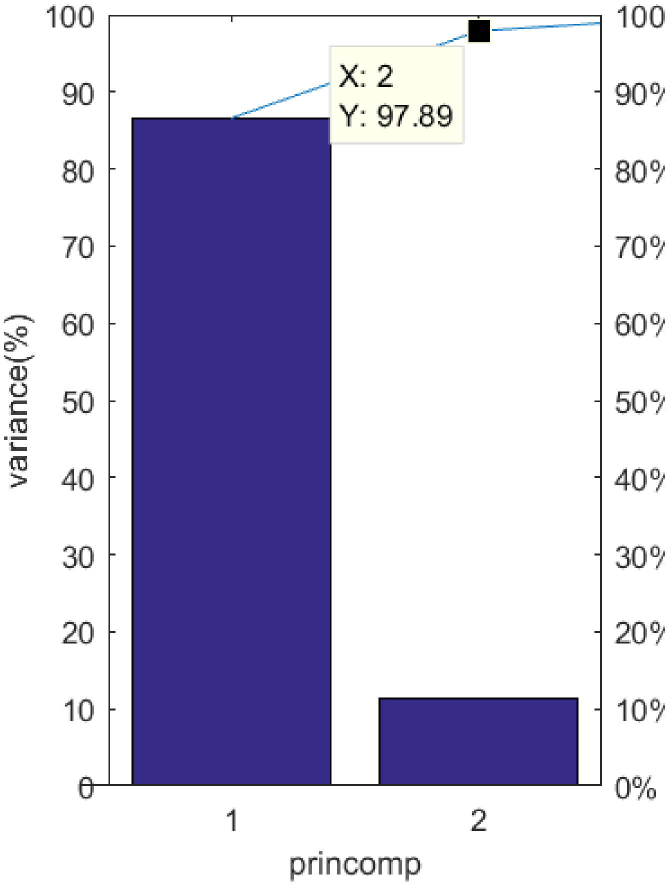


Figure 2.1: Scree plot

- Scree Test Procedure: Plot represents Eigen values explanation or variance in descending order and will shows the saturation after Knee point. No. of PC's to be selected are in between high component to knee point.

From the fig.2.1 after knee point there is a saturation i.e., amount of variance added by next Eigen value almost nil. therefore we can consider no. of PC's selected is $m = 2$.

After Selection of m number of PC's, $\mathbf{X}$ can be expressed as below

$$\begin{aligned}\mathbf{X} &= \mathbf{T}\mathbf{P}_m{}^T + \mathbf{E} \\ &= \sum_{i=1}^{m} \mathbf{t}_i \mathbf{p}_i{}^T + \mathbf{E}\end{aligned} \tag{2.7}$$

Where $\mathbf{P}$ containing m number of eigen vectors or PC's and $\mathbf{E}$ is residual matrix containing noise components expressed by remaining $p - m$ PC's. Some malfunctions may not influence first few PC's but it will have more effect on remaining PC's therefore, always considering residual term will improve diagnostic performance.

2.3.3 Indices for daignosing the Faults

In multivariate process plant while we are monitoring on-line process data to detect the abnormalities or faults we are going to use statistical testing for quick detection and diagnosis. For the fault detection and diagnosis purpose, T^2-statistic , Q-statistic will be used as detection indices's [7]. To approach to this testing we will first perform PCA as shown in section 2.3.1 after that we will find value of T^2 and SPE for each and every observations as shown equations 2.8 and 2.10 and it will compared to threshold values calculated using equations 2.9 and 2.11 respectively.

- Hotelling's T^2 statistic can be used to find the new measured variation from the modeled data. If variation explained by new latent variables (PC's) is greater than the already existed model then fault is detected. For New measurement data vector $\mathbf{x}$ variation can be found using below T^2 statistic expression.

  $$T^2 = \mathbf{x}^T \mathbf{P} (\mathbf{\Lambda}_m)^{-1} \mathbf{P}^T \mathbf{x} \tag{2.8}$$

  Where $\mathbf{\Lambda}_m$ is the Eigen value of principle component vector matrix, $\mathbf{P}_m$ is the loading vector matrix associated with m eigenvalues. The upper bound value for T^2 statistic can be found using Fischer Snedecor charts(F-Charts).

  $$T^2{}_{m,n,\alpha} = \frac{m(n-1)}{n-m} F(m, n-1, \alpha) \tag{2.9}$$

  Where m=no. of principle components considered, α is the level of significance. Calculated value of T^2 should not cross the threshold value which we will calculated from F-distribution charts.

- Another statistical method is using Q-statistic approach nothing but squared prediction error (SPE). It is the measure of goodness of fit of new sample to the model [6] detection follows as below equation 2.10 .

  $$SPE = \left\| (1 - \mathbf{P}_m \mathbf{P}_m{}^T) \mathbf{x} \right\|^2 \leq Q_\alpha \tag{2.10}$$

  Where Q_α can be found which will be the upper limit for detecting the fault as below

  $$Q_\alpha = \theta_1 \left(\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1{}^2} \right)^{1/h_0} \tag{2.11}$$

Where $\theta_i = \sum_{j=m+1}^{p} {\lambda_j}^i$; $h_0 = 1 - \frac{2\theta_1\theta_3}{3{\theta_2}^2}$ Where c_α normal distribution value with α significance level.

Detection of fault can be done effectively by using above Q-statistic and Hotellings statistic but to diagnosed the fault contribution charts will helps. Contribution charts will represent contribution of each variable to Principle component. The Q-statistic value for j^{th} variable to k^{th} sample can be found by using equation 3.16.

$$Q_{kj} = {e_{kj}}^2 = (\mathbf{x}_{kj} - \hat{\mathbf{x}}_{kj}) \tag{2.12}$$

By observing the contribution plots of process variables we can detect the which process variables to PC's exceeds limit can be found.

Chapter 3

Methods for Batch Process Monitoring

3.1 Introduction

The main objective of any process industry is to achieve high product quality and in-time demand reaching capability. These two capabilities can be fulfilled by batch processes. Using Batch processes high quality products will be produced with less quantity. In semiconductor manufacturing , pharmaceutical , specialty chemicals, food and beverages industries are using effectively. And also that monitoring for both product quality and for plant safety compare to continuous processes its bit difficult, due its dynamic operating point changes. So here are the some methods found in literature are explained briefly.

3.2 Multi Way PCA

Batch and fed-batch processes are typically monitored by using MPCA and MPLS(Multi Way Principle Least Squares). These popular methods are developed by Nomikos and MacGregor in [8]. MPCA method is quite same as PCA which is applied on unfolded data set of Batch array. Basically Batch data will be three dimensional array $\underline{\mathbf{X}}$ as shown in fig.3.1.

This 3-D array can be unfolded in three ways

- Batches×variables in each time interval(Unfolding in time wise)
- Batches×times for each variable(Unfolding variable-wise)
- Variables×time for each batch(Unfolding Batch wise)

Unfolding through Time wise will helps to analyze trajectories of samples, and unfolding through variable-wise will helps to analyze each batch wise variables. And Unfolding through Batch wise can be done keeping every layer $I \times J$ placed side by side along K axis will helps to summarize the variability information of variables along time among different batches.

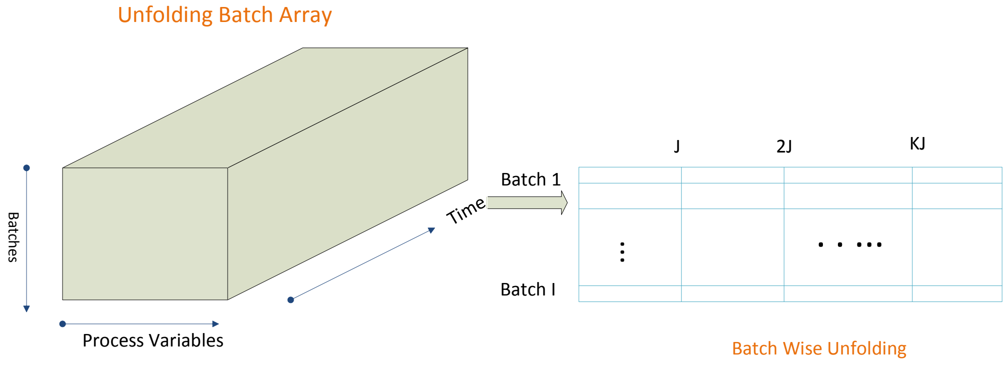


Figure 3.1: Batch-wise Unfolding

3.2.1 Variable Trajectory Estimation

In On-line monitoring of batch processes using MPCA method, we will have only few measurements from start to current time interval. So incomplete data in each batch will be estimated and filled for on-line monitoring. This can be done in three ways

- Zero deviation Unfilled remaining data points from current time to end of batch with zeros.
- Current deviation Unfilled data points will be filled with Current measurements
- PCA projection method

Suitable approach for on-line monitoring will be changed with respect to Batch process which we are monitoring. Most suitable approach is Current deviation method and projection method suggested by Nomikos and MacGregor [8]. While going for on-line batch process Monitoring using MPCA method, The performance will affected due to

- Estimation of future states or measurements.
- industrial noise and existing self-correlated information may contained in discarded PCs variance from process data.

