Application of Multivariate Data Analysis to Improve and Optimise Industrial Processes

by

Eesha Raut

A thesis submitted for the degree of

Doctor of Engineering at Newcastle University

Biopharmaceutical Bioprocessing Technology Centre

School of Chemical Engineering and Advanced Materials

October, 2016

Abstract

ABB, who is the sponsoring company for this research work, is a global leader in power and automation technologies based in St. Neots, Cambridgeshire. The thesis discusses the work carried out on a portfolio of projects as a part of the Engineering Doctorate programme. Application of multivariate statistical process control was central to the successful implementation of the projects.

The first project focussed on a Process Analytical Technology (PAT) software solution developed by ABB. The US Food and Drug Administration (FDA) have defined PAT as a process for designing, analysing and controlling manufacturing through timely measurements of Critical Quality Attributes (CQAs) of raw and in-process materials in order to achieve final product quality. The project's overall objective was to enable seamless roll out and maintenance of chemometric models for at-line testing across multiple worldwide locations. The work presented in the thesis discusses a solution that allows global maintenance of at-line analyser measurement stations whilst providing 'real time' quality data at the right business level to enable more efficient business decisions. This required optimising the software during the preliminary stages which included developing hierarchical Partial Least Square (PLS) Models, maintaining a process within control and exporting data using the Model Data Exporter plug-in. Likewise the project involved development of a combination of test sets that could assess and improve the robustness of the product. Following the Factory Acceptance Test (FAT) and Site Acceptance Test the product was successfully commissioned at customer site.

The second project investigated a recurring uncharacteristic event in the polymerisation process. This unusual phenomenon led to downgrading of the batch further causing a loss of revenue. Previous investigations indicated that the most likely reason for this unusual behaviour was due to the occurrence of crystallisation in the polymerisation reactor. These batches were identified by monitoring a 'kink' in the heat up profile during the polymerisation process. The root cause of this crystallisation was initially examined by monitoring the rate of reaction and analysing the behaviour of one variable at a time. However, these approaches were unsuccessful to identify the underlying issue with the crystallised batches. This body of work illustrates a series of steps developed using multivariate analysis techniques to identify unusual batches in the polymer reactor. Exploratory data analysis using Principal Component Analysis (PCA) and Multi-way Principal Component Analysis (MPCA) was performed on the historic batch data (quality, process and Overall Equipment Effectiveness (OEE)) to identify

the root cause of the problem and develop a well defined method that can be used by the operators to identify abnormal batches.

Table of Contents

Acknowledgements

First and foremost I would like to express sincere gratitude to my industrial supervisor Chris Hobbs from ABB for all the guidance and support throughout my Engineering Doctorate. I could not have imagined having a better advisor and mentor for my research. I would also like to take this opportunity to especially thank Dr. Jie Zhang from Newcastle University for his insightful comments that helped me reach my goal and finish my thesis. In addition I am also thankful to ABB customers for trusting in me and without whom neither of the projects would have been successful. A big thank you to all my ABB colleagues who provided me with such a memorable experience that I will cherish for life.

I would also like to thank Prof. Elaine Martin and Prof. Gary Montague for giving me this incredible opportunity to be a part of the Engineering Doctorate programme. I would like to thank the BBTC staff for all the help that they provided throughout the course. I would also like to thank my friends and colleagues from BBTC who enriched my doctorate experience.

Also I am extremely grateful to my sister Sharvari and lovely friend Anu for making time in their busy schedule and proof read my work. Special thanks to my best friend and fiancé Chathura for all his love, patience and understanding during the final stages of the Eng.D.

And last but not the least I would like to thank my parents for all their love and encouragement throughout my academic career. It is through their support that I ventured into engineering in the first place followed by embarking onto the doctorate journey.

And finally I appreciate the financial support provided by Newcastle University and EPSRC for funding my research projects.

List of Figures

List of Tables

Abbreviations and Acronyms

800xA ABB's Process Control System

ADI Analyser Device Integration

API Active Pharmaceutical Ingredient

ATR Attenuated Total Reflectance

B Batch

BDF Difluorobenzophenone

CPP Critical Process Parameters

CQA Critical Quality Attributes

CUSUM Cumulative Sum

DPS Diphenyl Sulphone

ECS Enterprise Connectivity Software

EEM Excitation Emission Matrix

ERP Enterprise Resource Planning

EWMA Exponentially Weighted Moving Average

FAT Factory Acceptance Test

FDA Food and Drug Administration

FTIR Fourier Transform Infrared

HQ Hydroquinone

MES Manufacturing Execution System

MD Mahalanobis Distance

MLR Multiple Linear Regression

MPCA Multiway Principal Component Analysis

MSC Multiplicative Scatter Correction

MSPC Multivariate Statistical Process Control

MW Molecular Weight

NaN Not a Number

N&M Nomikos and Macgregor

NIR Near Infrared

PAC Process Analytical Chemistry

PAT Process Analytical Technology

PC Principal Component

PCA Principal Component Analysis

PCR Principal Component Regression

PDL Production Data Log

PLS Partial Least Squares

PMV1 Pre-Melt Vessel 1

PMV2 Pre-Melt Vessel 2

Poly Polymer reactor

QbT Quality by Testing

QbD Quality by Design

SAP Systems Applications and Products

SAT Site Acceptance Test

SDS System Design Specification

SNV Standard Normal Variate

SPC Statistical Process Control

SPE Squared Prediction Error

SVD Singular Value Decomposition

UCL Upper Control Limit

1 Introduction

1.1 Motivation and Objectives

Background

The work reported in the thesis discusses two projects with novel application of Multivariate Statistical Process Control (MSPC) methods. For the first project MSPC method has been utilised for the implementation of Process Analytical Technology (PAT) within industry. This concept of PAT is in line with the FDA's Quality by Design (QbD) approach which aims to move pharmaceutical manufacturing processes from a rigid regulatory approach to a more flexible science and risk based approach. The Analyser Device Integration (ADI) software solution addresses the data management challenges associated with the industrial implementation of PAT. ADI software allows integration of quality management systems such as System Applications and Products (SAP) software with process control system and third party chemometric softwares. This thesis mainly focused on three aspects of the ADI product: 1) the integration of existing hierarchical Partial Least Squares (PLS) models within the novel software solution, 2) developing a method or a recipe to take successful at-line measurements and ensuring method compatibility across various measurement stations and 3) optimising the Model Data Exporter (MDE) tool to introduce selection methods or filters in order to access the right information at the right time for process improvement activities.

For the second project Principal Component Analysis (PCA), an MSPC technique was used to identify the source of an unusual crystallisation problem in polymer reactors. While the primary aim of this project was to identify the root cause of crystallisation in the reactors, the work carried out in the thesis also attempts to identify correlation between a quality variable and a crystallised batch. One of early works in monitoring batch polymerisation process using MSPC was introduced by Nomikos and MacGregor (1994). Following this a number of other publications such Kourti and MacGregor (1995), Nomikos and MacGregor (1995) Martin et al. (1996), Dong and McAvoy (1996) and Wold et al. (2009) have also discussed novel MSPC methods to monitor batch processes. Most of this work discussed in literature has been around the application of MSPC methods to monitor and control the polymerisation process. However, to the best of author's knowledge application of MSPC to identify this unusual crystallisation problem in a polymerisation process is novel and has not been reported in literature.

Project 1 – Business Productivity Improvement through the Application of Analyser Device Integration

The ADI software aims to implement the concept of PAT introduced by the US FDA by finding a solution to data managing and data integration issues. Increasing number of pharmaceutical, biotechnology and food & beverages companies have adopted PAT for better understanding of QbD approach and real time release. Typically in batch processes the product quality is measured at the end of the process thus any deviation in the final quality of the product cannot be rectified and a batch has to be discarded. PAT encouraged the utilisation of analytical devices such as NIR probes that could monitor the quality of the product during the process operation. Thus PAT encouraged to move away from the typical Quality by Testing (QbT) approach to QbD approach. PAT aimed to revolutionise these industries by increasing the manufacturing efficiency, reducing cost and product rejects, and providing opportunities for continuous improvement.

Most of the work reported in literature has focussed on developing analytical methods to monitor product quality using advanced spectroscopic techniques. For example Chavez et al. (2015) have explored the application of NIR methods to quantify the API content in non-coated tablets and Schaefer et al. (2013) have utilised NIR spectroscopy to control API crystallisation in manufacturing synthesis process. Also Cárdenas et al. (2015) have developed an innovative tool to prepare calibration sets based on process spectrum and establishment of model space by Hotelling's T^2 and Q-residual statistics.

The successful implementation of PAT requires tackling the data management issues that arise from large sets of data generated from the analytical instruments, handling the models, process parameters and the associated meta-data. The next step for PAT is to integrate these data sets in a synchronised manner to the business systems such as SAP software to enable traceability and continuous improvement of processes. The work covered in this thesis gives a brief overview of the unique solution that addresses the data handling issues with emphasis placed on deploying method, model development and data export tool within ABB's 800xA control environment to successfully implement PAT solution at the customer site.

Project 2 – Using Multivariate Analysis to Monitor Crystallisation Issue in Polymerisation Reactors

The second part of the thesis looks at a unique problem encountered by the customer manufacturing speciality polymers. The manufacturing process for polymerisation reaction is controlled by monitoring the contents temperature in the reactor using a cascade control loop. An unusual 'kink' in the contents temperature profile is considered as an indication of a crystallised batch. The customer has tried mitigating this problem by studying the chemistry and monitoring the rate of reaction in the process and also by univariately investigating variation in process and quality variables. The work done in this thesis aims to identify the crystallisation problem using multivariate statistical methods. Most of the work reported in literature revolves around the application of multivariate statistical process control to control a polymerisation reactor. However, this thesis reports the novel application of MSPC methods used to investigate an unusual crystallisation problem that is particular to this manufacturing process.

1.2 Thesis Contribution

For the first project the following contributions have been reported in the thesis:

Method configuration and deployment for successful implementation of ADI

The author was responsible for the configuration of a 'Method' which includes analyser settings, prediction and background models as well as other inputs such as constants, variables and process values. A number of tests were devised by the author for the purpose of this project to successfully implement the ADI solution. With multiple measurement stations used all over the world a method was required to be compatible in order to acquire measurements across multiple analysers. The author was involved in the development of a novel solution to allow method flexibility across various locations while also addressing background validity issues associated with it.

Model deployment and generation of respective model and process alarms

A number of existing hierarchical prediction and background models were successfully integrated with the ADI product. This required the author to be in a unique position to understand chemometric model development as well as the working functionality of the ADI software. While developing unique test scenarios to ensure efficient performance of the models, the author was also involved in generating various process and model alarms. This was carried out by setting the limits within the model such that they would generate the

respective alarms with the simulated spectral data. The simulated spectra were set up to resemble the actual spectral data that would be collected during measurement of at-line samples on the plant.

The work also focussed on optimising the software so that it would appropriately handle Not a Number (NaN) within the system. The contribution was crucial for understanding of the ADI system and chemometrics for successful model deployment.

Two key contributions during the model development and implementation phase have been summarised below:

1) Savitzky-Golay pre-treatment was applied on the data set as a pre-processing method to smooth the data. Savitzky-Golay smoothing and differentiation filter is used to increase the signal-to-noise ratio. The filter optimally fits a set of data points to a polynomial to minimise the least squares error. Once the pre-treatment was applied on the measured data the number of columns in the raw spectral data was reduced and less number of columns were available to project the PLS model. Since the number of columns in a 'Method' was hardcoded into the 800xA control environment the author suggested enabling more flexibility within the system. The modified solution would take into consideration the pre-treatment of the spectra with reduced number of spectral columns following the application of Savitzky-Golay filter.
2) Large size of PLS models were being imported into the system in order to take at-line measurements. However the system was unable to handle the large amounts of data within the model which resulted in the system to crash every time a model was imported within the 800xA environment. The author's contribution to this problem included identifying the reason for the system crash and requesting a solution to reduce the size of the model without affecting its predictive capabilities.

Optimising the data export tool

The Model Data Exporter (MDE) was the most important tool for the customer to access their historical data in order to improve their processes. The thesis discusses the development of various selection methods that enabled the chemometricians to access the right data for process improvement through data mining. The work reported in the thesis also looks at the various tests scenarios that were developed in order to enhance the functionality of the data export tool.

For the second project the following contributions have been reported in the thesis:

Identifying a possible co-relation between crystallised batches and quality variables

1) Identify a co-relation between crystallised batches and quality variables using exploratory data analysis.
    a. PCA was used to establish a possible co-relation between contamcount variable and/or initial raw material charge of Diphenyl Sulphone (DPS) with the crystallised batch. Contamcount is currently used to measure the amount of black residue that scours of the sides of the vessel and it is believed that high amount of black residue could result in a crystallised batch. Insufficient amount of DPS is also believed to be a reason for unusual behaviour of a bad batch. Both these variables in addition to the remaining quality variables were analysed to determine a possible relationship.
2) Identify co-relation between crystallised batches and particle size of raw material.
    a. Smaller particle size of the raw material could increase the rate of reaction in the polymerisation process. This in turn could result in sudden increase in temperature that is associated with the occurrence of crystallisation. The work reported in the thesis analyses the particles size distribution of the raw material and attempts to identify a co-relation with the crystallised batches using PCA.

Examining the correlation between crystallised batches and process variables and further investigating the root cause of crystallisation

The work carried out in this thesis investigates a deviation in the process data that is particular to a crystallised batch. With process data being three dimensional in nature (Batches x Variables x Time) the data was unfolded into two dimensions using two different approaches. In the first approach that data was unfolded using the Nomikos and Macgregor (N&M) approach (Nomikos and Macgregor, 1994) or batch-wise unfolding approach to allow batch to batch comparison and in the second approach the data was unfolded using the Wold's Approach or variable-wise approach to monitor the trajectory as a batch progresses. In addition to contents temperature profile the study also establishes co-relation between sudden increase in KwRise after hold point 1 and/or out of control level after hold point 1 with occurrence of crystallisation after hold point 2.

Developing a method to fingerprint a typical crystallised batch

This thesis examines the work carried out for the development of a new method to identify crystallised batches using exploratory data analysis on the process data. The current fingerprinting method has been developed for one of the six reactors. This novel method would remove operator dependability on identifying the 'kink' in the temperature profile. The well defined technique would identify the crystallised batches analysing the process data and monitoring the statistics.

1.3 Publications and Conferences

Journal Publication

Raut, V., E, Hobbs D.,C. "Analyser Device Integration – the power of Analytical Data", Planned submission to Journal of Process Control in August 2016

Conference Talks

Hobbs, D., C, Raut, V., E. "Analyser Device Integration – the power of Analytical Data" ACHEMA 2015, June 15- 19 in Frankfurt

Raut, V., E, Hobbs, D., C. "Business Productivity Improvement through the application of Analyser Device Integration", EuroPACT 2014 conference, May 6-9, 2014 Barcelona

Conference Posters

Raut, V., E, Martin, B., E, Hobbs, D., C. "Application of Multivariate Analysis to identify the crystallisation issue in polymerisation reactors", BIOPRO Worlds Talent Campus, 2014, Denmark

Raut, V., E, Hobbs, D., C. "Business Productivity Improvement through the application of Analyser Device Integration", EuroPACT 2014 conference, May 6-9, 2014 Barcelona

Raut, V., E, Hobbs, D., C. "Business Productivity Improvement through the application of Analyser Device Integration", ChemEngDayUK, 7-8 April 2014, The University of Manchester

1.4 Layout of the Thesis

Chapter 2 outlines the introduction to multivariate statistical methods that are applicable for both the projects studied in this thesis.

Chapter 3 explores the concept of PAT and application of PAT tools within ABB's ADI product

Chapter 4 discusses in depth the working of Analyser Device Integration product and the contribution of the author in the successful implementation of the novel product.

Chapter 5 discusses the background knowledge of the crystallisation problem in the polymerisation reactors particular to the customer.

Chapter 6 looks at the exploratory data analysis performed on the various sets of data provided to identify the root cause of crystallisation.

Chapter 7 concludes the thesis and discusses the suggested future works with respect to both the projects.

2 Chapter 2 - Introduction to Multivariate Statistical Methods

2.1 Chapter Overview

This chapter discusses the literature review that would be applicable for both the projects that are a part of this thesis.

Section 2.2 gives an overview of Multivariate Statistical Process Control (MSPC) techniques and growing application of these techniques within the industry.

Section 2.3 describes in detail the Principal Component Analysis (PCA) method and Section 2.4 explains the Partial Least Squares (PLS) method.

Section 2.5 discusses the various pre-treatment methods applied prior to developing multivariate models.

Section 2.6 describes the various observation diagnostic tools that are used for analysing multivariate data.

Section 2.7 explains multi-way techniques used to unfold three dimensional data into two dimensions to further carry out MSPC.

2.2 Introduction of Multivariate Statistical Process Control

Statistical Process Control (SPC) is traditionally a univariate method used to predict the product quality by monitoring variables one at a time. However since most industrial processes are multivariate in nature they have more than one variable affecting the process behaviour at a given time. Moreover it is the interaction and correlation between these variables that causes deviation in a process. Thus looking at only one variable at a time by performing univariate analysis could misrepresent the underlying real behaviour of the process.

In order to tackle these limitations SPC further evolved into Multivariate Statistical Process Control (MSPC). In the recent years with the advance in computers almost every process variable on a manufacturing plant is measured and historised. The acquired data could be extremely valuable if used to optimise processes, improve safety and reduce environmental risks. This concept and method has been popular with the manufacturing industry to maintain a process within a state of SPC. A system is said to be in state of SPC if certain process variables remain close to their expected values and with 'common cause variation' being the only source of variation present in the process (Kourti and MacGregor, 1995). Unlike automatic feedback process control where unusual or new event is addressed by simply continuing the process by compensating for the deviation, SPC aims to identify the cause for process variation and implement long term improvements. Traditional SPC utilises historical data to monitor batch processes. However most of these methods tend to utilise SPC charts such as the Shewart (X (bar) and Range Charts), Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) to monitor and control process variables. These charts were deemed to be inadequate for most modern processes where large amount of data is being collected with variables highly correlated with each other.

MPSC techniques, such as PCA and PLS as well as their associated control charts were developed to overcome these issues by using reduced number of latent variables than the apparent dimension of the process represented by the number of measured variables. These techniques take into account the inter-relationship between variables. Both PCA and PLS are suitable for analysing large sets of correlated data. PCA mainly explains the variation in the X data matrix while PLS analyses both X data matrix and Y data matrix (Martin et al., 1996).

Batch and semi-batch processes constitute a large number of chemical and pharmaceutical industries manufacturing speciality chemicals and high value added products. With rise in competition these companies are under increasing pressure to get the production right at the

first time. This was evident in the projects further discussed in this thesis with the primary purpose of both these projects being continuous improvement and optimisation of industrial processes. The first project focused on the application of PAT in food and chemical industry while the second project attempts to address the crystallisation issue in polymerisation reactor using MSPC.

Traditionally the quality of the product in a batch process is verified once it has reached the end of manufacturing process. Process steps are successfully operated in a repeatable manner and validated after three consecutive successful batches (Boudreau and McMillan, 2007). Most of the pharmaceutical industries keep to this conventional recipe driven approach followed by offline lab based analysis of the products to ensure it is within the required specification. For a long time the US FDA believed that quality cannot be tested into products but has to be built into the process. Recently there have been significant advances to improve pharmaceutical development and manufacturing and enhance the product quality through advanced process control and process development. However rigid manufacturing procedures and requirement for a number of regulatory approvals for the introduction of a process change hindered the application of these innovative approaches. In 2004 realising this need to encourage innovation the US FDA introduced the PAT initiative which states that "*Process Analytical Technology (PAT) is a system for designing, analyzing and controlling manufacturing through timely measurements (i.e. during processing) of critical quality and performance attributes of raw and in-process materials and processes with a goal of ensuring final product quality"(FDA, 2004).* One of the most important tools used in PAT has been the application of multivariate techniques which have been further discussed in this chapter.

2.3 Principal Component Analysis

2.3.1 Introduction

Given a data matrix X, consisting of n rows (observations/samples/batches) and p columns (variables) PCA provides an approximation of the data table in terms of k new variables or principal components. These k new variables will account for variation in the p original variables. PCA is one of the oldest multivariate techniques first introduced by Karl Pearson in 1901. He formulated the analysis as finding "lines and planes of closest fit to systems of points in space" (Pearson, 1901). It was then transformed to its current form by Hotelling in 1933. The advent in powerful computers gave this technique a much needed boost so it could be applied in diverse fields such as chemistry, engineering, geology and sociology.