3.2.2 Fault detection using Multi Way PCA

Unfolding Batch array $\underline{\mathbf{X}}(I \times J \times K)$ to 2-dimensional data set can be done in three ways as previously mentioned. Where I indicates number of batches, J indicates number variables, K indicates number of time intervals in each batch. Typically batch-wise unfolding results as $\mathbf{X}(I \times JK)$, will be used for MPCA. As shown in fig.3.1, on unfolded data PCA will be performed as mentioned in section 2.3.1.

After Performing PCA on matrix $\mathbf{X}(I \times JK)$ with JK variables and I samples results in

$$\begin{aligned}\mathbf{X} &= \mathbf{T}{\mathbf{P}_m}^T + \mathbf{E} \\ &= \sum_{i=1}^{m} \mathbf{t}_i{\mathbf{p}_i}^T + \mathbf{E}\end{aligned} \tag{3.1}$$

Where $\mathbf{P}_m(JK \times m)$ is the principal component(PC) loading matrix and PC score matrix $\mathbf{T}_m(I \times m)$ and residual matrix is $\mathbf{E}(I \times JK).m$ is number of PC's selected using cumulative percent variance method as done in section 2.2. And PC's can be obtained from singular value decomposition(SVD) of covariance matrix $\mathbf{S}$.

$$\begin{aligned}\mathbf{S} &= \frac{\mathbf{X^T X}}{I-1} \\ &= \mathbf{P \Lambda P}^T\end{aligned} \tag{3.2}$$

where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_{KJ})$ is the diagonal matrix containing eigen values and their corresponding eigen vectors are ordered in decreasing order in $\mathbf{P}$. $\mathbf{P}$ can be separated into two parts one is PC's space $\mathbf{P}_m(JK \times m)$ and another one is $\mathbf{P}_m(JK \times (JK - m))$ residual vector space.

3.2.3 Process monitoring Indices

Hotellings T^2 and SPE are test statistics for monitoring process in the direction of PC's and residual space respectively. Calculation of statistics can be done as below.

$$\begin{aligned}&T^2 = \mathbf{x}_B \mathbf{P}_m \mathbf{\Lambda}^{-1} {\mathbf{P}_m}^T {\mathbf{x}_B}^T \\ &SPE = \mathbf{e}\mathbf{e}^T \\ &\mathbf{e} = \mathbf{e}_B(1, (k-1)J : kJ) \\ &\mathbf{e}_B = \mathbf{x}_B - \mathbf{x}_B \mathbf{P}_m {\mathbf{P}_m}^T\end{aligned} \tag{3.3}$$

where $\mathbf{\Lambda}_m$ is diagonal matrix containing first m eigen values similarly $\mathbf{P}_m$ contains corresponding m eigen vectors and $\mathbf{x}_B(1 \times JK)$ is newly monitored sample unfolded and normalized accordingly. e indicates the residual vectors at time instant k with J number of variables.

Process Control limits can be found as below:

- Hotelling's T^2 limit

  $$T^2 \leq \frac{m(I^2-1)}{I(I-m)} F_{m,I-m,\alpha} \tag{3.4}$$

  Where $F_{m,I-m,\alpha}$ is an F distribution probability values corresponding to m,$I-m$ degrees of freedom with an confidence limit α.

- Squared Prediction Error limit

  $$SPE_\alpha \leq \frac{v}{2w} {\chi^2}_{\frac{2w^2}{v},\alpha} \tag{3.5}$$

  In the above equation w and v are mean and variance respectively for calculated SPE values for modeling stage. ${\chi^2}_{\frac{2w^2}{v},\alpha}$ means χ^2 probability value at $\frac{2w^2}{v}$ with α confidence limit. Since calculated SPE values for modeled batches for each time interval follows χ^2 distribution.

3.3 Dissimilarity Factor Analysis to Batch Process

Dissimilarity factor analysis (DISSIM)(developed by kano et al.,) is popularly used for continuous process data monitoring since this method can effectively detect the changes in observed variables correlations. It helps to detects the change in operating point effectively in continuous processes, provides quantitative value to changes in operating point. Since operating point changes are more steady in continuous processes but it will change more randomly in Batch processes. Batch-Monitoring using DISSIM concept will give highly fluctuated values not follows any distribution and it will be tough to obtain control limits for monitoring. Extension of DISSIM method to Batch processes is termed as EDISSIM method [9]. In this method used variable moving window strategy on each batch. Below are the differences between DISSIM analysis to EDISSIM(Extension to DISSIM)

- DISSIM analysis will be helpful for monitoring continuous processes, EDISSIM for Batch processes.

- EDISSIM will helps to find batch-to-batch variation of process trajectories during same time interval. Where as DISSIM analysis will helps to find operating point changes along time axis.

- Multiple reference models required in each batch with EDISSIM analysis whereas in DISSIM technique only one reference model exists.

- Calculated DF(Dissimilarity factor values) will follow Gamma-distribution in EDISSIM. whereas in DISSIM method no control limits will be derived.

3.3.1 Dissimilarity Index Based Monitoring

The Karhunen-Loeve(KL) expansion technique is well known for feature extraction and dimensionality reduction. Here this method is used to obtain changes in processes data distributions. Since distribution represents operating point variatons with respect to time, DF will capture the correlations among variables and will detect the variations.

3.3.2 Calculation of Dissimilarity Factors

Consider two data sets X_1 and X_2 having K_1 and K_2 samples respectively and their resultant covariance structure S will be as below:

$$\begin{gathered} S = \frac{K_1}{K_1 + K_2} S_1 + \frac{K_2}{K_1 + K_2} S_2 \\ Where\ S_i = \frac{1}{K_i} {X_i}^T X_i\ for\ i = 1, 2 \end{gathered} \tag{3.6}$$

After diagonalizing 3.6 will results in diagonal matrix $\mathbf{\Lambda}$ and orthogonal matrix P_0

$${P_0}^T R P_0 = \mathbf{\Lambda} \tag{3.7}$$

Now the original data matrices are transformed into Y_i using transformation matrix P.

$$\begin{aligned} Y_i &= \sqrt{\frac{K_i}{K_1+K_2}} X_i P_0 \Lambda^{-1/2} \\ &= \sqrt{\frac{K_i}{K_1+K_2}} X_i P \\ Where\ P &= P_0 \mathbf{\Lambda}^{-1/2} and\ i = 1,2 \end{aligned} \tag{3.8}$$

Covariance matrices of transformed data set will be shown below and it will satisfies the equation 3.10

$$\begin{aligned} S_i &= \frac{K_i}{K_1+K_2} P^T R_i P \\ where\ i &= 1,2 \end{aligned} \tag{3.9}$$

$$S_1 + S_2 = I \tag{3.10}$$

After eigen value decomposition of covariance matrices from equation 3.9 will yields eigen vector w_i and corresponding to eigen value λ_i where superscript j indicates jth eigen value or eigen vector.

$$S_i {w_i}^j = {\lambda_i}^j {w_i}^j \tag{3.11}$$

From equation 3.10 and 3.11 we can written as below

$$\begin{aligned} &S_2 {w_1}^j = (1 - {\lambda_1}^j) {w_1}^j \\ &\Rightarrow\ 1 - {\lambda_1}^j = {\lambda_2}^j \end{aligned} \tag{3.12}$$

Therefore now the two datasets are having same PC's but reversely ordered as most important PC for data set one will be least prioritized PC for data set two,vice versa. Below is the formula for dissimilarity factor calculations between two data sets.

$$D = diss(X_1, X_2) = \frac{4}{J} \sum_{j=1}^{J} (\lambda_{_j} - 0.5)^2 \tag{3.13}$$

Where J denotes number of variables in process and λ_j will be eigenvalues of covariance matrix S.

$$\begin{aligned} For\ \lambda_j &\simeq 0.5 \quad \Rightarrow D = 0\ (Data\ set\ is\ similar) \\ \lambda_j &\simeq 0 \quad\ \ \Rightarrow\ D = 1\ (process\ data\ set\ is\ different) \\ So\ D &\in [0,1] \end{aligned} \tag{3.14}$$

3.3.3 Moving Window Technique for Batch data

In moving window method, typically window size will be fixed for entire batch. But due to fixed length window size initial process data will not contain that much information. Observing and calculating monitoring indices's will not be good idea for initial data. Since the calculated dissimilarity factors(DF) will be fluctuated and inference from factors is nil. And it will increase diagnostic delays will not be entertained for good FDD strategies. Therefore to avoid this confusion choose the big window size during initial times and after that make it small when the process reaches to one by fourth part. Still selection of this window size will be very ambiguous and it will be decided by experts in that particular plant. Below will be block diagram representation. Here L will be initial window size and K will be no. of windows in the batch. For each window Dissimilarity factors will be calculated and value will be checked with predefined limits calculated from model.