As mentioned earlier PCA generates new set of data consisting of principal components which are linear transformation of the original variables that are mutually orthogonal to each other. PCA decomposes the data set into certain number of Principal Components (PCs) and each PC is described by a scores vector (t_r) and loadings vector (P_r). The score is the distance from the origin of the plane along each PC and is calculated as the product of the loading vector and observation. The first PC explains the greatest amount of variation in dataset while the second PC explains the next greatest variation and so on. Highly correlated variables usually require lower number of principal components to explain the total variance in the data (Nomikos and MacGregor, 1994). For example a highly correlated spectral data set analysed using PCA would typically require 2 or 3 PCs to explain most of the variation in the captured data as compared to another set of process data which could need up to 6 PCs to explain the maximum variation in the data set. In theory as many principal components as original variables can be calculated however in practice one rarely needs to compute all the PCs since the major source of variability in the data set can be captured by a small number of principal components. Quite often the lower PCs explain subtle process variation otherwise not observed in the higher PCs. The number of PCs required and selected to explain the sufficient amount of variation in the data set is described further in Section 2.3.1.2. For PCA the linear combination of the dataset X can be written as shown in Equation 2-1 where X denotes the matrix of PCs whose columns are the scores vector (T_r) and loadings vector (P_r).

$$X = \sum_{r=1}^{R} T_r P_r^T + E \qquad \textbf{Equation 2-1}$$

where R is the total number of PCs retained in the model which is less than or equal to the number of variables (or observations if the number of observations is less than the number of variables) in the original data and E contains the residual matrix. If all the PCs are retained in the model the residual matrix would be equal to zero. The orientation of every sample/observation is explained by loadings vector P_r which defines the greatest variability while the score vector T_r represents the projection of each object on loading vector.

The calculations of the various components of the PCA model such as scores, loadings and residuals have been explained in Section 2.3.1.1.

2.3.1.1 Geometrical and Algebraic Interpretation of PCA

For a data matrix X, with n observations and k variables there can be as many dimensions as there are variables. Each co-ordinate axis would represent each variable. As seen in Figure 2-1 the three axes (X_1, X_2 and X_3) represent three variables. The data would usually be standardised by scaling it to mean zero unit variance. The first principal component is the line in the K-dimensional space that captures the main source of variation in the data in the least squares sense. The original variable observations are then projected onto this line to represent the new co-ordinate PC1. The second PC is also represented by a line in K dimensional space however it is orthogonal to the first PC and explains next greatest amount of variation in the data matrix. The co-ordinate value of each observation on the principal component space is called as a score. By plotting this projected configuration one is able to identify the relationship between observations on a lower dimensional space. Loading vector defines the orientation of the model plane hence direction of the PCs in the K variable space. In summary the location of each observation is explained by a score while the positioning of the principal components is defined by the loadings.

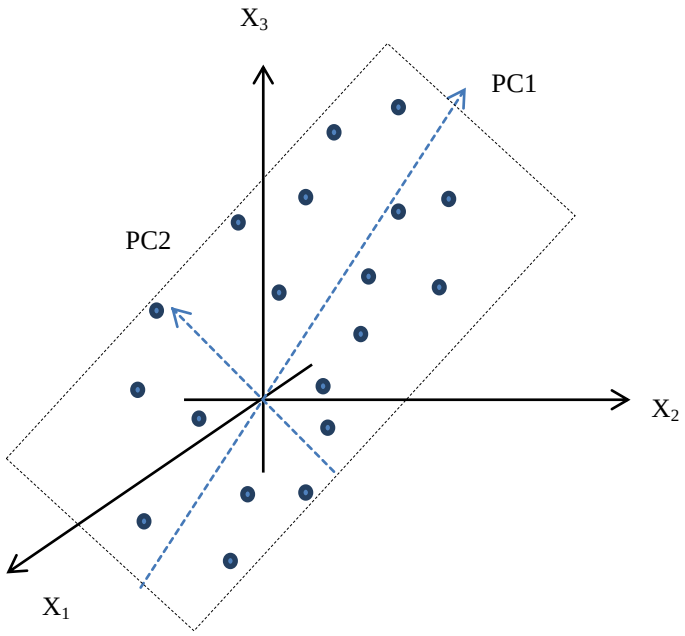

Figure 2-1: Geometrical Interpretation of PCA

A number of methods can be found in literature to derive principal components. However the most common methods used in practice are the Non-Iterative Partial Least Square (NIPALS) an iterative algorithm and the Singular Value Decomposition (SVD), a non-iterative algorithm. The NIPALS algorithm is normally applied to data with missing values. The

NIPALS tends to be faster than SVD if the number of rows and columns are large (Unscrambler, 2014).

Since data used in both the projects discussed in thesis consisted of large data matrices the NIPALS algorithm was utilised to develop the multivariate models.

Principal components can be used to provide approximation of data matrix X using Equation 2-1. Once a raw data set has been acquired it is then normalised by mean centring or auto-scaling. The normalised data set is then used to derive the covariance matrix as shown in Equation 2-2.

$$cov(X) = \frac{X_{Norm}^T X_{Norm}}{N-1} = \begin{bmatrix} C_{II} & \dots & C_{IJ} \\ \vdots & \ddots & \vdots \\ C_{IJ} & \dots & C_{JJ} \end{bmatrix} \quad \textbf{Equation 2-2}$$

where c_{ij} denotes the covariance between the i^{th} and j^{th} variable.

The eigenvectors i.e. the loading vectors are calculated from the covariance matrix and the corresponding eigenvalues denote the variance of the principal components (MacGregor and Kourti, 1995).

Once the loading vectors have been estimated they can now be used for calculating the score vectors as seen in Equation 2-3.

$$T = X \times P \quad \textbf{Equation 2-3}$$

Thus using the estimated scores vector the predicted data matrix $\hat{X}$ is calculated by $T \times P^T$.

The residual matrix E is the difference between X and $\hat{X}$ as seen in Equation 2-4.

$$E = X - \hat{X} \quad \textbf{Equation 2-4}$$

The data matrix X can also be written in terms of its vector components as show in Equation 2-5.

$$X = t_1 p_1^T + t_2 p_2^T + \cdots t_R p_R^T + E \quad \textbf{Equation 2-5}$$

2.3.1.2 Selecting the number of principal components

Once the total number of principal component's (R) are calculated it is important to determine the maximum number of principal components (r) to be retained in the model that could capture the major source of variation in the data. Including more number of PCs in the model than required could affect the sensitivity of the model since the lower order principal components with eigenvalues less than 1 may be representing the noise in the process. One of the methods to determine the number of principal components to be retained is to consider the cumulative percentage of variance explained by studying the eigenvalues or plotting the values against the principal component number. On mapping the eigenvalues any sudden drop in the plot indicates the number of principal components needed to explain the major source of variation in the data set.

Cross validation is another technique for determination of principal components described by Wold (1978). In this method each row from the data set is omitted once and PCA model is developed using the remaining data set. A predictive model for the omitted row is then developed using the calculated PCA model. Predictive Error Sum of Squares (PRESS) is then calculated for the omitted row and number of principal components to be retained is determined as that which gives the minimum residual error. Often when the number of samples is large they are split in groups. Each group is excluded when the PCA model is built. The PRESS is calculated for each excluded group and then are summed up to give the total PRESS.

2.4 Partial Least Squares

2.4.1 Overview

Partial Least Squares also known as Projection to Latent Structures (PLS) is a regression technique which can be applied when variables can be divided into cause/measured variables X and effect/quality variables Y. It can be used to model one effect variable or multiple effect variables at the same time. PLS method maximises the covariance between X and Y. Unlike PCA, PLS model uses both X and Y matrices to find the latent variables in X in order to predict the latent variables in Y. It is often used as an alternative technique to Multiple Linear Regression (MLR) since it is able to produce more robust models on addition of calibration samples from new population. The PLS model also tends to be more accurate than other algorithms when there is high correlation between cause variables (Geladi and Kowalski, 1986). The fundamental regression equation used for all the regression modelling approaches such as MLR, Principal Component Regression (PCR) and PLS is given by Equation 2-6.

$$Y = X\beta + E$$ **Equation 2-6**

where Y is the quality or predicted variable, X is the measured or predictor variable and β is the regression co-efficient.

Thus by creating a new set of latent variables PLS models show the correlation between measured variables (X) and quality variables (Y) where each latent variable is linear combination of X. The Non-linear Iterative Partial Least Squares (NIPALS) is the most commonly used algorithm to model PLS. The NIPALS algorithm can handle any missing values and is more suited to calculate a few latent variables.

2.4.2 Geometrical and Algebraic Interpretation of Partial Least Squares

As seen in Figure 2-2, J is the number of process variables in data set X and K is the number of variables in effect data set Y. The number of samples/observations for both the X and Y data set is given by I.

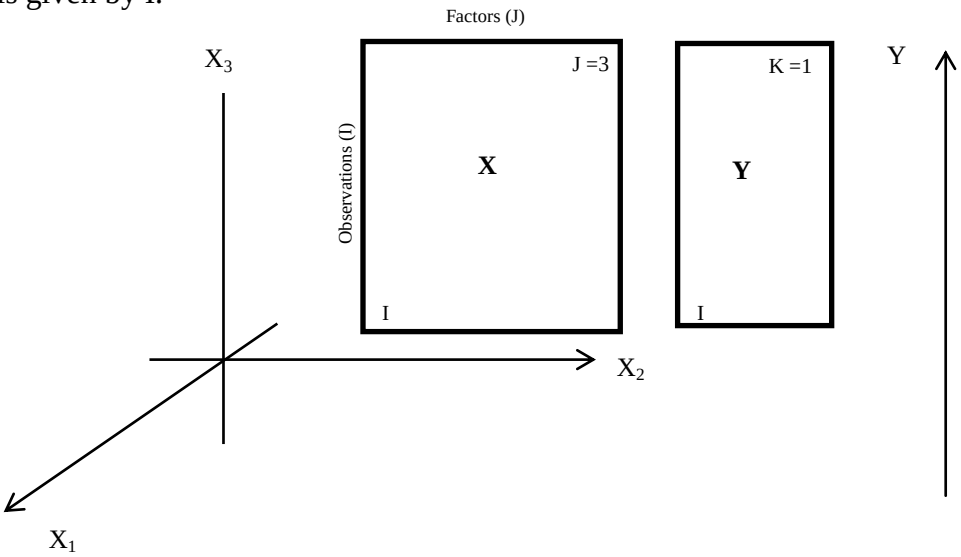

Figure 2-2: Geometrical Interpretation PLS

Each variable has one co-ordinate axis with its length defined by scaling it to mean zero unit variance. Once the data sets have been auto-scaled the first latent variable is calculated such that this factor is a line in the X space which provides a good correlation with the y vector. The score for the observations are obtained similar to PCA where the samples are projected on this line to obtain score t_1. The score vector t_1 can now be used to calculate the y estimate y_1 by multiplying t_i with weight of vector y.

The second principal component is calculated by projecting a line in the k-dimensional space orthogonal to the first PC. The second PC usually explains less percentage of data variation in

the data set as compared to the first PC. However if second PC indicates more correlation than first PC then according to Eriksson et al. (2006) this would indicate a strong structure in X which is not visible in the effect variable Y.

In terms of the mathematical representation for given independent data set X (with I samples and J process variables, X_j) and effect variable or dependent variable Y (with I samples and k cause variable Y_k) the factor of the cause data t_h (length I) and effect data u_h (length K) can be calculated using Equation 2-7 and Equation 2-8. Both these equations define outer relationship between data sets. E and F are the residual matrices for the X and Y data sets. An ideal model would be the one with zero residuals. Thus smaller residual is an indication of a good predictive model.

$$X = \sum_{h=1} t_h p_h^T + E \qquad \textbf{Equation 2-7}$$

$$Y = \sum_{h=1} u_h q_h^T + F \qquad \textbf{Equation 2-8}$$

An inner relationship can be produced by performing linear regression between the t_h and the u_h vectors as shown in Equation 2-9.

$$u_h = b_k t_h + \varepsilon_h \qquad \textbf{Equation 2-9}$$

The NIPALS algorithm is the most commonly used algorithm to estimate the PLS model. As seen before the PLS model is given by the Equation 2-6. However unlike PCA $T_h \neq XP_h$ but is calculated using the NIPALS algorithm as discussed below that was introduced by Wold et al. (1984).

1) Centre and scale X and Y data
2) Start with calculating vector u that is normally one of the columns of Y.
3) The x weights given by w: $w = X'u/u'u$
4) Normalise w to unit length: $w = \frac{w}{\|w\|} = 1.0$
5) Calculate X-scores, t: $t = Xw$
6) Calculate the y weights denoted by u: $c = Y't/t't$
7) Normalise c to unit length: $c = \frac{c}{\|c\|} = 1.0$
8) Calculate the updated set of u vector: $u = Yc/c'c$

9) Check if the output vector u has converged. This is tested in t i.e. if $\frac{\|t_{old}-t_{new}\|}{\|t_{new}\|} \ll \varepsilon$ where ε is small e.g. 10^{-6} or 10^{-8}. If converged then go to next step otherwise return to step 2. Incase of only one y-variable the algorithm converges in a single iteration (Wold et al., 2001).
10) Calculate the X loadings: $p = X't/t't$
11) Calculate $X_{new} = X - tp'$
12) Calculate $Y_{new} = Y - t'c$

Calculate the next component until the maximum number of variables have been calculated. The cross validation method explained in Section 2.3.1.2 can be used to indicate maximum number of principal components that would be required to explain maximum information about X in Y.

2.5 Pre-treatment

Data has to be pre-treated depending on what type of data is available and what method will be used to analyse the data set. Pre-treatment, although a minor part of data analysis it is an extremely crucial step that determines if a model is useful or not. Data can be pre-processed in a number of ways such as 1) outlier removal if a particular measurement is outside its limits due to human or measurement error, 2) transformation which is applied to normalise and linearise the data or 3) filtering the data which is normally applied to smooth the data by removing the noise (Candolfi et al., 1999). Outlier removal in combination with filtering the data using the Savitzky-Golay method was used in the ADI project discussed in Chapter 3 and Chapter 4.

With the advancement in analysers a number of different types of data such as process data, spectroscopic data and quality measurements can be collected with variables having different numerical ranges. Sections 2.5.1 and 2.5.2 mainly look into the basic pre-processing steps often applied to process data in order to transform it to a suitable form to carry out the analysis. Advanced pre-processing techniques required by the spectroscopic data have been discussed in detail in Section 2.5.3.

Variable with a large numerical range is expected to have more variance as compared to a variable having a smaller range. Taking this point into consideration it can be noted that pre-treatment is an extremely important step for PCA which is maximum variance projection method. Inappropriate data pre-processing would mean a variable with smaller variance would not be as expressed in the analysis as a variable with a larger variance.

2.5.1 Mean centering

Mean centering is one of the commonly used pre-processing techniques where the average value of each observation is calculated and then subtracted from each data observation. The goal of this method is to shift the variable trajectories to a common baseline. In other words Miller (2005) explains that this operation enhances the focus on response variation by eliminating the absolute intensity information from each of the variables. The author also indicates that not all data matrices would require mean centering and sometimes there can be unknown information hidden in the mean that could be critically relevant to develop a reliable model.

The general equation used to calculate mean centered data is given by

$$X_{mc} = X - \bar{x} \qquad \textbf{Equation 2-10}$$

Where X = original data

X_{mc} = mean centered data

$\bar{x}$ = mean response values

2.5.2 Auto-scaling

Scaling is normally the next step used in standard pre-processing method once the data has been mean centered. Auto-scaling procedure is mainly mean centering data followed by dividing by the standard deviation as shown in Equation 2-11

$$X_{as} = (X - \bar{x})/X_{SD} \qquad \textbf{Equation 2-11}$$

X_{SD} = standard deviation for individual variables

X_{as} = auto scaled data

This pre-processing method ensures that each variable has equal footing in the analysis and that each variable would exhibit similar level of variability. Auto-scaling is very much essential when unit of measurement is different for each variable or when different types of instruments have been used to capture data. If auto-scaling in not applied in these scenarios and data is analysed using variance maximisation techniques such as PCA then there is possibility of developing an inconsistent model. Sometimes no scaling is needed at all for example in the FTIR spectra where all the variables are expressed in the same unit and it might be important to retain the variance information to yield relative sensitivities of different wave numbers.

2.5.3 Advanced Pre-processing

Advanced pre-processing techniques are commonly applied on data collected from spectroscopic analysers such as NIR and Raman. Recent advancement in analytical chemistry has resulted in increasing amount of complex data. This raw complex data that is generated by the analysers needs to be transformed into 'clean' data by removing unwanted variation present in the data set and improving the linear relationship between spectra and analyte concentration (Engel et al., 2013). A number of experimental and instrumental phenomena such as scatter from particulates, molecular interactions, missing values, changes in sample size/path length etc., can cause deviation from the linear relationship established in Beer Lambert's law which states that absorbance is directly proportional to the concentration. A number of pre-processing methods that have been developed recently are broadly classified into two groups: scatter correction methods and derivative methods. The scatter correction methods include Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), Normalisation and Baseline Correction while derivative methods include Savitsky-Golay and Norris-Williams techniques (Rinnan et al., 2009b). Selecting an incorrect pre-processing method can have a detrimental effect on the final results of the data analysis. Engel et al. (2013) has reported three types of pre-processing selection methods namely, trial and error where a number pre-processing techniques are applied and the one with the best outcome is chosen, visual inspection and assessing the pre-processed data by quantifying the effect of quality parameters on the final outcome.

MSC and SNV methods have been used to counter the light scattering effects introduced by the presence of particles in the samples. Both the methods have similar equations and produce comparable outputs for the pre-treated spectra.

MSC was introduced by Martens et al. (1983) who developed this method to eliminate optical interference. MSC pre-treatment is a two-step process described in the following equations:

$$x_{org} = b_0 + b_{ref,1} x_{ref,1} + e \qquad \textbf{Equation 2-12}$$

$$x_{xcorr} = \frac{x_{org} - b_0}{b_{ref,1}} \qquad \textbf{Equation 2-13}$$

Equation 2-12 is used to estimate the correction co-efficient and the corrected spectra are calculated using Equation 2-13. In the equations bs' are the correction co-efficient, e is the unmodeled part, and x_{org}, x_{ref} and x_{corr} are the original, reference and the corrected spectra

respectively. The MSC technique was further expanded by Martens and Stark (1991) into extended MSC that included wavelength correction.

The basic equation of SNV correction introduced by Barnes et al. (1989) and normalisation has the same form as Equation 2-12. The difference between MSC and SNV corrections is that a reference spectrum is not required in SNV correction.

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \qquad \textbf{Equation 2-14}$$

For normalisation a_0 is always zero while a_1 depends on the type of normalisation that has been used.

Savitsky-Golay and Norris-Williams are the most commonly used derivative methods for pre-processing. Both the methods smooth the data so that signal to noise ratio is not reduced extensively in the corrected spectra (Rinnan et al., 2009a). Application of derivative pre-processing has the ability to remove the additive as well as the multiplicative effects in the spectra. Savitzky and Golay (1964) first suggested this method that smoothed the value for each data point by performing a polynomial regression. On comparing the Savitsky-Golay method with Norris-Williams, Rinnan et al. (2009a) suggests that the NW is a two-step process with derivations similar to finite difference method. These derivative methods do not generate similar results but aim to maintain acceptable signal-to-noise ratio.

In the first project further discussed in Chapters 3 and 4 the customer extensively used second order Savitsky-Golay pre-processing technique before developing the PLS models.

2.6 Observation Diagnostics

Chemometric tools such as PCA and PLS discussed in Section 2.3 and Section 2.4 can be used to extract information by converting a data matrix to a few plots. These tools identify unusual samples in a population by discovering them as outliers. Depending on the how extreme these outliers are they can be categorised as serious outliers or moderate outliers. Outliers can be identified in most of the process performance representation such as scores, loadings, residuals etc. that have been discussed in detail in the following sections.

This section looks into various graphical model parameters that can be used to optimise and improve processes. These model parameters enable better understanding of the underlying behaviour of variables and establish relationship within samples as well as between samples and variables.