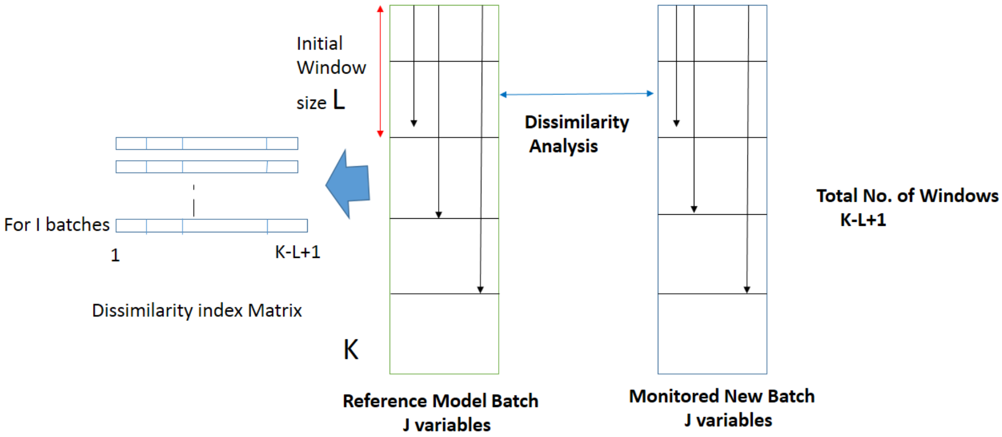


Figure 3.2: Variable Moving Window

Calculated DF values will not follow typical normal distribution, after observing the DF values for calculated windows follows follows γ distribution [9]. And to find the cause for fault operation contribution plots are analyzed.

3.3.4 Fault Contribution Variable

Dissimilarity factor values are calculated from eigen values of transformed covariance matrix as shown in equation3.13. So for calculation of most contributed variable, inverse transform needed and find the norm of that score matrix will be found as below.

$$\begin{aligned}
&\mathbf{Y}_i = \mathbf{X}_i\mathbf{A} \\
&Where\ A = \sqrt{\frac{N_i}{N_1 + N_2}\mathbf{P}} \\
&Score\ vector\ u\sin g\ \mathbf{Y}_i \\
&\Rightarrow \mathbf{t^T}{=}\mathbf{Y_i}\mathbf{w}_i{}^T \\
&\qquad = \sum_{j=1}^{J} \mathbf{x}_j(\mathbf{A}\mathbf{w}_i{}')_j
\end{aligned} \tag{3.15}$$

Here $\mathbf{x}_j$ represents j^{th} column vector of $\mathbf{X}_i$ and $(Aw_i{}^T)$ will represents j^{th} element of $(Aw_i{}^T)_j$ and for determining most contributed variable to score vector can calculated by finding norm of score vector as shown below.

$$C_j{}^{[D]} = \left\|\mathbf{x}_j(\mathbf{A}\mathbf{w}_i{}^T)\right\|_j \tag{3.16}$$

Above equation 3.16 will give the contribution of variable j to the fault.

3.3.5 Flow chart for DF method to batch process

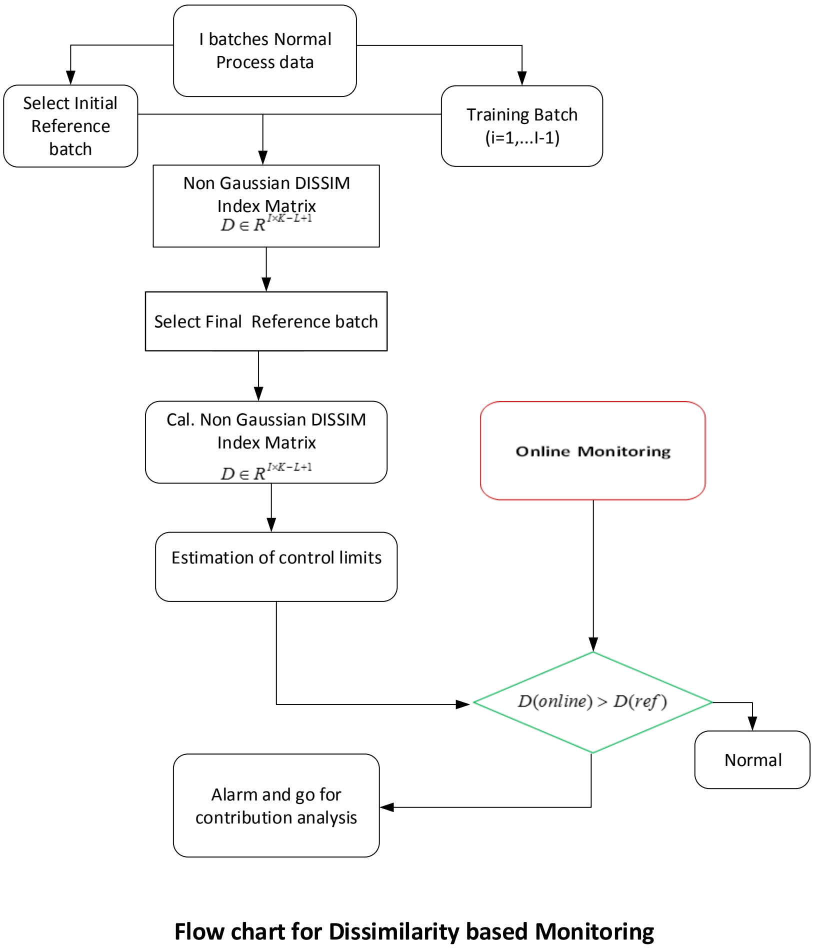

I batches Normal Process data
Select Initial Reference batch
Training Batch (i=1,...I-1)
Non Gaussian DISSIM Index Matrix $D \in R^{I \times K-L+1}$
Select Final Reference batch
Cal. Non Gaussian DISSIM Index Matrix $D \in R^{I \times K-L+1}$
Estimation of control limits
Online Monitoring
$D(online) > D(ref)$
Normal
Alarm and go for contribution analysis
Flow chart for Dissimilarity based Monitoring

Figure 3.3: Flow chart for DF based Method

Above flowchart explains step by step approach to EDISSIM method based monitoring to batch process. In Historical data base there will be I number of batches among that select one batch as reference. Choose initial window size L appropriately. Now calculate Dissimilarity factors(DF) for each window as shown in figure3.3.3. Calculated Non gaussian DFs will be having size $I \times (K - L + 1)$. In the above Index matrix find average of all the DFs and choose final reference batch. Now again repeat to calculate new dissimilarity factors and estimate control limits assuming gamma distribution. When new batch is monitoring compare calculated Dissimilarity factors with reference value. If new DF exceeds reference value go for contribution analysis which will give fault contributed eigen value, otherwise normal operation.

3.4 Pattern matching method

Main motivation behind this method is to locate small portion of monitored samples in a historical data base which will helps to identify faulty batches quickly. This can be done firstly by selecting candidate pool and process expert or experienced operator can evaluate exact patterns. The process of doing step by step method can be done as shown in block diagram 3.4. Initially Pattern matching methods application to process monitoring first introduced by Johannesmeyer et al [10]. He proposed similarity factors based on PCA technique(Refer to section2.2.2). According to PCA multivariate statistical technique most variability directions will be given by principal components(Eigen vectors) corresponding to high Eigen values.

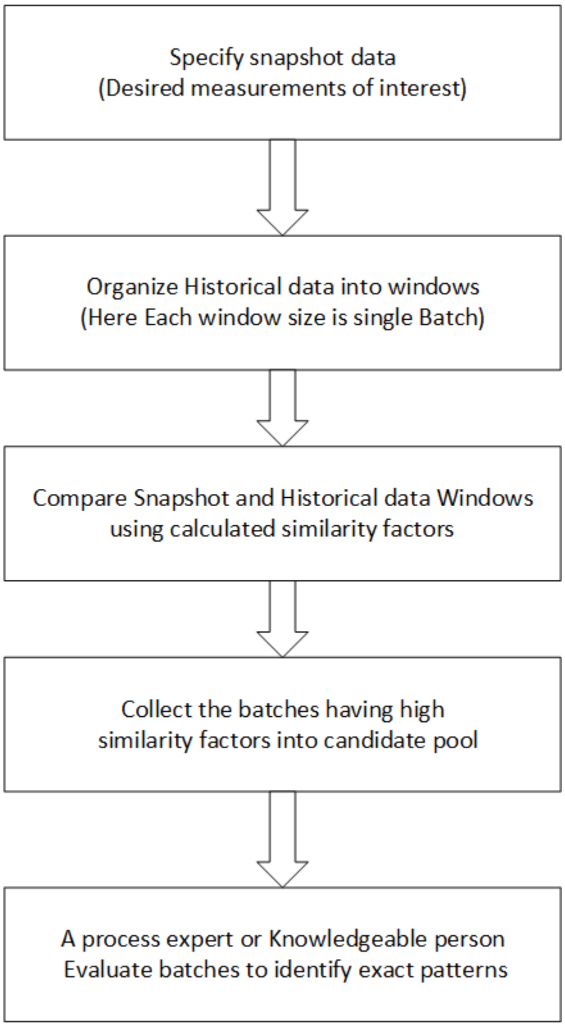

Specify snapshot data
(Desired measurements of interest)
Organize Historical data into windows
(Here Each window size is single Batch)
Compare Snapshot and Historical data Windows
using calculated similarity factors
Collect the batches having high
similarity factors into candidate pool
A process expert or Knowledgeable person
Evaluate batches to identify exact patterns

Figure 3.4: Flow diagram Pattern matching method

3.4.1 Similarity factors

Measure of Similarity between two data sets first introduced by Krzanowski(1979) using PCA technique. According to this method data sets having same number of variables and irrespective of their number of measurements it will give quantitative measure of similarity between two data sets. Consider two data sets having same n variables and having k number of principle components selection can be done by cumulative percent variance method2.2. And quantifying similarity between two data sets will be calculated using below expression.

$$S_{pca} = \frac{trace(L^T M M^T L)}{k} \tag{3.17}$$

In the above equation 3.17, L and M are having size $n \times k$ which are obtained from the snapshot data set S and historical data set H respectively. And their geometrical interpretation is given by sum of the cosine angle between each principle component vectors from snap shot data set to historical data set and weighting with inverse of number of PC's selected as shown below.

$$S_{PCA} = \frac{1}{k} \sum_{i=1}^{k} \sum_{i=1}^{k} \cos^2 \theta_{ij} \tag{3.18}$$

Instead of keeping same weight for each similarity factor as shown in equation3.17 its better to keep choosing weights according to their variability explained. E.seborg and Ashish Singhal(2001) proposed modification for equation 3.17 as shown below.

$$\begin{aligned} S^{\lambda}{}_{PCA} &= \frac{trace(R^T T T^T R)}{\sum_{i=1}^{k} {\lambda_i}^l {\lambda_j}^m} \\ where\ R &= L\Lambda_l \\ T &= M\Lambda_m \\ \Lambda &= diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_k}) \end{aligned} \tag{3.19}$$

${\lambda_i}^l$ and ${\lambda_i}^m$ are eigen values corresponding to the i^{th} PC of L and M respectively with decreasing order. Due to summation term in the denominator calculated similarity factors will be always less than one. Below is the geometrical interpretation of equation 3.19.

$$S^{\lambda}{}_{PCA} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} ({\lambda_i}^l {\lambda_j}^m) \cos^2 \theta_{ij}}{\sum_{i=1}^{k} {\lambda_i}^l {\lambda_j}^m} \tag{3.20}$$

Here θ_{ij} is angle between i^{th} and j^{th} principle component of L and M.

3.4.2 Distance similarity factors

This similarity factor are needed because, sometimes process data sets will have same principal components but distance between them is far apart. Mahalnobis distance(Θ) is the measure of distance between two data sets. In this context it will give the distance measure between historical data set $\mathbf{X}_H$ to snapshot data set $\mathbf{X}_S$. Mahalnobis distance calculations of above data sets are given as below subsequent equation.

$$\Theta = \sqrt{(\overline{\mathbf{x}}_H - \overline{\mathbf{x}}_S)^T \mathbf{\Sigma}_s^{*-1} (\overline{\mathbf{x}}_H - \overline{\mathbf{x}}_S)} \tag{3.21}$$

where $\overline{\mathbf{x}}_H$ and $\overline{\mathbf{x}}_S$ are mean vectors of historical and snapshot data sets respectively. $\mathbf{\Sigma}_s^{*-1}$ is the pseudo inverse of singular matrix containing selected number of PC's. And distance similarity factor is given by the probability of least possible distance between $\overline{\mathbf{x}}_H$ and $\overline{\mathbf{x}}_S$.

$$S_{dist} = 2 \times \frac{1}{\sqrt{2\pi}} \int_{\Theta}^{-\propto} e^{-z^2/2} dz \tag{3.22}$$

3.4.3 Validation measures

Here are the some validation metrics one is pool effeciency, another one is pattern matching effeciency proposed by Johannesmeyer and seborg(1999) [10]. Effectiveness of this method depends on these two effeciencies.

- Pool effeciency will represents with how much accuracy we can able to find monitored pattern in candidate pool N_p.

$$\begin{aligned} p &= \frac{N_1}{N_p} \times 100 \qquad\qquad \text{Where } N_P = N_1 + N_2 \\ N_1 &= \text{ Successful records} \\ N_2 &= \text{ Unsucessful records} \end{aligned} \tag{3.23}$$

- Pattern matching effeciency will represent the how effectively we can able to locate the similar patterns in historical data base. And also it is difficult to locate all patterns in historical data base which are actually similar to snap shot patterns is difficult ($N_{DB} \neq N_1$).

$$\eta = \frac{N_1}{N_{DB}} \times 100 \quad \text{where } N_{DB} = \ Actual\ no.\ of\ patterns\ similar\ \ to \text{ Historical data base} \tag{3.24}$$

- If the pool size is very less ($N_P = N_{DB}$) then the pattern matching efficiency is also less. Maximum possible theoretical effeciency is as shown below. If ($N_P = N_{DB}$) max possible efficiency is 100 percent.

$$\eta_{\max} = \frac{N_P}{N_{DB}} \times 100 \tag{3.25}$$

As shown in block diagram 3.4 after selection of candidate pool N_p process operator or expert will evaluate those batches N_1 which are having high similarity factor(nearer to one). This evaluation measure is pool efficiency calculated as in equation 3.23. Pattern matching efficiency η calculated as shown in equation 3.24.

3.4.4 Remarks

- MPCA method: Major False alarms and Identifying the sever of deviations can be done using MPCA with SPE and Hotelling test statistic effectively. But it fails to identify the incepient faults or slowly gradually incremental false which will degrades product quality. MPCA method will increases size of dataset so computationally lengthy process. So quick decision making cannot be done while monitoring online.

- Dissimilarity method using moving window approach: Variable time moving windows which are introduced to evaluate batch data sets using quantitative measure called EDISSIM. This method is effective for batch process monitoring but selecting control limits based on distribution of EDISSIM values for each operating condition is difficult.

- Pattern Matching Approach Using PCA : PCA can handle linear correlations among data so effective dimensionality will be decreased. Quick Identification of snapshot data in historical batch data can be done effectively without much computational efforts. But Candidate pool selection and evaluating based on operator experience may mislead decision making.

Chapter 4

Case study using Pattern matching

In this chapter Pattern matching strategy to Butanol fermentation batch process data is discussed. Main concentration here is to get better solution without candidate pool selection. Selection of candidate pool and their evaluation will take much time. Instead of that, pattern evaluation in decreasing order of their similarity factors will helps to find few patterns in historical data base. Both on-line and off-line monitoring strategies had been discussed for Butanol fermentation batch process data.

4.1 Case study:Batch Acetone-Butanol fermentation process

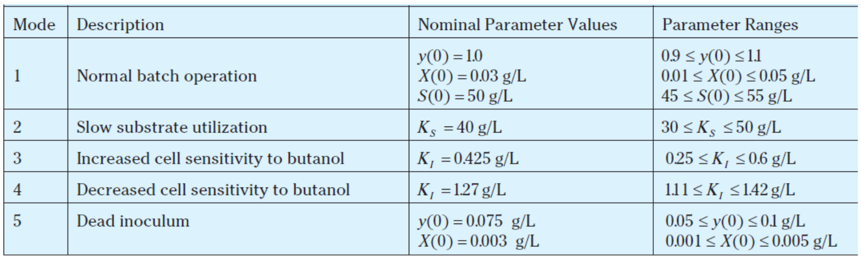
Batch Acetone-Butanol fermentation process produces acetone, ethanol and butanol as end products. Simulation study is performed on physical model of this process. Model consists of ten non-linear ordinary differential equations having ten measurements. Detailed description of this process and mathematical modeling discussed in Vortruba et al [11]. Through simulations extensive generation of data is done. This data used as Historical data base. In table 4.1 list of measured variables and parameters and their sampling time is shown.