2.6.1 Principal component scores and loadings

Scores plot are used to interpret the relationship between various observations or samples. It is the projection of data on subspace that reflects the sample location along the principal component. Thus each sample has a score on every PC. Plots could either be univariate where every score is plotted along the sample on the x-axis or a bivariate plot with PC1 and PC2 defining the co-ordinate axes. Figure 2-3 and Figure 2-4 shows a typical univariate and bivariate scores plot for wine production in various countries. This model was generated using an example data set in PLS_Toolbox software. A total of 10 samples and 5 variables were used for the analysis. The first two principal components capture approximately 78% of variation in the data set. In a bivariate scores plot samples close to each other would indicate similarity between the observations while the ones which are far away from one another would indicate difference between those samples.

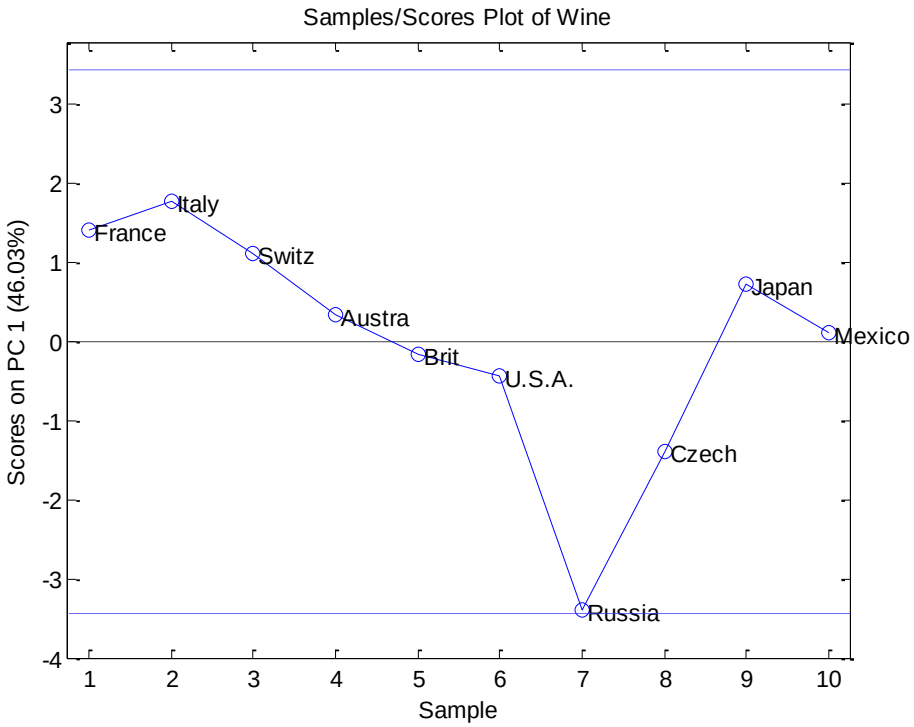

Figure 2-3: Univariate scores plot

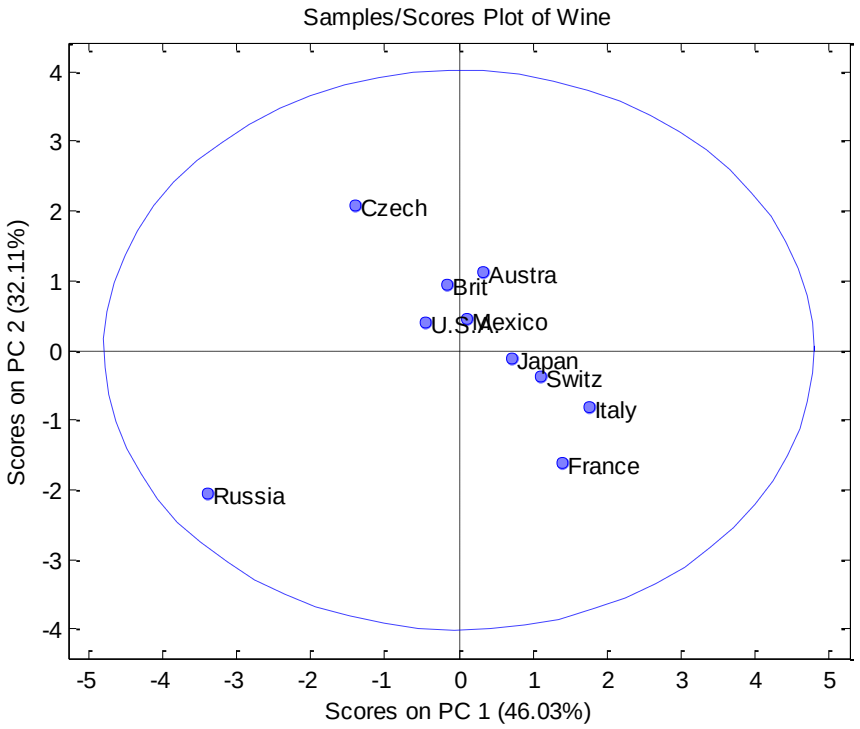

Figure 2-4: Bivariate scores plot

The loadings can be used to further investigate which variables determine the positioning of these samples in the principal component space. Every variable in the data matrix projects a loading on PC which explains the variation contained in the variable. A bivariate loadings plot for dominant PCs can detect correlation between X-variables. In a typical bivariate

loadings plot samples which lie in same quadrant would be positively correlated while the ones lying in diagonally opposite quadrant would be negatively correlated. For example Figure 2-5 shows that Beer and Liquor are negatively correlated to each other while a slight positive correlation between Wine and Life Expectancy can be observed. This behaviour must however be consistent across all the PCs in order to conclude a particular finding. Also variables close to the origin in a bivariate loadings plot have a lesser impact on the model as compared to variables that rest further away from the origin.

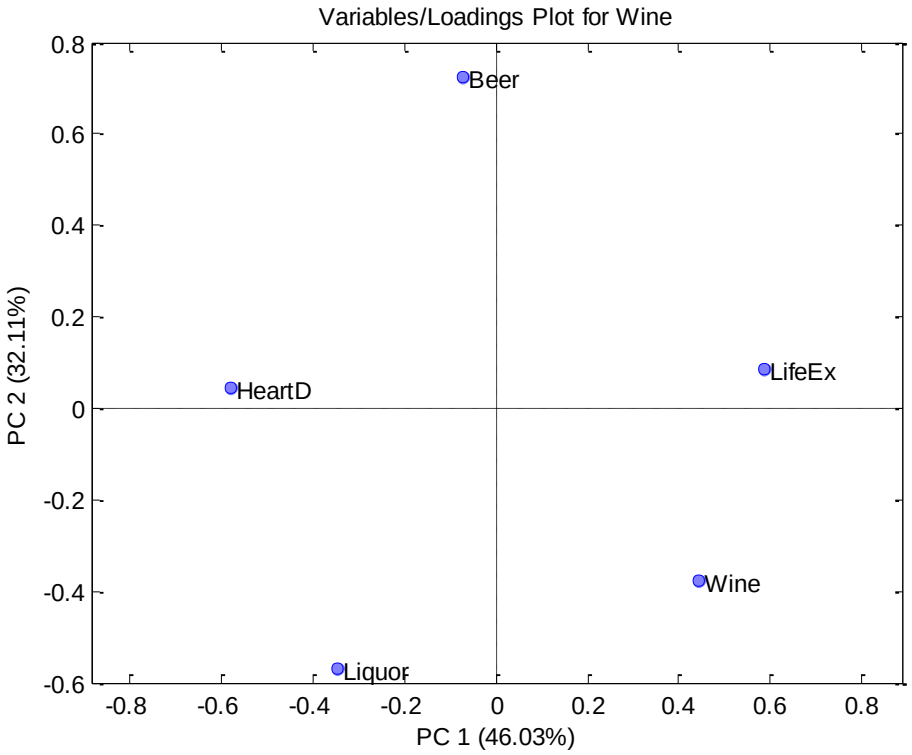

Figure 2-5: Bivariate loadings plot

Loadings and scores are complimentary to each other and have to be interpreted together when analysing a data set. A bi-plot which is 2 dimensional scatter plot superimposes scores and loadings data on the same graph. This allows for simultaneous interpretation of sample behaviour and variable relationship. An example of bi-plot can be seen in Figure 2-6.

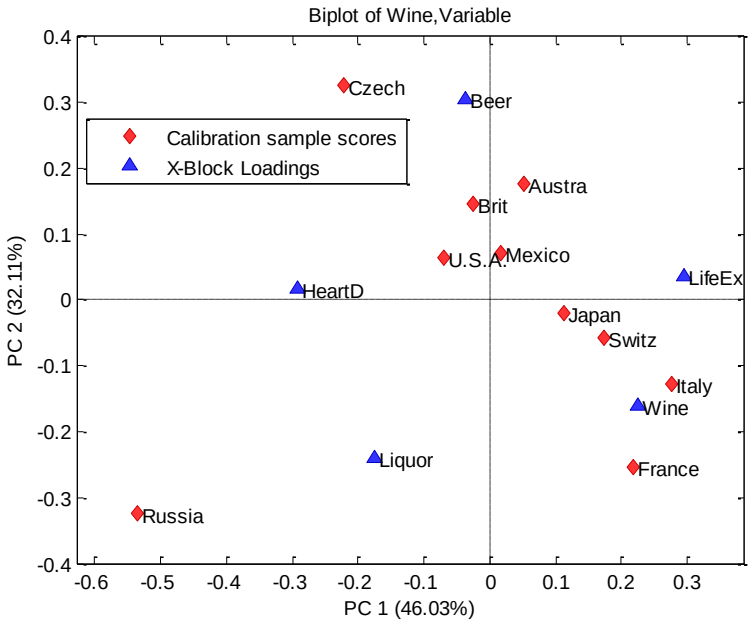

Figure 2-6: Biplot of scores and loadings

2.6.2 Residuals

A Chemometric model is developed using the optimal number of PCs in order to describe maximum amount of process variation in the data set. Thus the remaining nature of the unmodeled information is explained in the form of Residuals Q-statistic or Squared Prediction Error. In other words it is used to determine how well the new samples and variables fit the model. The Q statistic is defined as the sum of squares of the residual values at each variable in each sample of the data set. Keithley et al. (2009) have mentioned a number of advantages of using the Q-statistics for analyses such as monitoring quality control, interferent identification and outlier detection. Using the residual statistics one is able to analyse goodness of fit of the training dataset used to develop a calibration model. Samples with high residuals indicate that the sample may be not extreme however it does not fit the model well. This statistic can be used to identify a new behaviour in the process that is not explained in the reference data used to develop in control model. Such new observation can be identified using SPE as seen in Equation 2-15 for a PCA model and Equation 2-16 and Equation 2-17 for a PLS model. The SPE for the ith sample from the PCA model is calculated as shown in Equation 2-15.

$$SPE = \sum_{j=1}^{p} (x_{ij} - \hat{x}_{ij})^2 \qquad \textbf{Equation 2-15}$$

where p is the number of variables, x_{ij} and $\hat{x}_{ij}$ (the ith sample on the jth variable) are the elements of X (Original Data Matrix) and $\hat{X}$ (Estimated Data Matrix).

For a PLS model however the *SPE* can be calculated for the dependent variables (SPE_y) as seen in Equation 2-16 as well as for independent variables (SPE_x) as seen in Equation 2-17.

$$SPE_y = \sum_{i=1}^{q} (y_{new,i} - \hat{y}_{new,i})^2 \qquad \textbf{Equation 2-16}$$

where q is the number of variables in y data matrix, $y_{new,i}$ are actual values and $\hat{y}_{new,i}$ are the model predicted values.

$$SPE_x = \sum_{i=1}^{m} (x_{new,i} - \hat{x}_{new,i})^2 \qquad \textbf{Equation 2-17}$$

where m is the number of variables in x data matrix and $\hat{x}_{new,i}$ is estimated from reference PLS model.

For a normally distributed data set the control limits for SPE are given as follows (Jackson and Mudholkar, 1979):

$$Q_a = \theta_1 \left(\frac{Z_a \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right)^{\frac{1}{h_0}} \qquad \textbf{Equation 2-18}$$

where,

$$\theta i = \sum_{k=R+1}^{J} \lambda_k^t \qquad i = 1,2,3 \qquad \textbf{Equation 2-19}$$

and

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \qquad \textbf{Equation 2-20}$$

Z_a is the standard normal deviate corresponding to the upper (1-α) percentile, α is the confidence interval, λ_k is the eigenvalue of the residuals, R is the number of principal components retained in the model.

2.6.3 Hotelling's T^2

The Hotelling's T^2 test is centered around the concept of Mahalanobis Distance (MD) which is based on the measurement of distances between observations. The MD can be calculated in

the original variable space as well as in the PC space. In the original variable space the MD can be calculated using the equation as follows:

$$MD_i = \sqrt{(x_i - \bar{x})S_x^{-1}(x_i - \bar{x})^T} \qquad \textbf{Equation 2-21}$$

S_x is the covariance matrix.

The Hotelling's T^2 statistics is used to measure if for retained PCs in a model the variation in the quality variables is greater than the common cause variation. In the PC space The Hotelling's T^2 can be calculated for R number of retained PCs using the Equation 2-22.

$$T_R^2 = \sum_{i=1}^{R} \frac{t_i^T t_i}{S_{t_i}^2} = \frac{t_R^2}{S_{t_i}^2} \qquad \textbf{Equation 2-22}$$

where S_{ti}^2 is the estimated variance of t_i.

The Hotelling's T^2 once computed using Equation 2-22 is then compared to the Chi-squared table with (p-1) degrees of freedom. The Upper Control Limit (UCL) for Hotelling's T^2 is calculated using the Equation 2-23 (De Maesschalck et al., 2000).

$$T_{UCL}^2 = \frac{(n-1)^2}{n}\beta_{(\alpha::\frac{p}{2},\frac{n-p-1}{2})} \qquad \textbf{Equation 2-23}$$

2.6.4 Leverage

Leverage is another statistic similar to Hotelling's T^2 that has a large influence on parameters such as response, regression co-efficient and standard error (Davies, 1995). Equation 2-24 is used to compute the leverage calculations which is similar to MD equation.

$$h_i = \frac{1}{n} + \frac{(MD_i)^2}{n-1} \qquad \textbf{Equation 2-24}$$

It can be seen from Equation 2-24 that leverage (h_i) is directly proportional to Hotelling's T^2 statistics. A sample with high leverage would be significantly influential on the model and could be likely outlier. Such samples need to be further assessed by analysing the raw data and if necessary remove from the model.

2.6.5 Influence Plot

This is a plot of Q-residual versus Hotelling's T^2 or Leverage and represents two different types of outliers. With the residual statistics on the y-axis it describes the distance of the sample from the model while Hotelling's T^2 or Leverage on the x-axis describes how well does the sample fit this model. For example in Figure 2-7 batches with high residuals such as B38, B40 and B39 are not very well described by the model while batches lying to the right of plot such as B30 and B50 are described well by the model however they are more influential. A sample with high residual and high Hotelling's T^2 is a dangerous outlier as it is not described by the model as well as is influential in the model for example B35 in Figure 2-7.

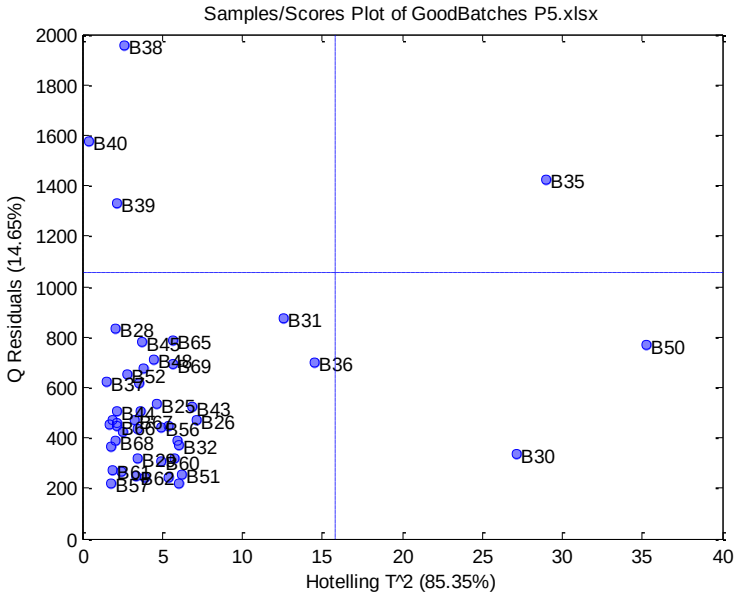

Figure 2-7: Example Influence Plot

2.6.6 Contribution plots

With the advancement and abundance of online sensors in the market there is a wealth of historical data available to engineers. Successful application of multivariate statistical methods such as PCA and PLS have simplified fault detection by reducing the heavily correlated data available from the sensors to a smaller set of uncorrelated variables (Kherkhof et al., 2013). Multivariate charts are able to identify process deviation but they are not able to detect the cause of process disturbance once it has been detected using the MSPC techniques. Thus further analysis into the model to analyse a particular variable or set of variables that may be responsible for a process deviation is performed using the contribution plot. Contribution plots are different to the loadings since loadings represent the variability across the data set being analysed while contributions look at unusual causes in the underlying data which may be particular to process variables that were peculiar for the process behaviour

(Martin, 2014). A high contribution for the process variables is indicative of problem with those variables.

The principle component scores can be written as a weighted sum of the process variables given by Equation 2-25.

$$t_{id} = \sum_{j=1}^{k} x_{ij} p_{jd} \qquad \textbf{Equation 2-25}$$

where p_{jd} is the loading for variable j^{th} variable at time instant k.

Contribution plots of process variables are also measured for D-statistic or Hotelling's T^2 and the Q-statistic when there is process disturbance. The D-statistic explains the systematic part of the process variation while the Q-statistic the residual part of the process variation. In case of process deviation one of these statistics will be out of the pre-defined confidence limits although the model could still be valid. Contribution of variables to these statistics should be investigated to identify the cause of process deviation.

Out of control Q-statistics would mean an unknown event has been found in the process. The contribution of process variable j^{th} at time period k to this event is calculated using Equation 2-26.

$$c_{jk}^{SPE} = \left(e_{new,jk}\right)^2 = (x_{new,jk} - \hat{x}_{new,jk})^2 \qquad \textbf{Equation 2-26}$$

where $x_{new,jk}$ and $\hat{x}_{new,jk}$ are the actual and predicted values of j variables. Plotting all the contributions for c_{jk}^{SPE} one is able to identify which time in a batch and for which variables process deviation occurred (Westerhuis et al., 2000).

The contribution to D-statistic for each process variable have been summarised by Nomikos (1996):

$$c_{jk}^{T^2} = \sum_{r=1}^{R} S_{rr}^{-1} x_{new,jk} p_{r,jk} t_{new,r}^{T} \qquad \textbf{Equation 2-27}$$

Equation 2-27 calculates the contribution of every element $x_{new,jk}$ to the Hotelling's T^2 statistic which is summed over the retained PCs. $S_{rr}^{-1}$is a diagonal matrix, $x_{new,jk}$ is the vector of new observation, $p_{r,jk}$ is the loading vector for j variable and $t_{new,r}^{T}$ is the score vector for

the new observations. A diagonal matrix is a square matrix where all the non-diagonal elements are zeros.

The control limits for contribution plots of D-statistic cannot be calculated using the F distribution but have to be obtained using the jack knife method (Westerhuis et al., 2000). Since only high contribution force the D-statistics out of confidence bounds only the upper control limit needs to be calculated. The UCL for contribution of every process variable is calculated as the mean of the contribution plus three times standard deviation of the contributions at each time.

Contribution plots aim to address one of the major weaknesses in the MSPC projection method where once an abnormal batch is identified there is no information available for the cause of a new event or disturbance once the process is out of control limits. By plotting the contributions for the scores or the statistics one is able to diagnose the special event which can allow operators to track and control the process variables leading to disturbance in the process.

2.7 Multiway Techniques – Data Unfolding

The advent of computers led to vast amounts of process data being collected and stored for purpose of process improvement and optimisation. The concept and methods of SPC gave an excellent opportunity to use this data to monitor processes such that process performance would remain in a state of SPC. SPC charts mentioned in Section 2.6 further allowed detection of any event that could cause process disturbance thus enabling long term process improvements to the final product quality. However the application of SPC was limited to univariate monitoring of process variable. Least squares regression technique such as PLS provided the much needed answer to analyse multiple variables at a time. Nomikos and MacGregor (1994) attribute the success of PCA in number of different areas due to the NIPALS algorithm (Geladi and Kowalski, 1986) which facilitated simple, fast and effective way to extract principal components in a sequential manner. PCA has been adopted in a number of scientific application such as chemistry, biology, geology, process and quality control, image analysis to name a few. In many cases such as image analysis and batch processes the data is available in the form of three way arrays. For e.g. in a typical batch process the data consists of j variables measured for k interval times which is collected for i number of batches.