Table 4.1: Model Variables and parameters of the Above Batch process

Variable/parameter	Description	Sampling Period
y	Dimensionless cellular RNA concentration	Not measured
X	Reactor cell concentration	30 min
S	Reactor substrate concentration	1 min
BA	Reactor butyric acid concentration	1 min
AA	Reactor acetic acid concentration	1 min
B	Reactor butanol concentration	1 min
A	Reactor acetone concentration	1 min
E	Reactor ethanol concentration	1 min
CO_2	CO_2 concentration	1 min
H_2	H_2 concentration	1 min
K_s	Substrate uptake saturation constant	Not available
K_I	Butanol inhibition constant	Not available

4.2 Historical data base and Data preprocessing

Historical data base is developed for different operating conditions. Model parameters and initial values are changed for every batch. Historical data base will contain both normal operating conditions(NOC) and abnormal operating conditions. In abnormal batches key parameter varied randomly from batch to batch. Time span of each batch is 30 hours and sampling time is one minute. Gaussian measurement noise is added to measurements with signal to noise ratio is approximately ten to mimic with practical industry data. Below table 4.2 provides nominal values for snap shot data and different operating conditions. One of the measurement out of nine measured values will be measured for every 30 minutes. So to make it uniform sampling period to all measured variables, using linear interpolation method intermediate data is generated with 1 minute sampling time. Entire historical data base will contain 9 lakh measurements for each process variable. Total number of batches are 500. For each operating condition simulated for 100 times with initial values shown in figure 4.1.



Mode	Description	Nominal Parameter Values	Parameter Ranges
1	Normal batch operation	$y(0) = 1.0$ $X(0) = 0.03$ g/L $S(0) = 50$ g/L	$0.9 \le y(0) \le 1.1$ $0.01 \le X(0) \le 0.05$ g/L $45 \le S(0) \le 55$ g/L
2	Slow substrate utilization	$K_S = 40$ g/L	$30 \le K_S \le 50$ g/L
3	Increased cell sensitivity to butanol	$K_I = 0.425$ g/L	$0.25 \le K_I \le 0.6$ g/L
4	Decreased cell sensitivity to butanol	$K_I = 1.27$ g/L	$1.11 \le K_I \le 1.42$ g/L
5	Dead inoculum	$y(0) = 0.075$ g/L $X(0) = 0.003$ g/L	$0.05 \le y(0) \le 0.1$ g/L $0.001 \le X(0) \le 0.005$ g/L

Figure 4.1: Different Operating modes for Acetone-Butonal process(Adapted from [12])

4.3 Results and Discussions

Generated snapshot data set will be searched in Historical data base to identify similar patterns. Which will helps to find abnormal situations. If the process under out of control then the snapshot data will be more similar to the batches where the abnormal operating condition available in historical data base. This will helps to take preventive action against faults. Developed snap shot data with normal operation condition nominal values, will be more similar to operating condition 1. Below graph represents the similarity factors(calculated using eqn. 3.17 and 3.19) which are nearer to one for both the factors and almost all the blue dots will replicate normal operating condition.

And for another operating condition if we simulate plant with operating condition 5 as shown in figure 4.2. Below figure 4.3.1 showing that batches having high similarity belongs to operating condition 5.

Off-line monitoring using this technique effectively detect the patterns in historical data base depends on operating condition. To avoid the confusion to select the patterns from both the similarity factors, here its the combined similarity factor $S_{comb} = {S_{pca}}^{\lambda} + (1 - \alpha)S_{pca}$ where α varies from 0 to 1. Here ${S_{pca}}^{\lambda}$ will have more weightage since its more effective similarity factor compare to other. Take typical value of α is 0.6. Below figure 4.3 represents similarity factor is high for below 100 batches belongs to Normal operating condition.

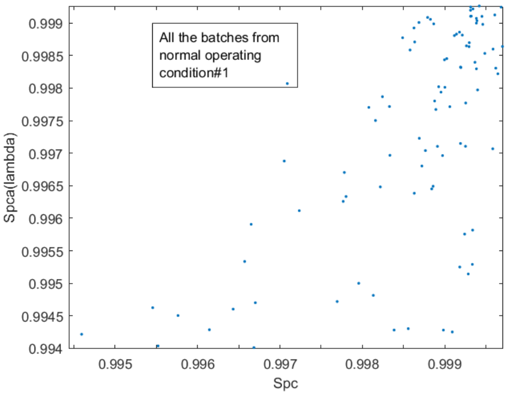


Figure 4.2: Batches having high similarity between snapshot data to operating condition 1

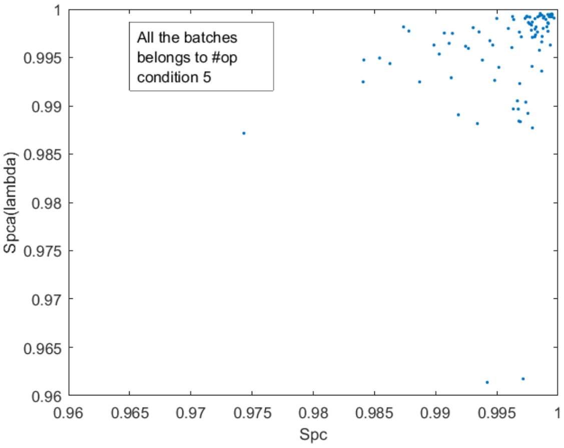


Figure 4.3: Batches having high similarity between snapshot data to operating condition 5

4.3.1 On-line monitoring

In on-line monitoring method we will not have data points from current instant to end of the batch. So for monitoring we need entire batch data. In literature there are three solutions popularly available as mentioned in section 3.2. Zero filling method is not good for this particular process monitoring. This method will misguide to take fault instruction. For this particular monitoring method we filled with previously available data (if the previous set of data resembles Normal operation data).

Above displayed figure 4.5 will be generated as below:

- After getting each sample in snapshot data, remaining data points will be anticipated with previous data.
- Now this entire batch is used for searching similar patterns in historical data set.
- For each time (sample) 500 similarity factors calculated and sorted in decreasing order. Selected only 10 out of hundred which are having high values
- When the process is significantly active (after 3 hours in this case) similarity factors will be high for corresponding operating conditions.

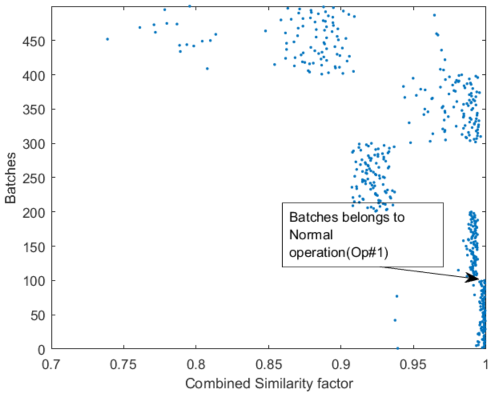


Figure 4.4: Batches having high similarity between snapshot data to operating condition 1

Since we simulated for snapshot data with nominal values, figure 4.5 showing that most of the similarity factors will fall under first 100 batches which are normal operating condition(NOC) batches. Below figure is for samples 1000 to 1800 and each sample will be sorted to 10 samples, like that nearly with 8000 similarity factors plotted with respect to 500 batches.

Obtained faulty snapshot data using operating condition 5 parameter ranged values is effectively matching in historical data base. Above figure indicates operating condition 5 which is faulty condition. Selected top ten similarity factors which are very close to one for each sample(During On-line monitoring) in historical data base is plotted.

- In this document using similarity concept made an attempt to apply for on-line monitoring. Choosing initial sampling instant and the method used for anticipating unfilled data will effect the monitoring.

- Online monitoring performance mainly depends on method of un filling the data, and initial window where operation will be in most active mode(peak operation). In this case nearly above 1000 samples will give right status about the fault. So nearly 16 hours it will take to identify the operation condition.

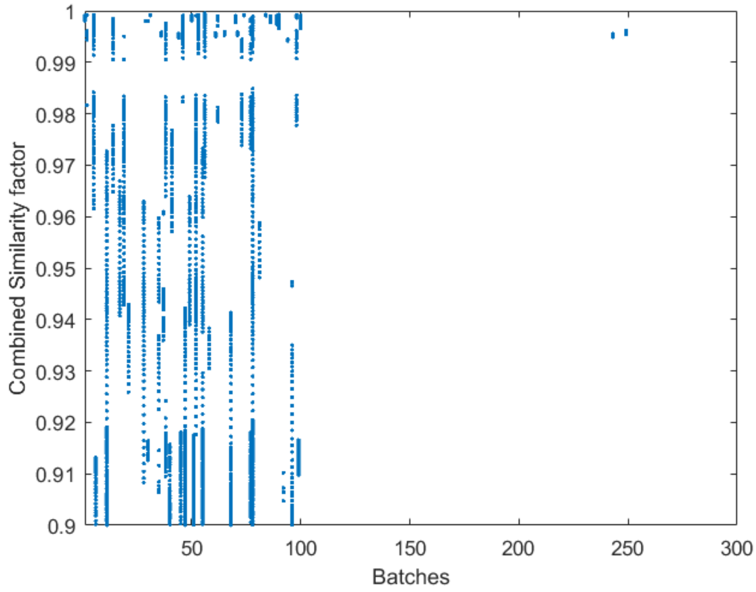


Figure 4.5: Batches having high similarity between snapshot data to operating condition 1

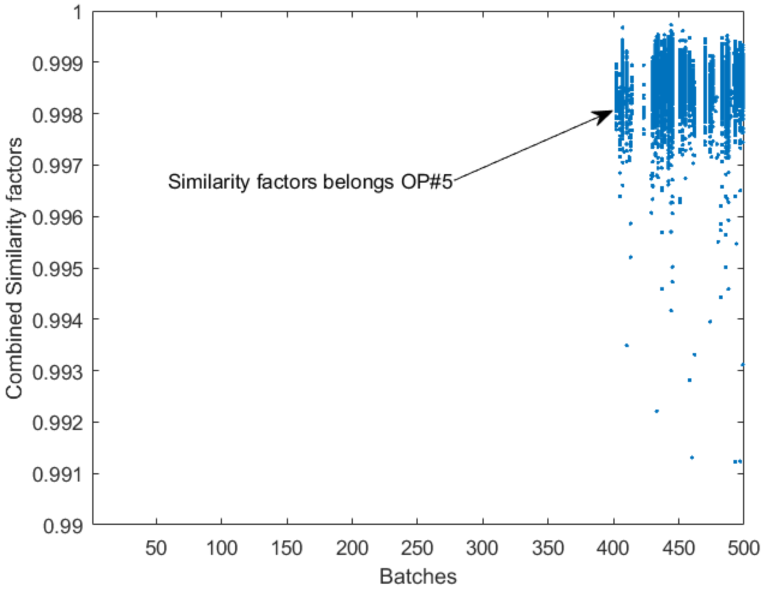


Figure 4.6: Batches having high similarity between snapshot data to operating condition 5

Chapter 5

Average PCs Method

In a Process plant there will be huge data, which will be collected for every sampled interval through sensors. This wealthy information should be well processed for taking diagnostic actions on the plant. For this purpose one of the Multivariate statistical technique called Principle component analysis(PCA) will helps to find the most significant information in the huge dimensional data. So dimensionality greatly reduced which ultimately increases diagnostic capabilities of a process plant. Insight to number of PC's retained can be decided by Cumulative percent variance(CPV), Cross validation, Scree test methods.

5.1 General similarity factor approach

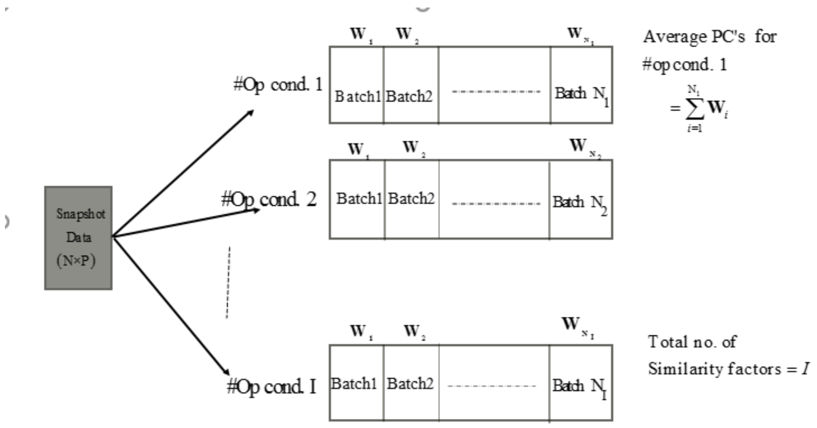
Basically this approach is belongs to process history based methods since it needs huge historical data set priorly. Pattern matching approach to batch process using similarity factor method done by Ashish singhal and E.Seborg(2001). This pattern matching is done between snapshot data set to Historical data set using Similarity factors. To identify the exact patterns of snapshot data in historical data batches, need to check the similarity factor values which are nearer to one(i.e. approximately more than 0.95) in between the batches. And operator will take the decision based on those batches having high similarity factors.

In every batch process there will be different operating conditions and Every operating condition duration will be different to each other. Suppose for every operating conditions there will be N_i number of similarity factors are calculated. Similarly for I number of operating conditions total number of similarity factors are $\sum_{i=1}^{I} N_i$. So to take decision based on similarity factors need to make it into descending order. And operator will evaluate those batches having highest similarity factors and take the diagnostic decisions.

5.2 Average PC's Model Approach to find Similarity factors

As mentioned earlier similarity between two data sets can be determined by angle between their main axises(Principle axises). In above section 3.10 equations for measuring quantified value for similarity between two data sets are mentioned. In the literature of pattern matching approach to batch process monitoring huge number of Similarity factor calculations are needed between snapshot data to each batch in every operating condition in historical data set [12]. Calculating similarity nothing but angle measure between the main axis of two datasets. So instead of calculating angle from snapshot Principle axises to each and every Batch Principle axises in each operating condition in historical data sets its better to calculate angle between snapshot dataset Principle axises to average of all batches PC's for each operating condition. Which will drastically decreases the efforts of operator to take diagnostic actions without any ambiguity. And also computational efforts greatly decreases so that quickly locating monitored patterns in historical batches is possible for various purposes like plant safety, maintenance etc.

As shown in figure 5.1 for every operating condition batches Principle components are calculated using singular value decomposition or Eigen value decomposition method and average of those $\frac{1}{N_1}\sum_{i=1}^{N_1}\mathbf{W}_i$ utilized in the model. So number of similarity factors calculation needed is only I equal to number of different operating conditions.Operating condition PC's which is giving highest similarity factor value(nearer to one) to snapshot data PC's will be chosen. Now that snapshot will be classified as that particular operating condition data set.



$\mathbf{W}_1$ $\mathbf{W}_2$ $\mathbf{W}_{N_1}$
Average PC's for #op cond. 1
$=\sum_{i=1}^{N_1}\mathbf{W}_i$
#Op cond. 1
Batch1 Batch2 ------------ Batch N1
#Op cond. 2
Batch1 Batch2 ------------ Batch N2
Snapshot Data (N×P)
#Op cond. I
Batch1 Batch2 ------------ Batch NI
Total no. of Similarity factors = I

Figure 5.1: Block diagram Representation of Average PC's Model

5.2.1 Flow Chart

In figure 5.2 flow chart for average model method. After calculating PC's for different operating conditions using training batches new batch or snapshot will be projected on this average model. After calculating similarity factors decide whether this new batch belongs to which operating condition(already existed) or completely different faulty operating condition batch(similarity is very less for new fault). So this method useful for model updating and incremental model building ultimately increases the efficiency of the model. When the similarity factor value is greater than the 0.9 value between new batch to existed operating condition batches considered as already existed operating condition batch. And this new batch is belongs to operating condition which is giving high value(nearer to one).This new batch utilized for incremental model building. If the similarity factor values are less to all the operating conditions PC's then new batch considered as new faulty operating batch which utilized for model updating.