The monitoring of batch processes using the classic PCA and PLS is difficult due to three dimensional nature of the process data. Data from a batch process consists of I batches measuring J variables over K intervals of time. In order to monitor such a three dimensional process the data matrix has to be converted to lower dimensional by unfolding the data matrix. The three dimensional data matrix can be unfolded in three possible ways: 1) Time wise 2) Batch wise and 3) Variable wise unfolding. Time wise unfolding analyses the variability among the samples, batch wise unfolding looks at variability in batches and variable wise unfolding analyses the variability among batch variables (Lee et al., 2004).

MPCA encompasses all the information about the batches onto a lower dimensional space. Like PCA, MPCA decomposes the matrix X as the sum of the product of scores vectors and loading vector plus the error or residual E as shown in Equation 2-1. Thus using historical batch data MPCA can monitor progress of the current batch and detect any deviation from expected trajectory.

For some other exceptional situation other multiway methods have been discussed such as the Tucker model (Geladi, 1989), PARAFAC model (Harshman and Lundy, 1994, Bro, 1997), the canonical decomposition (Carroll and Chang, 1970), three mode factor analysis and the tensor rank method (Sanchez and Kowalski, 1990). These methods are tri-linear approaches to monitor three-way matrices with the intention to retain the original dimension of the data matrix.

2.7.1 Batch-Wise Unfolding Approach

Nomikos and MacGregor (1994) unfolded the three way matrix in such a way that the vertical slices (I and J) are placed side by side to form a two dimensional matrix of the form I x JK. Each horizontal slice (J x K) represent the time history/trajectories for all variables of a single batch (I) and every vertical slice (I x J) is a matrix representing all the variables of all the batches at a particular time interval (k).

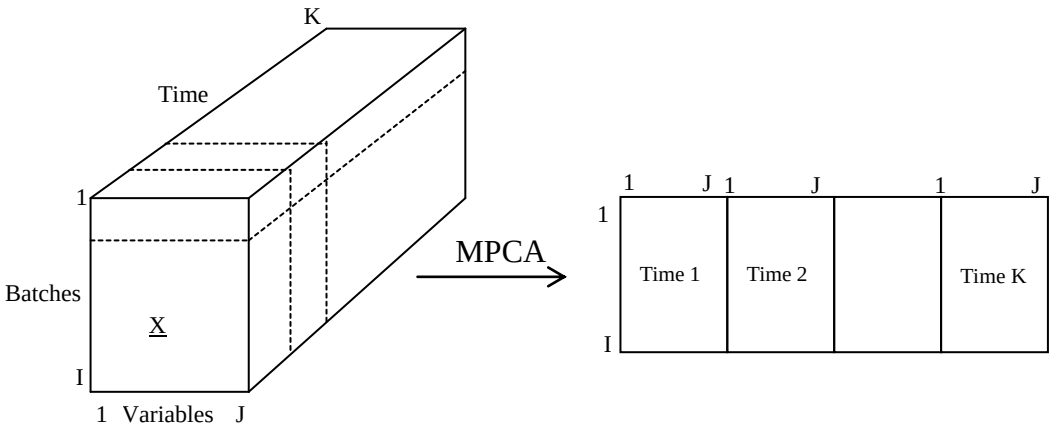

Figure 2-8: Batch-Wise Unfolding Approach

Multiway-PCA (MPCA) can be used for the analysis of historical batch data to understand the major source of batch to batch variation. Good quality batches can be used for nominal model development and any future batches could be monitored by analysing the correlation structure and to see if they lie within the defined limits.

2.7.1.1 Analysing historical batches - Offline and Online

A MPCA model is developed using a set of historical good batches. A good batch can be defined as one which follows expected process trajectory without indicating possible deviation that would in turn affect the final product quality. The loadings matrix obtained from the model developed can then be used to test for any unusual event by predicting the t-scores and residuals for the new batch.

Once a historical set of three-way batch data is obtained it is first unfolded to form a 2 dimensional matrix. Pre-processing such as centering and scaling is then applied to remove nonlinear and dynamics elements thus eliminating time trajectory of each variable present in the data set. A nominal PCA model is developed with control limits defined for control charts. For new batch projected on this model if the t-scores are closer to the origin with small

residual it would mean the operation of new batch is similar to the reference model developed.

However if a process variation is observed in the new batch it would be more helpful if the deviation is identified online than offline using historical data base. In order to test the data sequentially in time one needs to have access to the entire batch history. But for online monitoring this is not possible as the batch is still incomplete and data is only available from the start of the batch to the current time. Nomikos and MacGregor (1995) have discussed three methods to be able to estimate unknown data x_{new} between the current time and the end of batch.

The first method assumes that the batch would continue operating at the desired level and the future observation would be in line with the mean trajectories of the reference data base. Thus the unknown data x_{new} is filled with zeroes assuming that the batch will operate normally for the remaining duration. The major disadvantage of this approach is predictive ability at the start of batch run. The second approach presumes that the batch will continue to deviate in a similar manner as present and retain similar SPE for the rest of the batch. A rather pessimistic approach as compared to the first method as the unknown part of x_{new} is filled with the offset values at time point k. The third and final approach utilises the ability of PCA to handle missing data by regarding the future unknown observations of the batch as missing values. The loadings data matrix that is available up to time k can be used to predict scores and residuals given by Equation 2-28

$$t_{R,k} = (P_k' P_k)^{-1} P_k' x_{new,k} \qquad \textbf{Equation 2-28}$$

where P_k is the loadings matrix for all the retained PCs upto time point k and x_{new} is the vector containing measurement known up to time point k. The orthogonal property of the loadings vectors make the term $(P_k' P_k)^{-1}$ an identity matrix. Although this method has been reported to perform better than the other methods mentioned previously it could predict unusually large scores with limited information available at the start of a batch. Depending on the nature of the process operation one can use either one of the methods or even combination of methods in order to achieve the desired predictive capabilities with the MPCA model.

Once the scores and SPE have been predicted using either of the methods the batch will be monitored using the t-scores chart and SPE charts. The Hotelling's T^2 for a new batch is calculated using the Equation 2.29.

$$T_k^2 = (t_{new,k} - \bar{t}_k)^T S_k^{-1} (t_{new,k} - \bar{t}_k) \quad \textbf{Equation 2-29}$$

where $t_{new,k}$ is the scores of the new batch up to time point k, $\bar{t}_k$ is the means of the columns of the score matrix T_k and S_k is the covariance matrix of T_k. The control limit of Hotelling's T^2 for independent t scores is derived under the assumption of normality given by Equation 2-30

$$T_{UCL}^2 \sim \frac{R(I^2-1)}{I(I-R)} F_{R,I-R,\alpha} \quad \textbf{Equation 2-30}$$

where I is the number of nominal batches, R is the number PCs retained in the model and $F_{R,I-R,\alpha}$ is the critical value of the F-distribution with R and I-R degrees of freedom for a significance level α.

The SPE for new batch is calculated using the Equation 2-31.

$$SPE_k = \sum_{j=1}^{J} e_{jk}^2 \quad \textbf{Equation 2-31}$$

where e_{jk} is the prediction error at time point k. Box (1954) calculated the control limits for SPE given by Equation 2-32.

$$SPE_k = g x_h^2 \quad \textbf{Equation 2-32}$$

Although the weight g and the degrees of freedom h can be estimated quickly equating the mean and variance of $g x_h^2$ to the sample mean (m) and variance (ϑ) of the SPE sample at each time k. The control limit at significance level α for time interval k are given by:

$$SPE_\alpha = (\frac{\vartheta}{2m}) x^2_{\frac{2m^2}{\vartheta},\alpha} \quad \textbf{Equation 2-33}$$

where $X_{2m}{}^2/\nu,\alpha$ is the critical value of the chi-squared variable with $\frac{2m^2}{\vartheta}, \alpha$ df at significance level α.

2.7.2 Variable-Wise Unfolding Approach

The N&M approach or the batch wise unfolding approach discussed in Section 2.7.1 has a few drawbacks especially when a batch has to be monitored in real time. This is because in the batch-wise unfolding approach data is available only upto the current time and for online

monitoring the test data should be completed until the end of a batch (Lee et al., 2003). Also the batch-wise unfolding approach requires the batches to be of equal lengths which is not always possible within industry. A number of ways have been discussed in Section 2.7.1.1 which can estimate the trajectory till the end of a batch. This section looks into another approach introduced by Wold et al. (1998) which aims to preserve the direction of variables when unfolding a three-way matrix.

The data is unfolded as shown in Figure 2-9 where each row consists of data for all variables for an individual batch at a particular time point. The data is then auto-scaled to mean zero unit variance where scores matrix describes the mean trajectory of the evolving batch. The model equation for the unfolded data is similar to PCA.

$$X_{jk} = \sum_{k=1}^{r} T_r P^T + E \qquad \textbf{Equation 2-34}$$

Once the scores matrix is obtained they are re-arranged variable wise as shown in Figure 2-9. Confidence limits of +/- 2 and 3 standard deviation are then calculated for each time point.

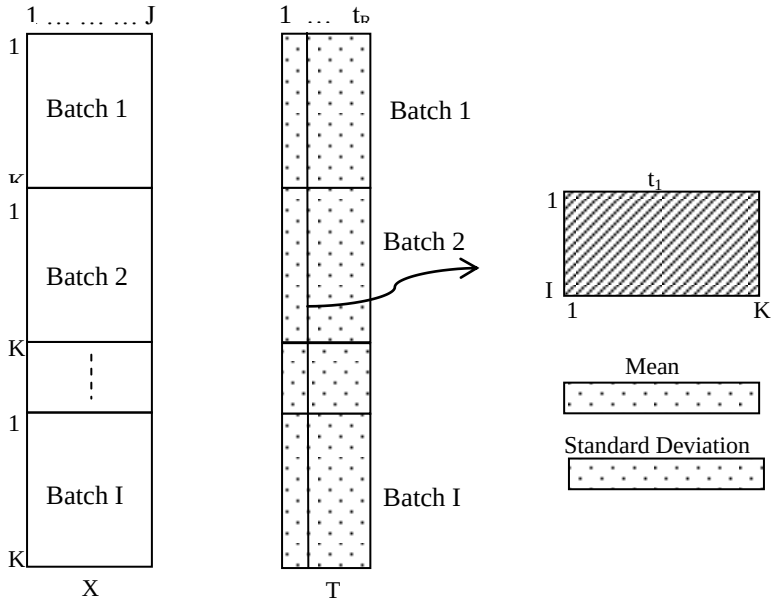

Figure 2-9: Variable-wise unfolding and re-arrangement of scores matrix

Thus MPCA is a technique which is statistically and algorithmically consistent with PCA achieves the same goals and benefits as the latter. It overcomes the drawbacks poised by PCA method for dealing with three-dimensional batch process datasets.

2.7.3 Multiway Partial Least Squares

Batch processes are non-linear in nature which limits the application of PLS which is a linear regression method (Marjanovic et al., 2006). The technique of Multi-way Partial Least Squares (MPLS) is similar to MPCA but MPLS includes the quality or effect variables Y. The Y block is two way matrix I x 1 and does not require unfolding. Once the data matrix X is unfolded both the X and Y blocks are then normalised to remove any trajectories that they follow. PLS is then applied on the data to develop a regression model predicting the quality variables based on the measured variables from the available batches.

2.8 Chapter Summary

This chapter introduces various methods and techniques that are parts of MSPC. These methods form the basis for the research work that has been undertaken in the two projects discussed further in the thesis. PLS regressions methods have been applied in Chapter 3 and Chapter 4 which discuss PAT software solution developed by ABB. The PCA and MPCA methods have been used in the second project to analyse crystallisation issue in the polymerisation reactors in Chapters 5 and 6.

Project 1: Business Productivity Improvement through the Application of Analyser Device Integration

3 Chapter 3 - Process Analytical Technology – a tool for continuous improvement

3.1 Chapter Overview

This chapter looks at the concept of PAT introduced by the FDA in 2004 in order to encourage continuous improvement especially within the pharmaceutical industry. A number of tools were introduced to encourage the application of this concept within the industry. This chapter looks at how these tools have been used with respect to the ADI software.

Section 3.2 gives a brief introduction about the PAT initiative and its application to current date in various industries.

Section 3.3 discusses the PAT tools that were introduced by the FDA and how these tools have been used as the basis for the developing the unique software solution for implementation of PAT.

Section 3.4 addresses the data management challenges faced by the industry that need to be addressed to make the concept of PAT into a practical solution.

Section 3.5 gives the summary of the chapter.

3.2 The concept of Process Analytical Technology

The concept of PAT evolved from the older concept of Process Analytical Chemistry (PAC) that has existed for many years in the chemical and food and beverage industry. The main objective of PAC was to enable better process understanding in order to discover cost effective, traceable and environment friendly ways to achieve the required final end product quality. The ultimate goal of PAC was to improve the production efficiency, improve safety, reduce waste and achieve consistent production. Baughman (2005) has described the successful implementation of the PAC within the petrochemical industry. The author claims that many of the process analytical instruments that are currently used in various industries were originally developed in the oil and petrochemical industries to enable high throughput by on-line measurements.

The US FDA popularised the concept of PAT to increase process understanding to enhance control and enable continuous improvement of the manufacturing process.

The FDA defines PAT as:

1. A system for the analysis and control of manufacturing processes based on timely measurement of critical quality parameters and performance attributes of raw materials and in-process materials

2. A process to ensure acceptable end-product quality at the end of the process (CDER, 2013).

The FDA also states that PAT involves:

1. The optimal application of process analytical chemistry (PAC)

2. Feedback process control strategies

3. Information management tools and/or product-process optimization strategies for the manufacture of pharmaceuticals (CDER, 2013)

This concept of PAT was in line with the FDA's Quality by Design (QbD) approach which aimed to move pharmaceutical manufacturing processes from a rigid regulatory approach to a more flexible science and risk based approach. The QbD approach aimed to build quality in the process by defining a flexible and acceptable design space for process operation by monitoring the CQAs and Critical Process Parameters (CPP's). A CQA is defined by the US FDA as physical, chemical or microbiological property that should be controlled within the defined limits to ensure the quality target product profile is achieved. Quality risk assessment (ICH Q9) can aid in linking the quality attributes of a product to CPP's that can have an impact on the product quality. As stated in the ICH-Q8 (2009) guideline this understanding in processes can be obtained by formal experimental designs, prior knowledge of the process and PAT. The successful implementation of PAT would require tools that can provide effective and efficient means of gathering data to allow for continuous improvement and development of risk mitigation strategies. Rathore (2014) has indicated that although the application of QbD has been somewhat successful within the target industry the adoption of PAT has not met the required expectation. This may be attributed to the fact that cost and risk involved in the successful implementation of PAT is too high and needs to be addressed by academia and industry to enable real time decisions for better quality control.

According to van den Berg et al. (2013) the food industry also faces strict regulatory demands in terms of quality, safety and traceability. The author deems this to be major challenge considering the high level of variation that exists in biological processes and believes that PAT would also revolutionise industrial quality control in food processing. Thus, the concept of PAT encourages shifting process control operations from a typical post-problem or feedback control strategies to during-process or model predictive control strategies as seen in Figure 3-1.

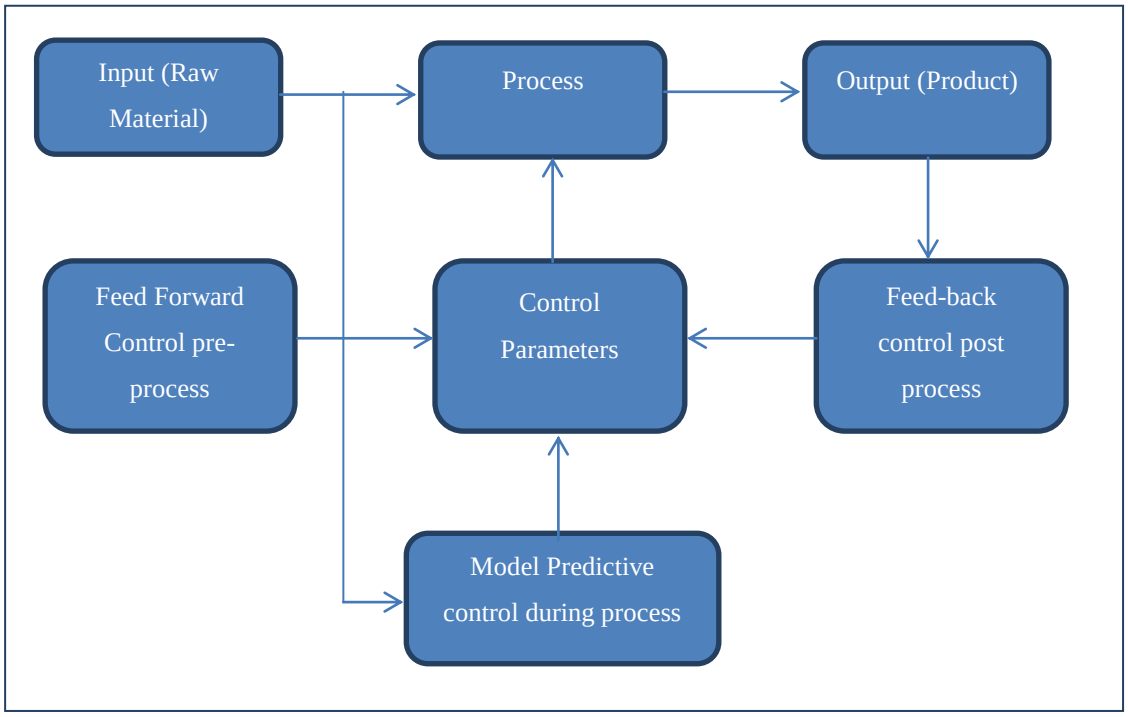

Figure 3-1: Schematic View of Process Control Strategies in manufacturing (van den Berg et al., 2013)

In order to ensure successful implementation of the PAT, the FDA introduced a number of principles and tools. These tools provide means to control 1) CQAs which are the product properties that define a good product and 2) CPP's which are product parameters having critical effect on the final product quality for example; monitoring the concentration of sugar at the start of fermentation for alcohol would be the CPP that would affect the final alcohol content which is a CQA. The PAT tools which enable the process understanding of pharmaceutical manufacturing and quality assurance are explained in Section 3.3

The application of PAT within industry is a four stage process:

1. **LEARN** – Gather data from analysers and process instruments to better understand the process

2. **PREDICT** – Create models to predict Product Quality Attributes. Use the predicted value to advise operators when quality deviation occurs or when a process is complete. Validate with lab tests

3. **CONTROL** – Develop feedback, feed forward models for controlling the process to achieve the desired quality.

4. **CONTINUOUS IMPROVEMENT** – Optimise process based on the data collected

3.3 PAT Tools

The FDA has defined four PAT tools such as 1) Multivariate Methods 2) Process Analysers 3) Process Control and 4) Continuous Improvement that can be used for the successful application of this technology within industry. This section of the report looks into how each of these tools were used within the scope of the ADI project with the customer.

3.3.1 Multivariate Tools for Design and Analysis

Industrial quality control has changed over the past few years with the advancement in sensor technology and its integration with the data analysis technology which is more commonly known as chemometrics. Chemometric tools establish relationship between different variables through the application of mathematical or statistical methods. Industrial processes and products are multi-factorial in nature and traditional methods of analysing one variable at a time are gradually progressing towards multivariate analysis to establish correlations between the product and the process. A number of these multivariate methods used in industry to monitor and control processes were explained in detail in Chapter 2.

Multivariate methods are used to analyse large amounts of data sets by extracting meaningful information. This is normally done by plotting the data and enabling visual interpretation of hidden information within the data sets. Multivariate analysis in conjunction with Design of Experiments, Response Surface Methodologies, Process Simulation and Pattern Recognition tools can be used to enable real time release of products. The next section of the report describes the multivariate tools used in the ADI project.

3.3.1.1 Hierarchical Modelling

For the ADI project an extension of PLS model which is a Hierarchical PLS model was applied using CAMO's Unscrambler software. Hierarchical modelling is not a method of analysis but a combination of multivariate models joined together using logic statement in order to arrive at a single unique result. These models can be used for projection, classification as well as regression methods. A hierarchical model can be developed to

classify samples in a stepwise manner or when developing non-linear models over large concentration range (CAMO, 2013). A more detailed explanation of the Hierarchical PLS modelling method that was used in the project is explained in the next section.

Overall Workflow

In order to use a prediction model for at-line testing a user was required to develop and validate a set of classification or regression rules so as to understand the boundaries/ambiguities. The following steps are involved in developing a final hierarchical model.

1. Develop a global multivariate model to understand if there are any ambiguities or non-linearities in the system.

2. If ambiguities or non-linearities exist, develop sub-models that can handle these. If there are subclasses, also develop and validate models to handle such situations.

3. Validate all models against a suitable validation set to ensure that the results project/predict/classify as expected.

4. Develop the hierarchies as determined by the results of the individual models during the training stage and enter the logic required to take the model to the next level. Also define the conditions that will result in a premature termination of the hierarchy. These will be defined as alarm conditions.

5. Alarm conditions will be defined as, a. Primary: These will result in termination of the method b. Secondary: These will allow the hierarchy to proceed however, the results that do not meet some predefined criteria will be marked for investigation.

3.3.1.2 Setup

HM set up within Unscrambler environment works as a cascading tree of decision making. It is expected that all projection, prediction and classification models generated in the Unscrambler can be used for HM development. The HM module supports up to 10 levels of hierarchy and multiple models could be included at each level.

Based on the output from the previous level one or more models could be defined within each level. Alternatively, if the output is satisfactory and reported, or it may be ambiguous or out of limits, in which case a warning can be displayed or the HM be told to exit. This behaviour is completely at the hands of the user, who has to make sure that the provided sequence of steps and the limits used are sensible.

Also, for each model within each level, an ordered list of logical conditions are specified by the user and executed in an IF-ELSE manner. This means that if the first condition is satisfied, any remaining conditions will not be executed. It follows that the order of the conditions is important. If for instance condition 1 finds that the predicted response is out of limits, a condition 2 testing for e.g. leverage of the predicted sample will never be executed.

3.3.1.3 HM Prediction

The Unscrambler software has another feature called the 'HM Predict' which allows applying the hierarchical model with a complete sequence of multivariate model and reporting of the results. This enabled for testing the model in Unscrambler before it is imported into ADI. The detailed report and example situations of the models developed for the ADI project have been discussed further in Chapter 4 Section 4.4.1.

3.3.2 Process Analysers

Predominantly process analysers were mainly involved in taking univariate measurements of physical variables such as pH, temperature, pressure etc. However over the past few decades an increased appreciation for the value of collecting data to measure quality of the product during the process has led to ground breaking development in this area. Bakeev (2010) defines process analysis as a field deployable instrumentation for real time analytics and chemometrics for monitoring the CQAs. Process analyser measurements that contain the information related to the biological, physical or chemical attributes could either be at-line, online or in-line measurements

The difference between these measurement techniques is described as follows:

At-line: Measurements where sample is removed, isolated, and analysed in close proximity to the process stream

On-line: Measurements where the sample is diverted from the manufacturing process, and may be returned to the process stream

In-line: Measurement where the sample is not removed from the process stream and can be invasive or non-invasive

Spectroscopic sensors have been popular in the industries adopting the PAT approach. They allow establishing a statistical relationship between the measured signals and reference analysis on a number of factors which affect the final product quality.

3.3.2.1 Types of Spectroscopic Analysers

van den Berg et al. (2013) suggests the most commonly used spectroscopic techniques for process monitoring are the Ultraviolet-Visible (UV-VIS) absorption, Near Infrared (NIR), Infrared (IR) absorption, Raman scatter and Excitation Emission Matrix (EEM). Out of these five the application of UV-VIS and EEM are somewhat limited as it can be only used to analyse a smaller select group of molecules. The main difference between NIR and MIR spectra is the difference in wave numbers that causes the molecules to respond differently. NIR spectroscopy involves the studying of absorption of compounds usually O-H, C-H and S-H bonds in the NIR range which lies in the electromagnetic spectrum ranging from 10000 – 4000 cm^{-1}. Broad overlapping peaks and large baseline variations make the interpretation of NIR spectrum very complex. Since NIR spectra penetrate the sample more than MIR spectra the NIR bands are 10-100 times less intense that MIR bands. This allows for direct analysis of highly absorbing material and porous samples with no need for sample preparation. It is difficult to identify IR frequencies with a specific chemical group in NIR range. Robust calibration and statistical techniques are therefore required for precise analysis of NIR spectra. Multivariate statistical techniques such as PCA and PLS previously discussed in Chapter 2 can be used for spectral analysis making it simple and easier to interpret and draw meaningful conclusion to a complex spectra. The PAT desired requirements for a spectroscopic sensor are 1) the sensor should be able to measure/predict critical control parameter of interest, 2) the measurement frequency of the sensor should be high in order to be able to analyse the rapidly changing process and 3) ideally the sensor should be non-invasive and guarantee product quality and integrity.

Presently only NIR spectroscopy is able to satisfy all of these requirements. It is also affordable and off the shelf technology making it the predominant sensor technology used in the manufacturing industry. As discussed in Chapter 4, the customer uses FTIR analysers coupled with ATR sampling technique for at-line sample testing. The next section looks more into detail in the working of FTIR analysers.

FTIR Analysers

The simplicity, sensitivity, versatility, speed of analysis and high throughput has led to FTIR spectroscopy being applicable not only to chemists and spectroscopists but also a number of specialists and non-specialists from various backgrounds. FTIR analysers are single beam and collect background and sample measurements at two different times. The actual spectrum of the sample is a ratio between the single beam spectrum and the background spectrum that is

obtained without the sample. Figure 3-2 illustrates the sequence of steps involved in obtaining a sample spectrum. Despite of a number of advantages showcased by FTIR spectroscopy its application is limited due to its other inadequacies. Subramanian and Rodriguez-Saona, 2009 have reported some of the shortcomings with the FTIR spectroscopy method such as 1) its inability to detect atoms and monoatomic ions, elements and inert gases, 2) complications due to overlapping peaks in biological samples and 3) signal masking important signals especially for biological samples that contain water which has a strong absorption band for FTIR method thus requiring extensive sample preparation to reduce the water effect.

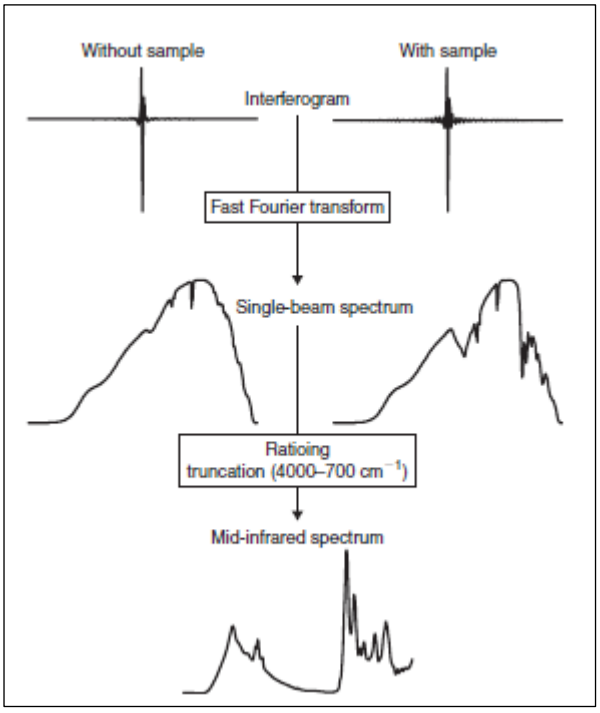

Figure 3-2: Illustration of how MIR spectrum is obtained from InterferogramSampling Techniques (Subramanian and Rodriguez-Saona, 2009)

The customer has been using two types of at-line FTIR analysers namely Thermo Nicolet iS10 and Thermo Nicolet 350. Figure 3-3 illustrates a typical Thermo Nicolet iS10 analyser used by the customer for at-line measurements. The standard applications of this type of analyser have been to analyse biodiesel blending, polymers, plastics pharmaceuticals and food industry. Samples are measured directly through vials and materials are characterized quickly and easily.

Figure 3-3: Thermo Nicolet iS10 (Thermo Fisher Scientific, 2015)

A number of sampling techniques such as transmittance, reflectance, Attenuated Total Reflectance (ATR) and diffuse reflectance exist for FTIR spectroscopy (Thermo Fisher Scientific, 2015). Figure 3-4 shows a schematic diagram of various FTIR sampling techniques currently used in industry. The ATR sampling technique used by the customer has been further discussed.

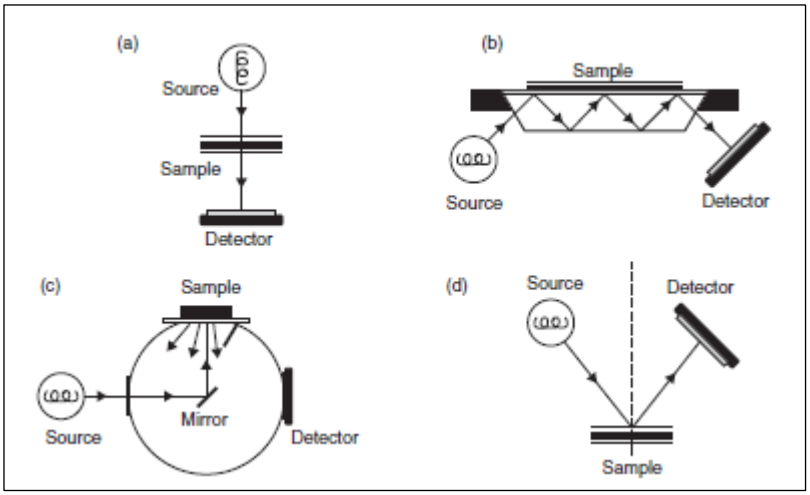

Figure 3-4: Schematic diagram of various FTIR sampling techniques: (a)transmission, (b) attenuated total reflectance. (c) diffuse reflectance in an integrating sphere, and (d) specular reflectance (Subramanian and Rodriguez-Saona, 2009)

Attenuated Total Reflectance (ATR)

ATR is one of the most common sampling techniques with little or no sample preparation required for speeding up the process of sample analysis. In order that the technique is successful the sample should be in direct contact with the ATR crystal and the refractive index of the crystal must be significantly greater than that of the sample in order for the light to be internally reflected in the crystal. As described by PerkinElmer (2005) the ATR

accessory functions by measuring the changes that occur in a totally internally reflected infrared beam when the beam comes in contact with a sample (Figure 3-5). As the infrared beam is directed onto the dense crystal with a high refractive index the internal reflectance creates an evanescent wave that extends beyond the surface of the crystal only a few microns onto the sample. In the region of infrared spectrum where the sample absorbs the energy the evanescent wave is attenuated or altered which is then passed back to the IR beam which exists at the opposite end of the crystal and is passed to the detector in the IR spectrometer. Some of the usual crystals used for ATR analysis include ZnSe, Ge and diamond. Compared to ZnSe with spectral range down to 550 cm^{-1} and Ge with spectral range down to 650 cm^{-1} diamond has wider spectral range down to 200 cm^{-1} or less. However the high cost associated with the use of diamond has limited its use within the industry.(Thermo Fisher Scientific, 2015)

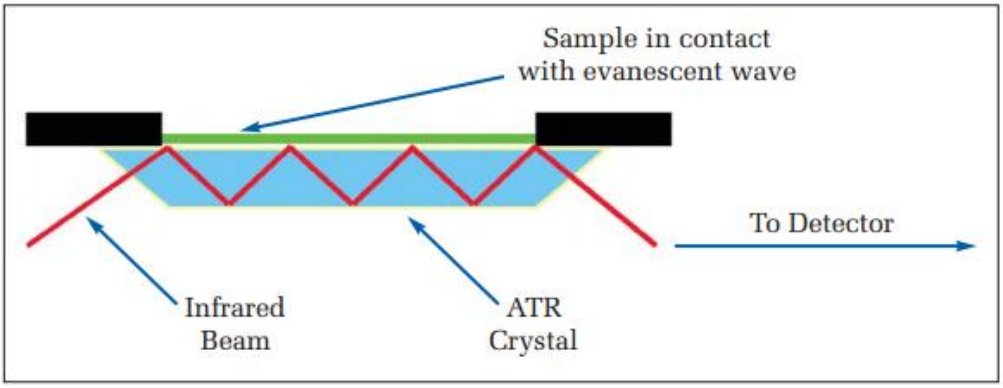

Figure 3-5: Multiple reflection ATR system ATR-FTIR (PerkinElmer, 2005)

The Thermo Nicolet iS10 analyser with the ATR accessory can be seen in Figure 3-6. This accessory has a durable, high performance diamond ATR that provides high quality spectral data used to verify materials.

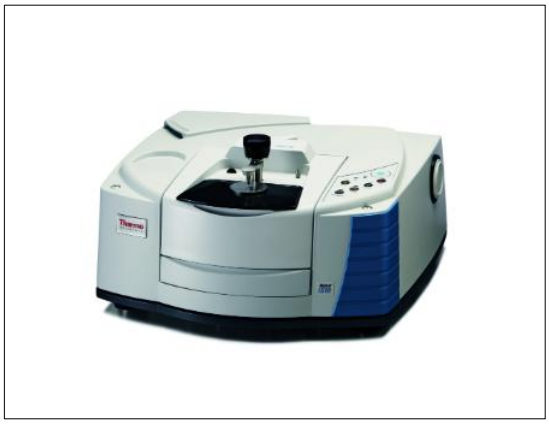

Figure 3-6: Nicolet iS10 ATR accessory (Thermo Fisher Scientific, 2015)

3.3.3 Process Control

Process monitoring and control strategies which intend to monitor the process in order to maintain it to desired state should be able to accommodate attributes of input materials, the ability and reliability of process analysers to measure critical attributes, and the achievement of process end points to ensure consistent quality of the output materials and the final product (FDA, 2004). Ganguly and Vogel (2006) indicate the importance of monitoring, controlling and reporting of CQAs so as to implement the concept of PAT successfully. Whether the measurements are taken online, inline or at-line integration with process control and automation is necessary to ensure so that right data is in the right place at the right time to facilitate manufacturing decision making.

The automation pyramid as seen in Figure 3-7 describes the various layers and functions of automation starting from plant floor and actuators that can extend all the way upto Manufacturing Execution Systems, (MES) and Enterprise Resource Planning (ERP) level that have been described in Sections 3.4.2 and 3.4.3.

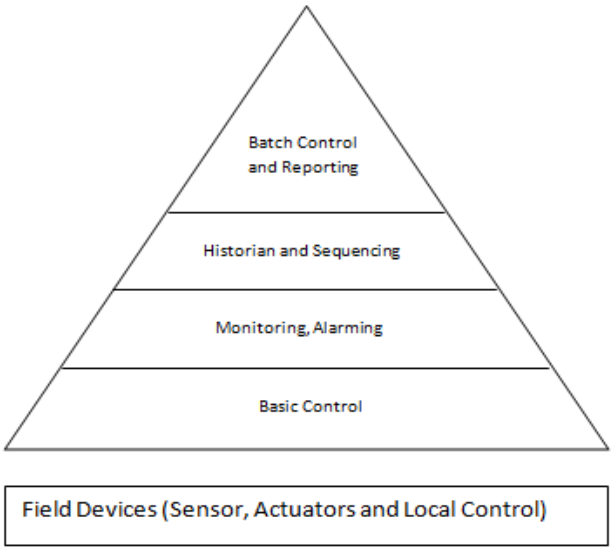

Figure 3-7: Automation Pyramid (Ganguly and Vogel, 2006)

3.3.4 Continuous Improvement through Data Management

FDA encourages continuous improvement of processes through the collection and analysis of data that can justify the post approval changes. They believe data acquisition should be further supported by information technology systems which are valuable for the manufacturers. Once the data has been collected a number of opportunities could then be identified to improve the usefulness of the available product and process knowledge. This available knowledge is most beneficial only when there is better understanding of the relevant multifactorial relationships

and there are means to evaluate these relationships. FDA believes that today's information technology infrastructure makes development and maintenance of this knowledge base practical.

ADI platform providing PAT and automation solution is based on ABB's 800xA control system. It provides an integrated environment for ease of engineering, data and process visualisation, data management, multivariate advanced SPC and enterprise connectivity. Figure 3-8 illustrates ABB's integrated PAT solution based on the backbone of 800xA control system.

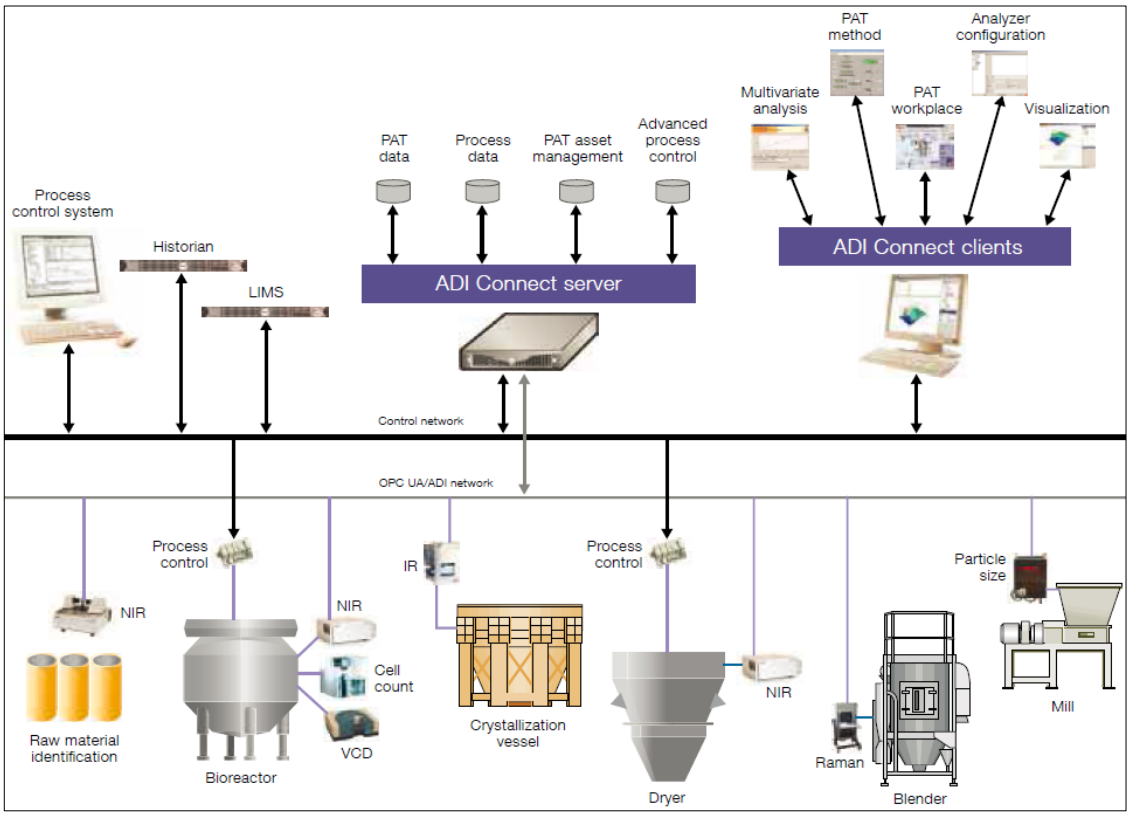

Figure 3-8: ABB's Integrated PAT solution

3.4 Data Management Challenges

The FDA's PAT initiative encourages chemical and pharmaceutical industries to adopt a risk based approach by promoting real time release of quality in process and final product based on process data. For real time release the CQAs as well as the CPP's for products can be monitored and controlled using direct or indirect process analytical methods. The FDA considers real time release of product comparable to alternative analytical procedure for the final release of products.

However with the ability to measure PAT data online comes an increase in real time measurements being made which in turn can lead to data overload. This coupled with the fact

that the data sources often produce differently formatted data using a variety of data export tools makes it a major challenge for those interpreting the data. Once collected the data is often comprised of large arrays which must be further processed to access the beneficial or useful data within. This enables a better understanding of the inter-relationship of process parameters affecting the final product quality which provides the opportunity to control the quality around variations in the process.