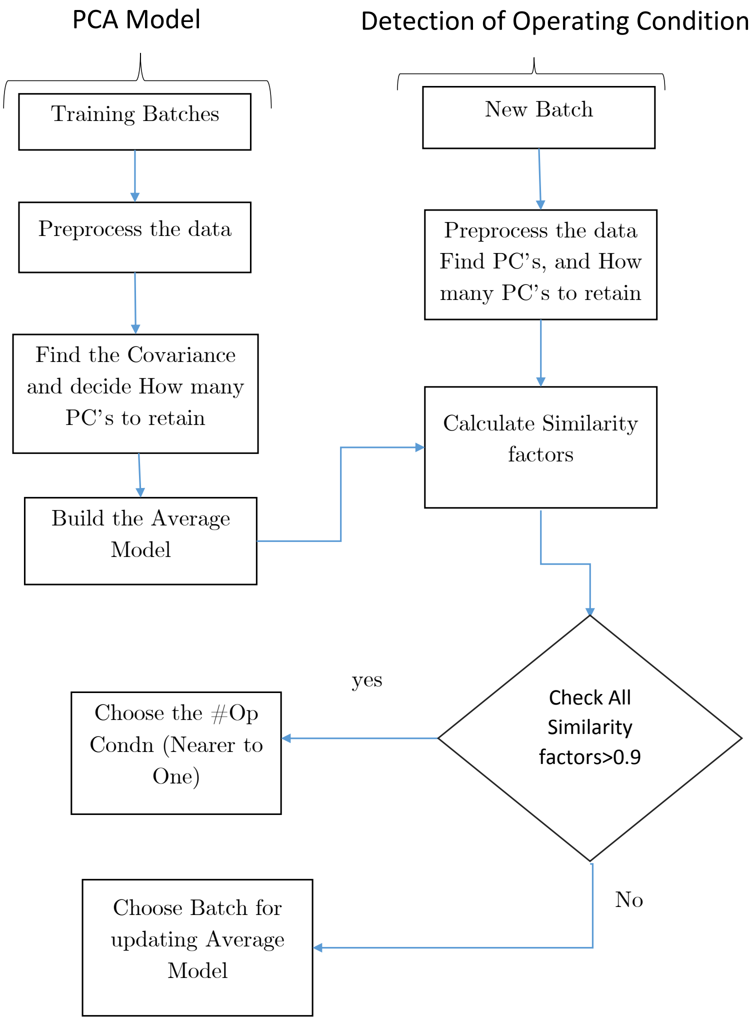

PCA Model
Detection of Operating Condition
Training Batches
Preprocess the data
Find the Covariance and decide How many PC's to retain
Build the Average Model
New Batch
Preprocess the data Find PC's, and How many PC's to retain
Calculate Similarity factors
Check All Similarity factors>0.9
yes
Choose the #Op Condn (Nearer to One)
No
Choose Batch for updating Average Model

Figure 5.2: Flow diagram of Average PC's model method

5.3 Results of case studies

In every batch process there will be different operating conditions. Each operating condition duration may or may not be same. In historical data set for every operating condition there will be different batches of data collected multiple times when process plant under running condition. This historical data used for future reference to quickly locate snapshot patterns. To verify this method Batch Acetone Butanol Fermentation Process used as case study. Mathematical model of this process simulation has been done to get amount of batch data for different operating conditions used for develop the model and to test the model.

- Total 112 batches used as training data set to develop model from different operating conditions(56 Normal,14 from four faulty operating condition batches).
- From figure 5.2 to 5.6 are results for each Monitored snapshot data corresponding operating condition.

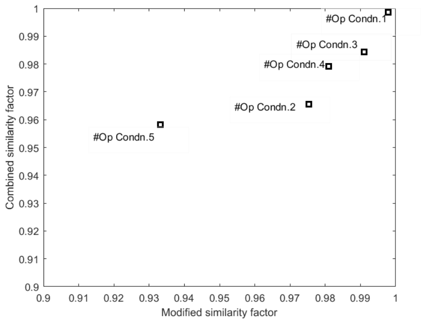


Figure 5.3: Monitored Snapshot data similar with Normal batch operating condition

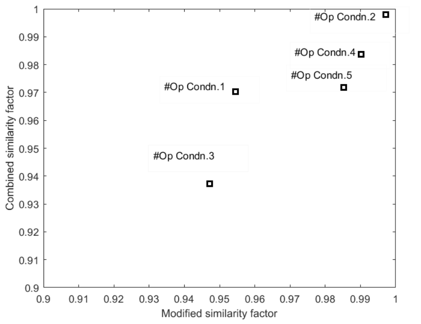


Figure 5.4: Monitored Snapshot data similar with Slow substrate operating condition

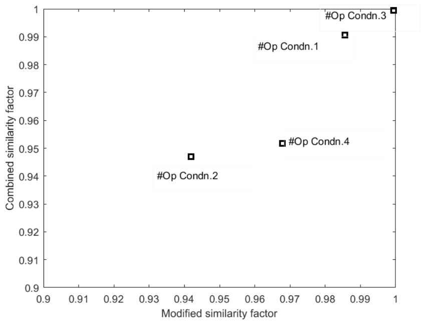


Figure 5.5: Monitored Snapshot data similar with Increased Cell sensitivity to Butanol operating condition

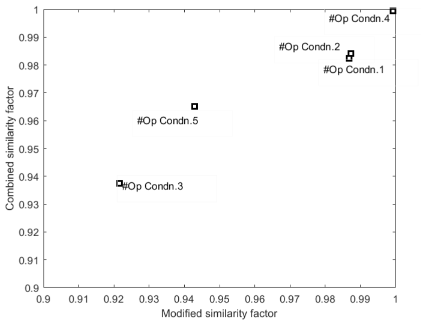


Figure 5.6: Monitored Snapshot data similar with Decreased Cell sensitivity to Butanol operating condition

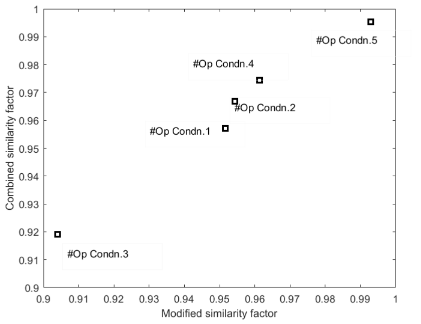


Figure 5.7: Monitored Snapshot data similar with Dead inoculum operating condition

- In Table 5.2 performance of average model method for 18 test batches(6 Normal batches, 3 from each faulty operating condition) are shown. Symbol cross indicates in table 5.2, simulated snapshot data fall under that particular operating condition. Mismatches indicated with 'O' symbol.

Table 5.1: Modified similarity factor Values for 18 Test batches from different operating Conditions

#OP Cond	N	N	N	N	N	N	F1	F1	F1	F2	F2	F2	F3	F3	F3	F4
N	0.95	0.97	0.97	0.96	0.97	0.97	0.91	0.92	0.94	0.87	0.92	0.83	0.92	0.96	0.94	0.6
F1	0.93	0.93	0.93	0.93	0.92	0.92	0.96	0.96	0.96	0.89	0.87	0.81	0.90	0.93	0.91	0.6
F2	0.91	0.89	0.89	0.89	0.89	0.87	0.81	0.84	0.86	0.88	0.92	0.92	0.82	0.82	0.79	0.5
F3	0.94	0.93	0.95	0.95	0.95	0.96	0.92	0.92	0.93	0.82	0.86	0.75	0.96	0.97	0.97	0.7
F4	0.74	0.72	0.76	0.77	0.75	0.76	0.79	0.76	0.75	0.65	0.69	0.56	0.82	0.80	0.83	0.9

Table 5.2: Performance of Average model for 18 test batches 'X' indicates matching 'O' indicates Mismatch

#OP Cond	N	N	N	N	N	N	F1	F1	F1	F2	F2	F2	F3	F3	F3	F4	F4	F4
N	X	X	X	X	X	X					O							
F1							X	X	X	O								
F2											O	X						
F3													X	X	X			
F4																X	X	X

5.3.1 Remarks

During initial stage of development of process monitoring(for a cold start of the plant) using pattern matching method needs huge historical data base. So collecting enough data about health of plant required some amount of time depends on batch duration. During collection of enough data parallel diagnostic method is needed. For above mentioned batch process each operating condition takes 30 hours, To proceed with averaging method here needed training data set. Till the development of average model needs to rely on any of the available methods in the literature.

- PCA can handle linear correlations in the data and effective dimensionality reduction can be possible if the data is well correlated linearly. So PCA can't handle non-linear correlations [13]. And the fact about batch data is highly time varying, uncorrelated and non-linear in nature. Using PCA similarity factor approach to this particular Acetone Butanol fermentation process case study not providing good discrimination within different operating conditions.

- To address this non linearity and uncorrelated data Corresponding analysis(CA) may be useful. In chapter 6 briefly explained about CA and the results are displayed for above case study 4.1

Chapter 6

Corresponding Analysis based Pattern matching Approach

Using CA method to develop the better indices to look into the row column associations. Generally PCA decomposes entire variance in the matrix $\mathbf{X}_{m\times n}$ which is having m samples with n columns(no. of variables). But CA decomposes measure of row column associations, typically formulated as total chi-square value to capture dependencies [14]. Since the inherent nature of variables in this batch data is non-linear in nature. If the nature of the data is more non linear, discriminating analysis based on linear scaling may not be effective which is happen same with PCA. So to get desirable characters in lower dimensional space like self-aggregation and classification, needs to do non-linear scaling. In CA non-linear scaling of data has been done to get desirable advantages like row column associations and discrimination which better suits for process monitoring.