The ability to access the right information at the right time is imperative to enable real time release, the setting of product specification based on variability observed in the process, investigating out of specification product and further validation of the process. A cost estimate of inefficient use of the available data has been pointed out by Bakeev (2010). For a single product the annual costs are estimated as follows:

Capacity underutilization – US$25-50M

Lost batches – US$24-28M

Delayed market entry – US$10-20M

Supply chain overburden – US$4-8M

Regulatory actions - >US$500M

These cost estimates are mainly associated with the manufacture of pharmaceutical products which are high value goods.

Data collected from an analyser could be discrete, intermittent or continuous. Irrespective of the dimensionality of the data if it is easily accessible it can then be utilised to optimise and improve our understanding of the process and potentially lead to fewer non conforming products. There may be further benefits in the supply chain by being able to use varying qualities of raw materials. Global standardisation of plant design and collection of data provides a richer data source and a better pay back time for the investment made which in turn facilitates more timely and accurate corporate decisions.

Over the last few years there have been number of questions around how to manage multiple analytical platforms predicting different quality parameters to control, lab, quality and enterprise level systems? How to manage the prediction models needed to predict the quality parameter and for checking the integrity of the analytical device? And how to do all of this in a joined up approach over multiple geographical locations?

A good start is to have standards which define how analytical device will communicate their data and how they will communicate with model prediction packages such as CAMO Insights or Umetrics Simca-Q. In 2013 the OPC foundation released its ADI (ADI) standard through the collaboration of global automation, hardware and software suppliers. The standard defines a single model for analyser vendors to standardise their interface to other software packages.

'The OPC foundation is dedicated to ensuring interoperability in automation by creating and maintaining open specifications that standardise the communication of the acquired process data, alarm and event records, historical data and batch data to multi-vendor enterprise systems and between production devices' (Foundation, 2013). To see how this data is fully integrated with other higher level systems it is necessary to look at a specific implementation of the OPC ADI standard and analyser technology.

3.4.1 ISA95 – Enterprise Control Systems

ISA-95 is the international standard for the integration of enterprise and control systems. ISA-95 consists of models and terminology that can be used to determine which information has to be exchanged between systems for sales, finance and logistics and systems for production, maintenance and quality (ISA, 2010). Thus, the benefit of integration of enterprise and control systems is to improve communication between various departments within an organisation.

With advances in technology, exchange of this information is being automated making important information available at the right place and time with enterprise having access to real time information such as information about raw materials enabling optimum usage of storage capacity. Thus there are a lot of advantages with an automated interface between the office and the shop floor (ISA, 2010).

The international standard ISA-95 has been developed to address the problems encountered during the development of automated interfaces between enterprise and control systems. This standard can be used in manufacturing environments all over the world and can also be applied in all sort of industries and processes whether batch or continuous.

3.4.2 Manufacturing Execution System (MES)

Manufacturing automation has faced a significant change over past 20 years with the emerging internet society addressing new enterprise control and management integration for agile business to manufacturing purposes. According to Morel et al. (2003) automation engineering would soon have to adopt the system engineering approach in order to deal with the increasing complexity of integrating intelligence/information manufacturing automation

with networked manufacturing enterprise. MES and ERP are couple of enterprise systems amongst host other systems as seen in Figure 3-9 that aim to facilitate manufacturing chain with networked enterprise in order to manage goods and services as desired by the internet society.

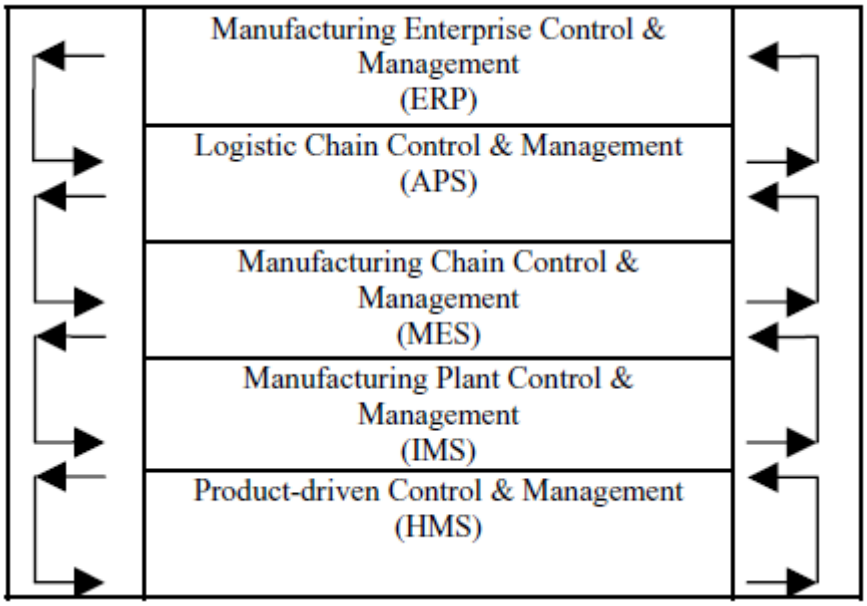

Figure 3-9: Manufacturing Enterprise Control and Mangement Systems (Morel et al., 2003)

According to Morel et al. (2003) a form of technical intelligence that has been embedded in the manufacturing systems and the products themselves that makes it possible to address the concept of agile business to manufacturing (B2M).

The primary goal of a number of industries using MES is for managing factory floor activities such as resource allocation, dispatching production units, quality management, operation planning, detailed scheduling, labour management, product tracking and keeping records for product genealogy. This information could then be used to optimise the production activities by:

- Improving communication inside a production facility
- Improving the communication capability between production and other activities in manufacturing enterprise such as product design, process planning, resource planning, supply chain management, service and sales and equipment control
- Monitor production to control important process parameters
- To better manage the production related data

MES is defined by Feng (2000) as "A system that consists of a set of integrated software and hardware components that provide functions for managing production activities from job order launch to finished products. Using current and accurate data, MES initiates, guides, responds to, and reports on production activities as they occur. MES provides production activity information to other engineering and business activities in the enterprise and its supply chain via bidirectional communications".

3.4.3 Enterprise Resource Planning (ERP)

Manufacturing systems have evolved from inventory control in the 1960's to material requirements planning (MRP) in the 1970's where a computer was used to calculate gross material requirements: from MRP II in the 1980's which incorporated the financial accounting system and financial management system along with manufacturing and materials management systems. In the early 1990's where MRP II was further expanded to include product design, information warehousing, materials planning, capacity, finance, and project management and given the term Enterprise Resource Planning (ERP) (Umble et al., 2003). Globalisation and revolution in information technology has pressurised companies to lower costs, improve the quality of products and provide realistic and reliable delivery dates through effective and efficient coordination of production and distribution activities. The ERP systems assure the means of integrating business functions under one database, one application and a unified and integrated interface across the business process. There have been a number of publicized failures with the implementation of ERP as this system does have moderate chance of hurting the business because of potential implementation problem (Umble et al., 2003). Out of the various number of problems that have been mentioned by Umble et al., 2003 poor management, change in business objectives during the project and lack of management support have been cited as the top three reasons associated with the implementation issues of the software. Within the ERP industry SAP and PeopleSoft are the most popular standardised solutions adopted by industry. However one of drawbacks with adopting these software solutions is that they impose their own logic on a company's strategy. As a result 50-75% of US firms have experienced failure to implement advanced manufacturing technology. There have been other occasions where for example Dell Computer Corporation initially started SAP implementation but later withdrew their interest in standardised software and designed a bespoke solution more suited to their organisation.

It is therefore extremely important for businesses to examine variety of critical factors for ERP implementation to be a success.

Important factors for ERP implementation would include:

1. Clear understanding of strategic goals
2. Commitment by top managements
3. Excellent Project Management
4. Organisational change management
5. A great implementation team
6. Data Accuracy
7. Extensive education and training
8. Focussed performance measures

With ERP systems imposing their own logic on company's strategy it is very important to select the most appropriate ERP system. The four most commonly used ERP systems are SAP, Oracle, PeopleSoft and Baan with each of these systems having 60-70% feature overlap making it difficult to accurately differentiate between the systems (Gupta and Kohli, 2006). SAP mainly dominates the ERP market with more sales than its three closest competitors.

The customer in this case uses SAP to manage not only their logistics and supply chain but also maintain quality control of their products. ABB's Enterprise Connectivity Software (ECS) that is an MES solution serves as the link between SAP and ADI.

3.5 Chapter Summary

The successful implementation of PAT solution will require a combination of all the tools discussed in this section. The ADI software aims to provide a complete package for PAT solution to meet the customers' process improvement and automation needs. With the ADI platform being both open and secure it enables tight integration with ABB's 800xA control system or any other third party control system as well. ADI would give the user the flexibility to use various analysers in conjunction with third party chemometric packages to build and develop predictive models for online, inline and at-line measurements. The model data exporter tool would further make it easy to access the data thus significantly reducing time consuming activities such as analyser integration and data synchronisation. Immediate benefits of QbD approach can be attained through appropriate quality management and implementation of risk based approaches. Validation being one of the key concepts for ensuring the success of QbD and PAT dedicated workflows can be configured to control operator interaction for at-line measurement taken for verification purposes.

The customer in this project also used SAP as their ERP system along with ABB's ECS as an MES solution. The MES solution bridges the vertical integration gap between the business and manufacturing systems by providing intelligent data access and viewing for different level within an organisation. Thus integrating data is of importance when analytical model is also being used to control a process using at-line measurements. The ADI system provided the customer with a single integrated data platform that simplified the integration of quality management systems, analyser integrations and process control for real time release of products. The next chapter discusses the working functionality of various components of the ADI solution.

4 Chapter 4 – Analyser Device Integration – ABB's PAT solution

4.1 Chapter Overview

This chapter looks at the various components of the ADI software that have been developed as per the customer requirements. This project required the combined resources of a project delivery team and a product development team as ADI product needed to be extended to deliver all of the customer's requirements.

With a PAT solution already pre-existing with the customer, Section 4.2 discusses the motivation for developing the ADI solution.

Section 4.3 briefly describes the PAT solution and explains the functionality of the various components of the solution.

Section 4.4 explains in detail various tests that were devised around certain components of the product in order to successfully commission the project.

Section 4.5 summarises the chapter.

4.2 Introduction

This project looks into the application of ADI software which is a scalable product meeting many of the requirements of a fully integrated PAT solution. ADI product has a flexible open architecture that is designed to integrate analytical devices and predictive modelling (Chemometric) software with process control and other higher level systems. One of the challenges to date as mentioned in Chapter 3 has been the wide information technology gap and lack of standards. Analysers and chemometric packages do not share a common user interface and data format and they do not offer the connectivity required to efficiently exchange data with plant and business system such as SAP. The first steps in overcoming this challenge were met by the introduction of the OPC Foundation standard for analytical device integration (OPC ADI). This standard introduced a common architecture through which analytical devices may exchange data with 'data users'. The standard also made allowances for integration with model development packages such as Unscrambler and Simca.

One of the unique features of ADI is its capability to integrate data from different analytical devices with quality and higher level servers. ADI bridges the gap between the businesses (ERP) and manufacturing systems (DCS) as it records, processes, manages, monitors, reports and stores the at-line spectral and predicted data along with the associated meta data.

The thesis mainly discusses the novel application of model integration within ADI, method deployment while enabling method compatibility by assessing the background validity issues and improving the model data exporter tool by implementing advanced selection methods to access data to optimise a model.

4.2.1 ADI and the Customer

The customer has been using the principles of PAT for more than 15 years. They currently have multiple plants worldwide using their proprietary PAT system. This system was ahead of its time but difficulties with bespoke design, continued support and the desire to add additional sites encouraged them to look for a system which was off the shelf and globally supportable. The overall objective was to enable seamless roll out of PAT worldwide by capturing and sharing best practices, thus allowing them to leverage the economic benefits of real time process analytics. The customer decided to work with ABB to extend ABB's existing PAT solution and hence enabling the customer to have an enhanced globally supportable platform.

4.2.2 Customers proprietary at-line system

The previous system managed the analytical and modelling components associated with at-line material tests. It was a centralised system with a dedicated network connection for PAT data to each of the sites involved. The majority of PAT test undertaken by the customer are quality driven and performed using at-line Nicolet FTIR analysers. There are some seventy plus tests that can be run as a part of the production process. Quality checks to be applied to a particular production batch are managed by a central SAP system. This system is responsible for the management of quality tests and their data and generate unique bar coded labels that identify individual samples. Each at-line analyser station has a barcode reader which uses the sample number from the barcode to execute the correct test method. The test method contains the necessary predictive models or equations necessary to perform the test. The spectral data from the analyser is passed to the 'PAT system' server where it is analysed using chemometric models. The results of the analysis are passed from the PAT system to the SAP server where they form part of a material inspection characteristic within the SAP quality module. The original spectral data are stored with the system server. The measurement is manual and involves sample preparation which is performed by an operator following a workflow on a dedicated screen locally beside the analyser.

While this system worked well it was necessary for the customer to standardise on a particular analyser type (Nicolet) and maintain a bespoke software driver. In addition, the lack of

integration with SAP meant that separate screens and spreadsheets were required to link the at-line lab with quality system. The collection of model development followed a specific fixed process and was not flexible.

4.3 ABB and Solution

The ABB solution facilitated to take a more open approach using the new OPC ADI standard providing an open platform to use different analytical platforms with tight integration to the quality and model development systems. As the customer has a global team responsible for model development and distribution it was important for them to maintain the ability to work centrally. ABB therefore implemented an architecture comprising of centrally based corporate level servers with a local site based client for each at-line analyser measurement location. The central servers manage model development, deployment and interaction with the SAP business and quality system whilst the local server and clients manage operator measurement workflows, analyser integration and data collection tasks. In keeping with the customer requirements only one central server communicated with the customers' business and quality systems (SAP). The sample number is scanned at the analyser station. ABB's Enterprise Connectivity Server (ECS) uses the sample number to interrogate SAP for the material number and test parameters required for that sample. The material number determines which method is selected to manage the test. The ADI method manages all the analyser settings, predictive and diagnostic models necessary to take measurement. Once a measurement is taken it is sent to a third party prediction engine (in this case CAMO) which provides the predicted value and associated statistics. These values can be visualised on the operator interface and also sent to central storage along with the sample and background spectral data. All of this data can be retrieved at a later date using the Model Data Exporter (MDE) tool. This report will look into detail about the various components of the system such as the ADI server, ECS server, Data Exporter, Information Manager (Historical Storage), Prediction server and Model development using hierarchical models that make up the system. Figure 4-1 gives the overview of the entire system and links between the various components of the software.

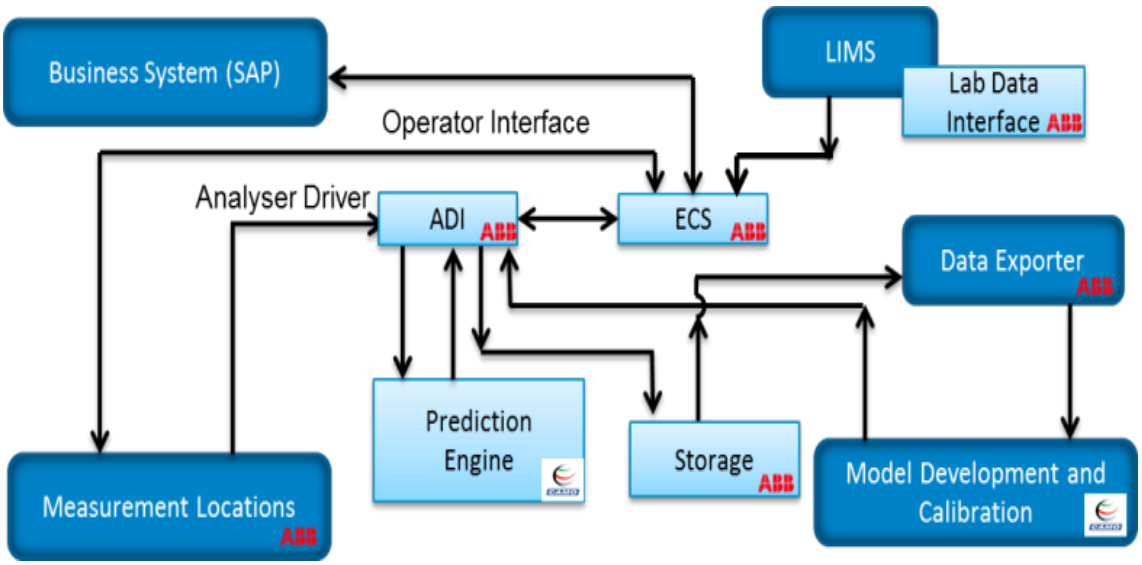

Figure 4-1: ADI System Overview

4.3.1 Analyser Device Integration (ADI) Component

The ADI Connectivity server is an 'add on' to ABB's 800xA control technologies platform. When used for online measurements it can be part of ABB's 800xA process control system or integrated through standard OPC interfaces to third party control systems. This project was concerned with at-line measurements only so there was no need to integrate it to the control systems. The ADI server manages the configuration and runtime for components associated with configuring and executing a method. When instructed by ECS it will co-ordinate the calling of the correct prediction model and manage the collection of spectral and predicted data. Model alarms and event handling is also managed by this component.

In order to take a successful measurement a method has to be configured within the ADI server. A method comprises of sample and background settings of the analyser, prediction models both for the background and sample measurements and other inputs such a constants and variables that are essential for the measurements. In order to manage at-line methods it is important to have control over the parts which can be edited which in this case is the method and the model. The Method is version controlled within the ADI server and the Model is version controlled by Subversion in repositories located in disk storage on the model development system.

4.3.1.1 ADI Method

An ADI method brings together all the parameters including models and analyser settings required to make a measurement. Configurations such as analyser settings, background age, sample models, background models, variables (batchID, materialID, etc) and constants (eg: co-efficient of variance as these are calculated previously and added to the method) are

attached here. ADI uniquely allows for parameterised setting of the analyser to be included in the method keeping fixed parameters such as analyser name and IP address separate. This allows all the parameters that affect the measurement to be managed and version controlled in one place.

Method Execution

Once a method has been created as part of the version control process, it then has to be approved before execution. A method can be executed from method runtime aspect faceplate in the ADI as well from operator interface i.e. from the ECS.

4.3.1.2 Analyser settings and third party systems compatibility

In a traditional set up each individual analyser is managed by its own dedicated client software which makes no distinction between parameters used for measurement (e.g. aperture size and number of scans) and parameters used to define the analyser (e.g. analyser name and network address). With ADI, parameters for measurement belong to the method object and parameters that define the analyser belong to the analyser object. This was a decision taken early on to allow the system to be model centric providing flexibility in how models can be used and transported between sites. As long as an analyser is compliant with OPC UA standards ADI allows for vendor independent integration of all the devices with centralised management and execution of models and methods using the aspect object technology. The user has complete control of the analyser settings and actions through the common user interface while having full transparency of all raw, processed and diagnostic information. ADI also allows for concurrent integration of analyser from multiple vendors and it is scalable from 1 to 25 devices in single system.

In the past, due to the constraints imposed by technology standards, the customer had limited themselves to one analyser manufacturer and one analyser type (FTIR). Implementing the new OPC ADI standard using an ABB driver enabled the customer to consider the use of other analyser platforms. In future they now have the flexibility to consider different analyser vendors provided that the vendors have adopted the ADI standard.

4.3.1.3 ECS Server

The Enterprise Connectivity Software bridges the vertical integration gap between business and manufacturing systems and delivering significant new opportunities to increase productivity by providing tighter integration with the customers business and quality systems. The use of this system allows, for example, the operator to view both quality and

measurement parameters integrated on to one screen. Previously multiple screens connected to different systems would have been required.

The ECS solution belonging to the ABB cpmPlus family comprises of ISA-95 model based platform which performs:

- as a storage model for keeping the states of all connected systems,
- as a definition template for specification of process workflows and data mappings and
- as a definition paradigm used for structuring and maintenance of communication channels towards external systems.

The integrated model-based approach uses object oriented software to define and maintain the manufacturing processes linking it with associated data stored in external systems.

In runtime, the user interface is provided by an ECS Client. The ECS Client is built on top of the ECS Client Framework, which provides an event based platform for development of the user interfaces for different environments (e.g. mobile devices, windows forms, web-based).