6.1 Formulation of CA Algorithm

Objective of the CA is to find the proper lower dimension space $\mathbf{S}$ which should be the approximation of $\mathbf{X}_{m\times n}$ in terms of its proximity to the row and cloud points [14]. This optimization problem to determine the space $\mathbf{S}$ can be obtained by solving minimization problem based on weighted Euclidean distance. Here Weighting matrix $\mathbf{D}_r$ for this particular optimization problem is the distance from row cloud which is diagonal matrix containing row sums defined $\mathbf{D}_r = diag(\mathbf{r})$. And similarly $\mathbf{D}_c = diag(\mathbf{c})$. Here $\mathbf{r}$ defined as row sums and similarly $\mathbf{r}$ also defined as column cloud sum calculated as shown below equations 6.1

$$\begin{aligned} \mathbf{r} &= [(1/g)\mathbf{X}]1 \\ \mathbf{c} &= [(1/g)\mathbf{X}]'1 \end{aligned} \tag{6.1}$$

where, g is sum of all the elements in the matrix $\mathbf{X}_{m\times n}$ where 1 will be taken as appropriate dimension matrix containing all 1's. $\mathbf{D}_r$ and $\mathbf{D}_c$ will be measure of total inertia of cloud of points. So solution to minimizing this optimization problem is decomposition of inertia associated with the row(or column) cloud [15]. i.e, nothing but SVD of weighted Inertia matrix as shown in equation 6.2. Here $\mathbf{X}$ is appropriately scaled matrix.

$$\mathbf{D}_r{}^{-1/2}[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T]\mathbf{D}_c{}^{-1/2} = \mathbf{A}\mathbf{D}_\mu \mathbf{B}^T \tag{6.2}$$

such that, $\mathbf{A}\mathbf{D}_r{}^{-1}\mathbf{A}^T = \mathbf{I}_{m\times m}$ *and* $\mathbf{B}^T\mathbf{D}_c{}^{-1}\mathbf{B} = \mathbf{I}_{n\times n}$.

Here $\mathbf{D}_\mu$ contains singular values in descending order. $\mathbf{A}$ and $\mathbf{B}$ contains PC's of the inertia matrix which are principle axis to column cloud and row cloud respectively. In PCA decomposition of variance of data matrix $\mathbf{X}$ need to be done whereas in CA inertia of the data matrix needed be decomposed. Here the method of finding PC's for row cloud and column cloud are dual in nature. Finally CA based similarity factors are calculated using PC's of row cloud. Case study of batch process has been discussed in section 6.2.

6.2 CA based Similarity factor Approach

As mentioned earlier to get quantitative value to mention similarity between two data sets, need to calculate angle between those principle axises. To select number of PC's to be retained in lower dimensional space can be done by percentage of cumulative sum (as explained in section 2.2) of eigen values of $\mathbf{D}_\mu$ taken not less than 95 percentage inertia, i.e. $k = \max(k_s, k_h)$ Here k_s and k_h slected number of PC's from snapshot(S) and Historical dataset(H) respectively. After selection of PC's from both snapshot and Historical data sets, obtained lower dimensional spaces used to compare the orientation among them. Below is the similarity factor equation.

$$S_{CA} = \frac{trace(\mathbf{B}_1{}^T\mathbf{B}_2\mathbf{B}_2{}^T\mathbf{B}_1)}{k} \tag{6.3}$$

Where $\mathbf{B}_1$ and $\mathbf{B}_2$ are PC's of row cloud matrices of Snapshot and Historical data sets.

Geometrical interpretation of above similarity factor can also realized as shown in equation6.4

$$S_{CA} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(\lambda_i{}^l\lambda_j{}^m)\cos^2\theta_{ij}}{\sum_{i=1}^{k}\lambda_i{}^l\lambda_j{}^m} \tag{6.4}$$

Where θ_{ij} angle between snapshot i^{th} PC to Historical dataset j^{th} PC.s and k will be number of PCs retained to get 95 percentage variability in lower dimensional space.

Calculated similarity factors can be used for pattern matching. Using CA based similarity factor approach to identify the correct patterns in historical batches case study has been done which is explained in below section.

6.3 Case study: Results and Discussions

The main objective to this case study is looking for good discrimination among different operating conditions using CA based similarity factors. For case study Acetone Butanol fermentation batch process is used, details and brief explanation of this process mentioned in section 4.1.

- To verify this method selected 18 batches(6 Normal, 3 from each faulty operating Condition) as test batches to verify pattern matching ability of CA on same number of train batches from each operating condition.

- In the Below table cross mark indicates matching datasets to their corresponding operating conditions. Here If the CA similarity more than 50 Percentage similarity is considered. From below table there are more than 30 percentage of mismatches for all operating conditions even though there is enough discrimination to other batches. And there are mis classifications occuring own operating conditions batches itself.

Table 6.1: Performance of CA based pattern matching approach to Batch Acetone Butanol fermentation process

Cond	N	N	N	N	N	N	F1	F1	F1	F2	F2	F2	F3	F3	F3	F4	F4	F4
N	X																	
N	X		X		X													
N			X		X													
N	X					X												
N						X												
N		X																
F1							X											
F1							X											
F1		X	X					X		X								
F2										X								
F2						X												
F2										X				X				X
F3																		
F3																		
F3			X		X								X	X		X		
F4																		X
F4																X	X	
F4																X		

6.3.1 Conclusion

Similarity factor approach to quickly locate similar patterns in data base computationally not required much efforts compared to other statistical approaches. So PCA and CA based similarity factor approach has been tested. Even though CA is able to discriminate, the pattern matching performance is poor to this particular case study. Almost in every operating condition succesful patterns are approximately 50 percentage where as in PCA based similarity factor approach successfully locate the patterns without much diversifying value. Both CA and PCA based approaches have its own merits and de merits to this case study.

6.4 Future Work

Since Univariate techniques are not able to detect the correlations and slight mean changes in a dynamic data. So to detect slight mean changes and correlations Multivariate techniques are more popular. One of the multivariate technique PCA and their variants(Kernal PCA, Independent component analysis, Sensitive component analysis) are more effective in process monitoring.

From the above results mentioned, using PCA for non linear data is not give good discrimination among the operating conditions. So PCA variants may provides better results which will ultimately decreases false decision making lack of discrimination based on similarity factors.

Independent component analysis(ICA) is one of the method to extract hidden structures of data if data containing independent components. And another PCA variant is KPCA, Since the batch data is non-linear,dynamic in nature instead of going with typical PCA based methods kernel based method will deal non-linear data effectively. Based on non-linearity in the available data choosing right kernel(radial basis kernel, polynomial kernel, gaussian kernel) may results in good performance.

Bibliography

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & chemical engineering* 27, (2003) 293–311.

[2] R. Isermann and P. Balle. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control engineering practice* 5, (1997) 709–719.

[3] T. Majozi. Batch chemical process integration: analysis, synthesis and optimization. Springer Science & Business Media, 2010.

[4] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & chemical engineering* 27, (2003) 327–346.

[5] R. Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control* 29, (2005) 71–85.

[6] M. Ahmed, M. Baqqar, F. Gu, and A. D. Ball. Fault detection and diagnosis using Principal Component Analysis of vibration data from a reciprocating compressor. In Proceedings of 2012 UKACC International Conference on Control. 2012 461–466.

[7] T. He, W.-R. Xie, Q.-H. Wu, and T.-L. Shi. Process fault detection and diagnosis based on principal component analysis. In Machine Learning and Cybernetics, 2006 International Conference on. IEEE, 2006 3551–3556.

[8] P. Nomikos and J. F. MacGregor. Monitoring batch processes using multiway principal component analysis. *AIChE Journal* 40, (1994) 1361–1375.

[9] C. Zhao, F. Wang, and M. Jia. Dissimilarity analysis based batch process monitoring using moving windows. *AIChE journal* 53, (2007) 1267–1277.

[10] M. C. Johannesmeyer, A. Singhal, and D. E. Seborg. Pattern matching in historical data. *AIChE journal* 48, (2002) 2022–2038.

[11] J. Votruba, B. Volesky, and L. Yerushalmi. Mathematical model of a batch acetone–butanol fermentation. *Biotechnology and bioengineering* 28, (1986) 247–255.

[12] A. Singhal and D. E. Seborg. Pattern matching in historical batch data using PCA. *IEEE Control Systems* 22, (2002) 53–63.

[13] J. Chen and J. Liu. Derivation of function space analysis based PCA control charts for batch process monitoring 56, (2001) 3289–3304.

[14] K. P. Detroja, R. D. Gudi, and S. C. Patwardhan. Fault Diagnosis using Correspondence Analysis: Implementation issues and analysis. In 2006 IEEE International Conference on Industrial Technology. 2006 1374–1379.

[15] M. Greenacre. Theory and Applications of Correspondence Analysis. Academic Press, 1984.