The cpmPlus ECS contains several modules responsible for different aspects of an MES solution. Each module is designed to plug into a central core allowing for a flexible configuration and cost. The following modules are utilized in this project.

- Execution module providing list of tasks, workflow management and traceability for execution and required information for operators
- Administration tool giving the ability to reconfigure the system (quality data, users, manufacturing instructions on particular work centres, workflows, etc.)
- Barcodes module providing integration with barcode scanners and printers
- Reporting for delivering required information from all system's modules to plant management

4.3.1.3.1 Operator Workflow Description

A barcode scanner is connected to measurement station computer via a USB port. Users are authenticated by the system using login and domain authentication. Depending on which group in the Active Directory (AD) the user belongs to, defines the access rights assigned. (e.g. the right to perform measurements on a particular analyser and privileges to execute

particular methods.) . This allowed the customer to use existing IT infrastructure and existing user accounts.

In order to perform at-line measurements it is necessary to know the relationship between the SAP material number of the scanned sample and the method which is to be executed. The barcode determines which type of type test has to be carried out. This can be Q (quality), 'P' production or 'X' (experiment – used for model checking and development).

Q-Type: A formal quality measurement where the results are reported to SAP. The barcode provides the sample number which enables the system to retrieve the material number from SAP. The material number determines which ADI method will be executed.

P-Type: A production measurement where results are reported locally. The barcode is a 'dummy' number which tells the system to skip the material number look up from SAP allowing the operator to select it from a predefined list instead. The material number determines which ADI method will be executed.

X-Type: An experimental measurement used during model development. This is a specialist menu driven measurement available to a limited number of users. The user will enter in the sample identification (e.g. sample number) and select the ADI method to be executed from a list.

Once the correct method is determined the operator will be taken through a workflow (step by step instructions) until two measurements have been made that pass a statistical t-test which is used to ensure that both samples have been prepared correctly. The average of the two measurements is then used as the result.

The method defines which data is to be collected and saved in the Information Manager and if a Q – type measurement reported to SAP. Figure 4-2 explains the various workflows used as a part of the ADI solution.

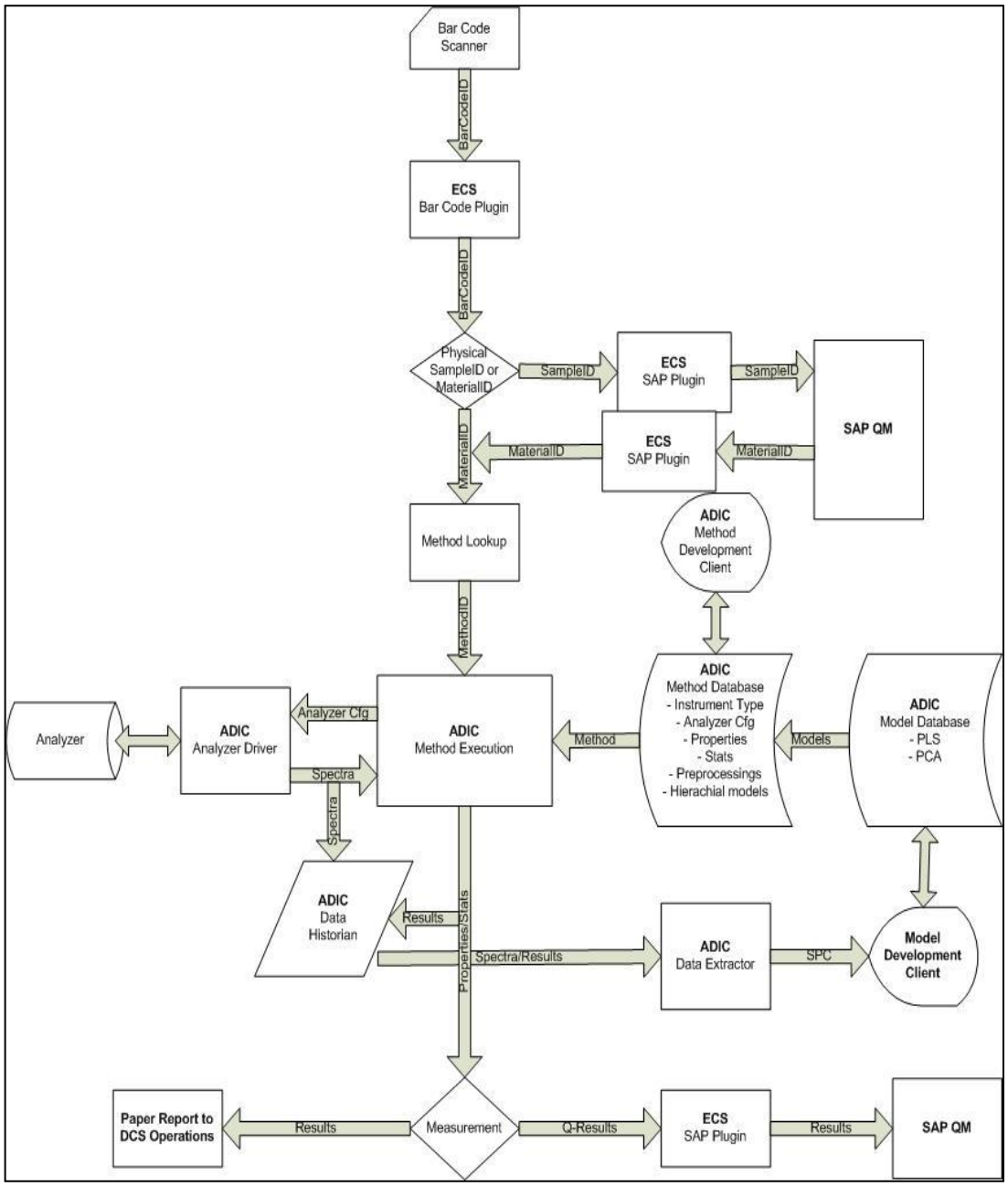

Figure 4-2: Workflow of at-line measurment

4.3.1.4 Storage - ADI Information Manager (Historian) Description

In order to continuously improve process and material understanding it is necessary to store as much information, as possible, about the measurement. Data must be stored in such a way that it can be extracted and presented to different users with different requirements at a future date. History services are therefore an integral part of the solution. For this project they are responsible for the storage of spectral, numerical, lab, process, alarm & events and method meta data.

Spectral data is a large data-array which in the case of NIR analyser represents the measurement and background interferogram. Each measurement spectra is stored in a ‘Profile

Logs' (array storage) located on the ABB 800xA Information Manager (IM) server. Spectral data is stored on event i.e. as a measurement occurs. Numerical data (vector data) is a single value time stamped data such as predicted value or associated statistic. Numerical data is stored on event (change) within a 'Property Log'.

Lab data is the data result from the empirical lab test used to verify that the analyser is predicting correctly. This data is stored in a special type of 'Property Log' called as 'Lab Log'. The main difference between the two types is that the 'Lab Log' requires a manual entry for the time stamp where as the 'Property Log' picks up the time the event occurred automatically from the system. The lab data is time stamped (with the time the sample was taken) and is retrospectively stored alongside the at-line measurement data. A lab log may be populated from external third party products such as MS Excel or LIMS. For the purposes of this project lab logs formed a part of the lab system interface.

Alarm and Event data is captured and stored within a message log located on the IM. In general terms 'alarms' are events that are used to indicate a problem that needs to be brought to the operators attention such as in instrument failure or model deviation whereas 'events' such as audit trail events are also recorded here. Methods 'meta' data is the data that describes the circumstances under the method ran. It contains for example the location, type and settings of the analyser as well as user and model information. It is directly related to the data storage described above. This data is stored into a Production Data Log (PDL). The PDL, in addition to meta data, contains links to spectral, numerical, lab and alarm/event data associated with a measurement.

The archive function allows for offline storage of all data held within the system. The customer required that measurement data be archived periodically to network drives for secure storage. This was achieved by archiving PDL logs along with their linked data. As seen in Figure 4-3 measurement data (eg: spectra) is stored as Profile Logs while the predicted properties and process data are stored under Numeric Logs. The alarm and event data is recorded as Message logs. All of the data stored under PDL can be accessed offline at any time in the future from the archived PDL.

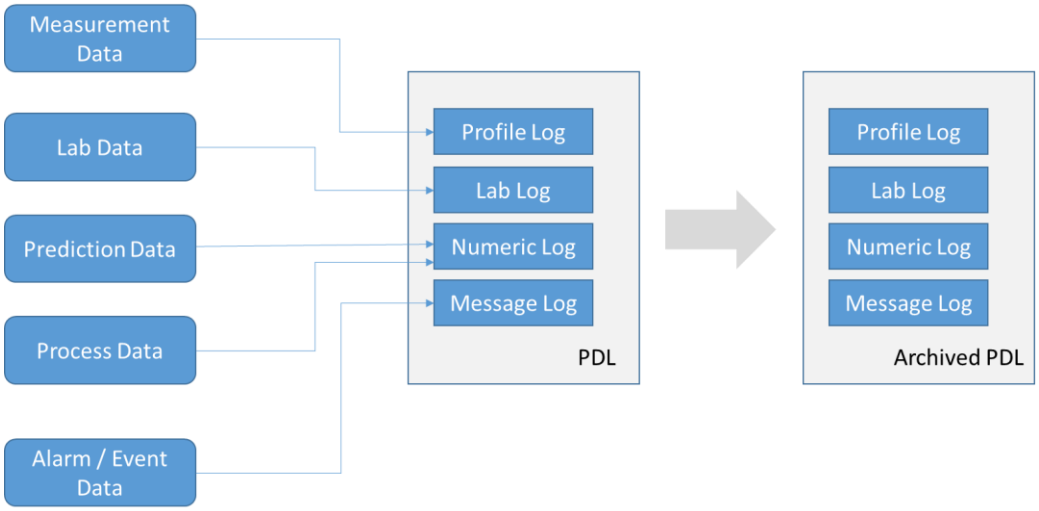

Figure 4-3: Data storage structure in ADI

4.3.1.5 Model Prediction Engine

For this project the prediction server was CAMO's Insight prediction server which is the online prediction engine associated with CAMO's Unscrambler model development software. The server is tightly integrated with ABB's 800xA ADI technology.

Models are developed within CAMO's Unscrambler environment. The run time model is imported into ABB ADI server where they are then downloaded to CAMO's Insight prediction engine. Measurement spectra is passed to the prediction engine which returns the predicted value, associated stats and any model alarms which may have occurred. At any one time there are a number of different models waiting to be called. The diagram in Figure 4-4 describes the relationship of the ADI Insight server with the complete 800xA configuration and Figure 4-5 defines how the ADI Insight server works.

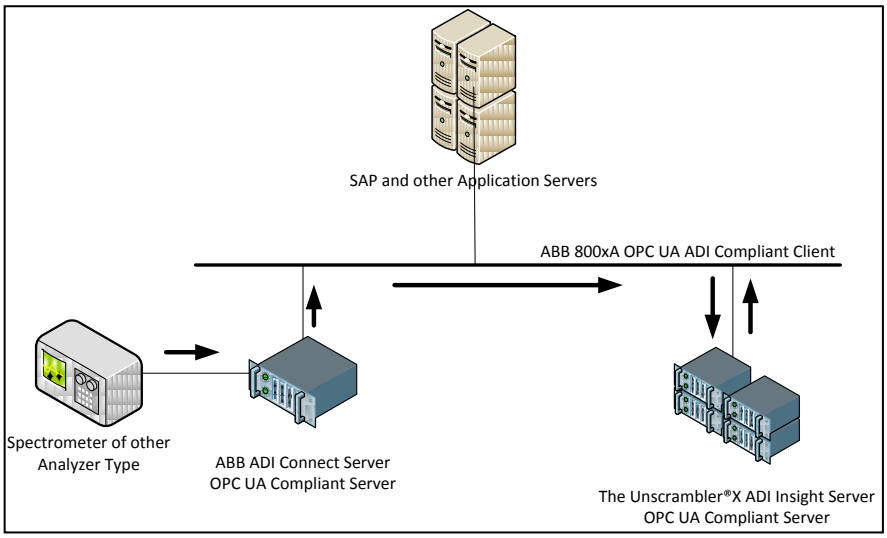

Figure 4-4: Information flow from ADI Connect Server to CAMO Software's ADI Insight Server

System	ABB System 800xA		The Unscrambler® X ADI Insight
Type	OPC UA ADI Compliant Client		OPC UA ADI Server
Information Flow			
Process 1	Send request to ADI Insight Server for Analysis and send Chemometric Model to load.	→	ADI Insight checks validity of Model sent to server and also check if it has already been loaded.
	List of outputs becomes visible to the user and is presented in a tree structure	←	Once validated, ADI Insight sends all relevant Model parameters back to 800xA client
Process 2	800xA OPC UA Client sends request for analysis along with new sample data to apply the chemometric model to.	→	ADI Insight server applies the sent Model to the data
	Results can be viewed in 800xA visualization modules or any other OPC UA ADI compliant client	←	ADI Insight server returns all of the prediction statistics to 800xA OPC Client

Figure 4-5: Workflow within The Unscrambler X ADI Insight Server

The data returned by The Unscrambler ADI Insight server during process 2 is:

1. The prediction of the property
2. Statistics
3. Hi / Lo alarms (property and/or stats)
4. Hi / Lo warnings (property and/or stats)
5. General Alarm (based on logical condition)
6. Normal (No Alarm)

4.3.1.6 Model Data Exporter

Once the measurements have been taken and properties have been predicted and sent to historical storage they can be accessed anytime in the future using the ABB's Model Data Exporter (MDE). Lab data along with spectral data and predicted values can be exported

using the MDE which can then be used by Chemometricians to develop new calibration models.

The MDE for ADI has adopted a model centric data export unlike its legacy xPAT which performed a batch centric data export. The filter applied on the MDE depend primarily on the model that has been initially chosen following which other filters can be applied such as:

- Model and Method Version
- Methods
- Method Instance ID
- Variables using expressions
- Measurement Locations
- Analysis Time
- Model Alarms
- Lab Value and Tolerances
- Property Values

Default as well as customised template can be used to select columns and ordering and constructors for row labels and array headers. Also the data can either be exported to standard files such as MS excel or directly to the supported chemometric package. One of other important features of the MDE is that data exported from different time zones is time aligned.

4.4 Development of test scenarios for Method and Model Deployment, Model Import and Model Data Exporter (MDE) within ADI method

As a part of this project various unique test scenarios were to be developed in order to ensure that the model development and model data exporter side of the project was successfully commissioned. This section of the report describes the tests developed for Factory Acceptance Test (FAT) of the product. Although there were various other tests conducted during the FAT this section will look in detail at the tests developed by the author for successful working of the ADI method and the MDE component. The report also details the critical analysis of various problems encountered during the testing stages of the product and discusses the novel methods that were developed in order to overcome the challenges to develop an integrated novel solution.

This section of the thesis has been divided into four parts: 1) Model development 2) Method configuration 3) ADI Alarms and 4) Data Export. For each of these sections the criteria for each test, the preparation required, the tests carried out and the outcome/result has been discussed. The detailed steps carried out in each test can be found in ABB's internal AT-Line FAT specification document.

4.4.1 Model Development and Model Import

The customer provided a list of PLS models (Table 4-1) to be developed within The Unscrambler version 10.2 environment. The original names of the properties have not been disclosed due to confidentiality reasons. The calibration models were developed as per customer requirements and prediction models were tested using simulated csv files containing spectral data. The models were developed to include model alarms that could be generated and visualised on an ADI method faceplate. The test scenario detailing the working of this application and outcome of the tests has been further discussed in this section.

Table 4-1: List of FAT models

Model Name	Model Type
Property 1	PLS model
Property 2	PLS model
Property 3	PLS model
Property 4	PLS model
Property 5	PLS model
Property 6	PLS model
Property 7 v17	Hierarchical PLS model
Property7 v17_low	Hierarchical PLS model
Property7 v17_high	Hierarchical PLS Model

Figure 4-6 is a screenshot of the heirarchical Property 7 model used in the FAT testing of the ADI system. As seen in the Figure 4.6 the Global Property 7 model is applied for the 1st level. If a model warning (in orange) or model alarm (in red) is generated at this level then values are sent to the ADI server which can be visualised on the 800xA faceplate. However if predicted values are normal (green) at the 1st level then depending on the range of value either a low Property 7 or high Property 7 model is applied and the predicted values are displayed at the 2nd level. The values generated at this level are then sent to the ADI method faceplate for the operators to view it.

The screenshot of three other models mentioned in Table 4-1 has been attached in the Appendix A.

Results		Level-1			Level-2					
		Y Predicted	Sample Leverage	X Sample Validation Residuals	Y Predicted	Sample Leverage	X Sample Validation Residuals	Y Predicted	Sample Leverage	X Sample Validation Residuals
		Property 7 Global			Property 7 v17 Low			Property 7 v17 High		
		1	2	3	4	5	6	7	8	9
POI17220_____99762290	1	18.9597	0.0205	0.6248	5.2302	0.1090	0.1867			
POI17220_____99762290	2	7.5538	0.0344	0.0305						
_____10100001902	3	23.7191	0.0237	0.0030	23.7268	0.0159	0.0044			
POI17221_____10100057217	4	13.1205	0.0172	0.0197						
LDN00026_____10100052848	5	49.3379	0.0082	0.0051						
PLZ76475_____10100066714	6	32.5297	0.0062	0.0151				40.1528	0.0674	0.0355
PLZ76475_____10100066714	7	32.9437	0.0062	0.0145				40.5533	0.0636	0.0353
HONO01003_____10100102877	8	53.7147	0.0089	0.0037						
HONO01003_____10100102877	9	53.9525	0.0092	0.0048						
ADKA0764A_CONC_40565	10	41.7669	0.0078	0.0048						
ADKA0765_10%___40571	11	48.6718	0.0078	0.0078						
ADKA0766_5%____40574	12	42.3973	0.0086	0.0040						
ADKA0768_CONC__40581	13	42.6346	0.0085	0.0051						
ADKA0771_CONC__40585	14	41.9922	0.0083	0.0054						
ADKA0771_5%____40586	15	45.1852	0.0068	0.0050						
ADKR0768_15%___40596	16	54.8506	0.0095	0.0124						
OMN09009_5%____41048	17	9.3021	0.0441	0.0112						
OMN09009_15%___41050	18	18.4582	0.0355	0.0076	19.5947	0.0157	0.0607			
PLZ76479_____10100183879	19	35.5591	0.0063	0.0117				42.0336	0.0511	0.0316
ADKR0743_5%____45502	20	42.3598	0.0073	0.0024						
PLZ96481_____10100285375	21	33.1968	0.0092	0.0109				39.3561	0.0752	0.0255
PLZ96481_____10100285375	22	33.2661	0.0094	0.0112				39.0948	0.0778	0.0240
PBZ93105_____10100346132	23	62.1790	0.0166	0.0044						
PBZ93105_____10100346132	24	62.5512	0.0174	0.0042						
ADKD0800_____10100357554	25	19.6407	0.0159	0.0016	20.6208	0.0146	0.0090			
PBZ63124_____10100448467	26	65.2956	0.0236	0.0063						
EAKA0068_____10100462256	27	4.1025	0.0304	0.0319						
CNZ15193_____10100470021	28	22.8694	0.0236	0.0018	23.5850	0.0157	0.0036			
LCKD0112_____10100541845	29	14.6365	0.0164	0.0047						
LCKD0112_____10100541845	30	15.0834	0.0161	0.0042	14.3010	0.0331	0.0059			

Figure 4-6: Screenshot of Hierarchical Property 7 model

In order to run and develop the model further it then has to be imported within the 800xA ADI environment. The FAT test for determining the successful import of a model is described in Section 4.4.1.1.

4.4.1.1 Sample Model Import FAT test

Purpose of the Test

To verify that available CAMO Unscrambler 10.2 prediction models can be imported to the ADI server. Successful import of models into the ADI system is one of first steps to be carried out

Test Method

CAMO Unscrambler 10.2 prediction models were imported into the system by navigating to the functional structure Root/ADIConnectObjects/ ADIConnectConfigurationObjects/ADIConnectModels/ADIModels/ TestCollection/ADIModelCollection and selecting the ADIConnectModelImporter aspect. Using the browse option the test models were then uploaded into ADI system. Once a model is successfully imported into the system it will appear as new object type in the ADI system.

Result

The models initially failed due to the problem described further in section 4.4.1.2 which was overcome and models were then successfully imported in the functional structure and could be verified under TestCollection/ADIModelCollection

4.4.1.2 Model import issue and solution

The problem: Initially while importing the models in the ADI environment the entire PLS model including the spectral data set that used to generate the model was imported into the system. However during optimization of the MDE it was identified that the large size of the model resulted in slowing down the system. In extreme cases it would also cause the import to crash the ADI system resulting in unsuccessful model import.

Solution: This problem was resolved in collaboration with CAMO. The author suggested to reduce the original size of the model so as to include only those components of the model that are required for the prediction of properties The solution therefore facilitated to only import the 'short' model' and to remove additional data that was not required for a run time prediction such as the raw spectra used for calibration and various other parameters used for model development. The result was a 'run time' or short HM model that enabled successful import of the models.

4.4.1.3 Validating Not a Number (NaN) within the test environment

It is possible under certain conditions for empty or bad data (defined as NaN by the customer) to be returned to the system. From earlier learning the customer required for the system to handle the NaN for two scenarios: 1) for wave numbers in the spectra for which the predicted values remained undefined and 2) for the system to handle unexpected missing values in the spectra. If either of these situations occurred the system was required to populate the data fields with NaN. In line with this requirement the ADI system was modified in order to deal with NaN. CAMO too made a number of changes to the way Unscrambler and its prediction engine performed when missing values or NaN values were present in the data set. CAMO reported the following changes to the Unscrambler and its prediction engines:

1. Modification of the Savitzky-Golay Derivative functionality to handle missing values in the calculation

2. Modification of the Predict functionality to process predicted values as non-predictions when a missing value is present at a significant regression co-efficient.

Following the changes models were developed and the simulated spectra were chosen such that it would demonstrate the working of the NaN functionality.

The PLS models developed for the testing are detailed below:

1. Global Property 7: Spans the range of 10 to 65% Property 7
2. High Property 7: Spans the range of 40 to 65% Property 7
3. Low Property 7: Spans the range of 10 to 35 % Property 7

As it can be seen from the above property ranges 35% to 45% is an undefined range. If the prediction property is within this undefined range then an alarm is generated by the prediction engine and the outcome of the result is reported to the ADI system.

The three major wave numbers selected for testing the missing values are listed in Table 4-2.

Table 4-2: Wave numbers to test NaN

Model	**Global**	**High**	**Low**
Wave number (cm-1)	836.996	1078.07	1020.21

The test scenarios listed in Table 4-3 were carried out and the pass/fail status of the tests was recorded. All the samples contained missing values at the start of the spectra and therefore acted as test cases for this system.

Table 4-3: Various test scenarios for HM Property 7 model

Sample ID	Test Definition	Expected Result	Acceptance (Yes/No)
7	Normal sample, Low Property 7	Sample will pass both Global and Low Property 7 Prediction with no warnings	Yes
7a	Fail sample at Global Property 7, (missing value at 836.996 cm-1)	Sample will fail at the Global Property 7 level 1 with no prediction possible.	Yes
7b	Fail sample at Low Property 7, (missing value at 1020.21 cm-1)	Sample will fail at the Low Property 7 level 2 with no prediction possible.	Yes
7c	Fail sample at Low Property 7, (missing value at 1018 cm-1)	Sample will fail at the Low Property 7 level 2 with no prediction possible. The missing value does not occur at a significant regression coefficient , but in a segment where the derivative converts it to a missing value at the significant coefficient	Yes
7d	Normal sample, Low Property 7 (missing value at 1029 cm-1)	Missing value close to a significant regression coefficient but in a segment that does not encapsulate the significant regression coefficient.	Yes
36	Fail sample, OOS low Property 7.	Sample will fail due to the predicted value being less that that expected for a Low Property 7 sample.	Yes
39	Fail sample, No Evaluation	This sample will fail since the predicted value lies within the No Evaluation region between 35 and 40% Property 7.	Yes
42	Fail sample, No	This sample will fail since the predicted value	Yes

	Evaluation	lies within the No Evaluation region between 35 and 40% Property 7.	
52	Fail sample at High Property 7 level 2.	This sample will pass the prediction value limits for Global Property 7 at level 1, but fail prediction for High Property 7 at level 2.	Yes
54	Pass sample, High Property 7	Sample will pass both Global and High Property 7 Prediction with no warnings	Yes

4.4.1.3.1 FAT to verify NaN is handled correctly within the system

This test has been devised to verify if the NaN was handled correctly within the Unscrambler environment.

Purpose of the Test

To verify that the NaN are handled correctly

Test Method

The test data is described in Table 4-4:

Table 4-4: Test Data for NaN Test

Test Configuration	Value
CSV Data File Name	TestSetMissingValues.csv
Method Name	ODProperty 7 Method
Material Code	111 111 111 3
Location Name	Location 1
Unscrambler Project File Name	NewTestProject with Missing

Following are the steps executed for the tests:

1. Log into the Aspect Server and open the ADIServer configuration application and confirm that the TestSetMissing.csv appears as connection string.

2. Open this CSV file and confirm that the NaN value is present in some row(s) of the file.
3. Log onto the Unscrambler Server and open Unscrambler. Load the <NewtestProjectwithMissing> model and create a new matrix. Populate the matrix with the csv file values. Select the created matrix and apply the HM prediction.

The screenshot in Figure 4-7 are the results generated from the test. The NaN appeared as an empty cell within the Unscrambler environment however within the 800 xA system it would appear as NaN.

The remaining part of the test that further confirmed the generation of model alarms on the faceplate by taking measurements in the ADI system can be found in the FAT internal documents.

Results		Level-1			Level-2					
		Y Predicted	Sample Leve	X Sample Va	Y Predicted	Sample Leve	X Sample Va	Y Predicted	Sample Leve	X Sample Va
		Property 7 Global			Property 7 v17 Low			Property 7 v17 High		
		1	2	3	4	5	6	7	8	9
7	1	19.2027	0.0162	0.0000	18.2339	0.0257	0.0000			
7a	2		0.0160	0.0000						
7b	3	19.2027	0.0162	0.0000		0.0159	0.0000			
7c	4	19.2027	0.0162	0.0000		0.0159	0.0000			
7d	5	19.2027	0.0162	0.0000	18.2339	0.0257	0.0000			
36	6	3.2294	0.0403	0.0000						
39	7	35.0427	0.0066	0.0000						
42	8	39.0373	0.0065	0.0000						
52	9	41.1602	0.0067	0.0000				39.6491	0.0149	0.0000
54	10	47.7048	0.0092	0.0000				48.9865	0.0103	0.0000
54a	11	47.7048	0.0092	0.0000					0.0103	0.0000
54b	12	47.7048	0.0092	0.0000					0.0103	0.0000
54c	13	47.7048	0.0092	0.0000				48.9865	0.0103	0.0000
84	14	66.2493	0.0295	0.0000						

Figure 4-7: Screenshot of predicted Property 7 model results to test missing values

4.4.1.4 ADI Background Tests

Purpose of the Test

To verify that the background acquisition functions is as specified in the SDS.

Test Method

The test data used to perform the FAT tests are described in Table 4-5:

Table 4-5: Test data for background tests

Method Runtime Configuration	Value
Material ID	3562100940
User ID	U_ID_123456
Method Name	Property 7
Batch ID	PB_ID_123456
Model Name	Background
Physical Stream/Analyser Name	AKX1001779
Location Name	Location2
Analyser Group	NicoletiS10
CSV file	Background.csv

A validated background is necessary to ensure that any background changes are accounted for in the sample measurement. For the analyser used in this project it was necessary to acquire a new background every 30 minutes into the process. The HM predict functionality was used on the background models to visualise the expected results which should appear within ADI. These results were used to visualise expected results for the test before they were used to acquire measurements. Figure 4-8 is an example of typical background model with the HMPredict functionality used to test its prediction capabilities. The alarms generated within model were visible to the operators on the 800xA faceplate.

Results		RMS 1		Peak Model 1		PCA 1			
		RMS	State	Value	State	Projected Ho	State	Projected Sa	State
		1	2	3	4	5	6	7	8
2	1	0.0033	Warning	285.9850	AlarmHigh	39.6643	Alarm	0.1900	Alarm
17	2	0.0057	Alarm	279.2423	WarningHigh	36.2053	Alarm	0.1738	Alarm
1	3	0.0014	Normal	241.9158	Normal	76.6830	Alarm	0.3630	Alarm
22	4	0.0032	Warning	207.6467	Normal	101.9460	Alarm	0.4810	Alarm
9	5	0.0012	Normal	187.0411	Normal	190.7850	Alarm	0.8962	Alarm

Figure 4-8: HM predict on Background diagnostic model

Results

For the purpose of this test the variables were populated in the Method Runtime Aspect (MRA) of the Location object. Following this the successful import of background models was confirmed in the prediction server. The model was configured to the 'Property 7' method and measurements were taken using the Main Faceplate. The background diagnostics was confirmed by looking at the extended faceplate to ensure the background measurements were acquired. The method was then stopped and the test was passed once the faceplate returned to its original idle state.

4.4.2 Method Configuration and Method Execution

A method for the ADI system is an Object type that allows a user to define the offline configuration of a recipe which will be executed by the ADI system. The configuration of a method would include analyser settings, prediction and background models as well as other inputs such as constants, variables and process values. The configurations are saved in method configuration aspect in 800xA system. The method configuration aspect supports versioning of the methods and the most appropriate configuration would be approved and utilized for at-line measurements.

4.4.2.1 ADI Method configuration Test

Purpose of the Test

To verify that a method can be configured using the ADI system as specified in the System Design Specification (SDS)

Test Method

The test data used for FAT testing can be seen in Table 4-6

Table 4-6: Test Data for Method configuration Test

Method Name	Property 7 Method
Analyzer group	FAT Analyzer
Measurement Location Group	Atline
Sample Model	FAT Model Test
Background Model	FAT Background Model Test

Results

The detailed steps for executing method configuration in FAT and results of the tests conducted can be found in at-Line FAT specification document. The ADI method configuration test was successful with the method being approved and ready for taking at-line measurements.

Figure 4-9 is a screenshot of a typical ‘ADIMethodConfig’ Aspect. Within this aspect the tabs for Analyser Settings, Inputs which would include the constants and process values and the Models tab which would consist of Sample model and Background model can be clearly seen. Once a method has been configured it then has to be Approved in order to be able use them to take at-line measurements. The Approved Method aspect can also be seen in the figure on the following page.

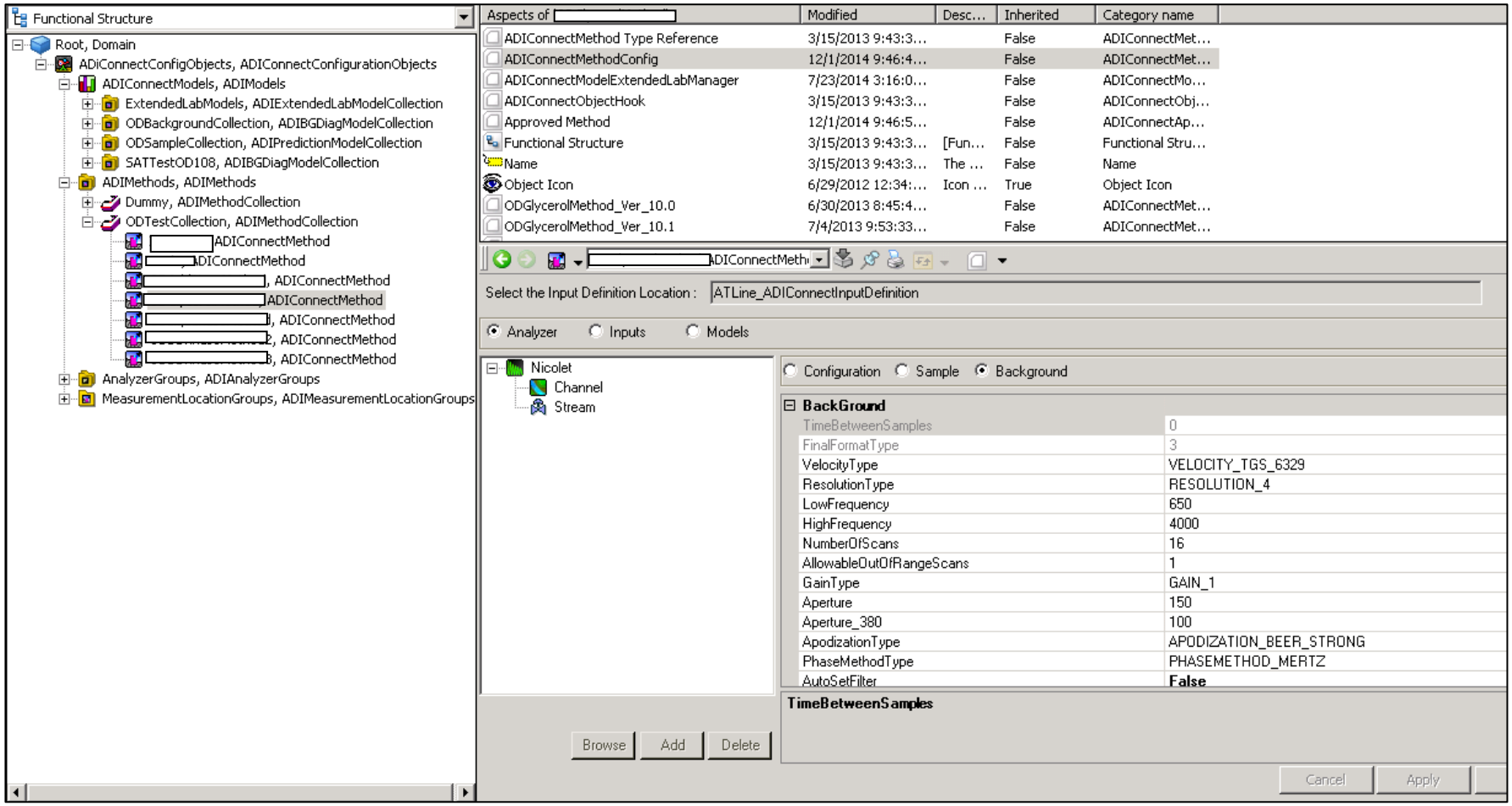

Figure 4-9: Screenshot of method configuration aspect

4.4.2.2 ADI Method Execution Test

Purpose of the Test

To verify that ADI method can be executed within the 800xA At-line ADI system as specified in the SDS.

Test Method

Table 4-7: ADI Method Execution Test Data

Method Runtime Configuration	Value
Material ID	3562100940
UserID	U_ID_123456
Method Name	ABBINMethodTest1
Batch ID	PB_ID_123456
Model Name	Property 7
Physical Stream/Analyser Name	Analyser1
Location Name	Location 2
Analyser Group	Nicolet 380

The Batch ID, User ID and Material ID were populated in the Method runtime aspect window which is opened by selecting the main faceplate aspect under the Location 2 aspect. After confirming the necessary configurations such as successful model import in the CAMO insight server, successful log configuration and if the correct version of method has been approved, the 'ABBINMethodTest1' method was executed using Main Faceplate for Location 2 object.

Results

The test was passed after taking successful at-line measurements from the faceplate as per test procedure and was deemed completed when faceplate finally returned to its original 'Idle' state.

4.4.3 Method Compatibility Issue

The customer has been using two different makes of Nicolet analyser: Nicolet iS10 and Nicolet 380 for at-line measurements and wanted the flexibility to use any of the two analysers' with any method no matter which measurement location or analyser group the method belonged to. This situation turned out to be critical since one particular analyser group (either iS10 or 380) needed to be configured in order to approve a method.

This problem was addressed by adding an additional setting for the analysers in the method and introduction of two separate analyser drivers for iS10 and 380. The FAT test carried out to ensure successful implementation of the change can be seen further in Section 4.4.3.1.

Also with one method now being used across more than one analyser, ADI also had to tackle background validity issue. The customer had specific requirements for the background expiry settings such as if the background on one of analysers expired this should not affect the background of the other analyser: with both analysers being used across the same method. Section 4.4.3.1 also describes the tests that were designed in order to verify the background validity across methods and analysers.

4.4.3.1 Method Compatibility Test with iS10 and 380 Analysers

Purpose of the Test

1. To verify that the same method can be used for both iS10 and 380 analyser instance.
2. Review measurement to confirm that correct instrument settings are used on each instrument.
3. Also confirm that the background is not shared with another measurement location.

Test Method

Table 4-8: ADI Method Compatibility Test Data

Test Configuration	Value
Method Name	ODProperty 6 Method
Material Code	5555555551
First Location	Location 1
Second Location	Location 2

Prediction Property Name	Property 3 Property 7 Property 1 Property 2
CSV file name	NewTestSet.csv

In order to be able to verify the method compatibility across analysers at different locations tests were performed to verify the analyser settings by monitoring the factory settings for each analyser. For Location 1 the factory setting were confirmed at C:\ProgramFiles(x86)\ABBIndustrialIT\ADIConnect\AnalyzerDriver\NicoletAdapter\config\. In the xml file the factory settings on the analyzer were confirmed by verifying that the Initial Value for parameter "Model" is set to "2" for analyzer 380. This would confirm that analyzer 380 has been assigned to Location 1. The same parameter for analyzer iS10 is set to a value of "4" to assign it to Location 2.

Results

The measurements were initiated from the ECS server. Prior to the initiation of measurement the aperture setting on the each of the analyzer was recorded by navigating to [FunctionalStructure]Root/ADiConnectConfigObjects/ADIMethods/TestMethodCollection/< *Method Name*>. The aperture setting can be found out under the approved method tab and expanding the analyser radio button. The recorded aperture setting is seen Table 4-9.

Table 4-9: Aperture size for analysers' iS10 and 380

Parameter	Value
Aperture_iS10	150
Aperture_380	100

In the ECS client the material code is assigned to the method to be executed as previously defined in Table 4-8. On selecting Location 1 to take measurement the sample barcode is inserted manually. A new background is taken for Location 1 with analyser 380 followed by 2

replicates of sample measurements. On completing this measurement cycle Location 2 is selected and the barcode inserted manually. On the next screen the background for this location was still invalid thus confirming that although same method is used for both the locations the background is not shared between the 2 locations.

4.4.4 To demonstrate that Separate Background for Methods with the same Analyser settings are not required

Following the initial FAT test it was observed that a new background was required for every method although the analysers used for both the methods were same. This phenomenon was raised as an observation deviation (OD). This issue was rectified for the next FAT and test specific to approve the functionality of the required background has been discussed further.

Purpose of the Test

To verify that methods sharing the same Analyser settings can have the same background on a particular instrument.

Test Method

The test data is described in Table 4-10:

Table 4-10: Test method to verify background validity

Test Configuration	Value
Method1 Name	ODProperty 4 Method
Method2 Name	ODProperty 7 Method
Material Code 1	1111111112
Material Code 2	1111111113
Physical Stream/Analyzer Name	AGL0500230
Background Age Method1	30 min
Background Age Method2	30 min
Prediction Property Method1	Property 5 Property 2
Prediction Property Method2	Property 7

Test Configuration	Value
Location Name	Location 1

The tests were carried out in a number of steps to ensure both the methods have the same analyser configurations and same background age. Initial steps designed to carry satisfactory working of this functionality have been discussed here.

1. Log into the Aspect Server) as User1, navigate to [Functional Structure]Root/ADiConnectConfigObjects/ADIMethods/TestMethodCollection/ *<Method1 Name>*, click on the Approved Method aspect and under the analyzer radio button expand analyzer setting icon and click on stream.
2. Confirm that the Background Age is set to *<Background Age Method1>*.
3. Click on the sample radio button and record the configuration of the *<Method1 Name>* in Table 4-11:

Table 4-11: Test method to verify background validity: Method 1 configuration

Parameter	Value
Velocity Type	VELOCITY_TGS_6329
Resolution Type	RESOLUTION_4
Low Frequency	650
High Frequency	4000
Numbers of Scans	4
Allowable Out of Range Scans	1
Gain Type	GAIN_1
Aperture	150
Aperture 380	100
Apodization Type	APODIZATION_BEER_STRONG

Phase Method Type	PHASEMETHOD_MERTZ
Auto Set Filter	FALSE
Low Pass Filter Type	LOWPASS_11000
High Pass Filter Type	HIGHPASS_20
Interferogram Maximum	9.8
Interferogram Minimum	2
Peak Minimum	10
Absorbance Peak Minimum	-0.50
Absorbance Peak Maximum	3.00

4. Navigate to [Functional Structure]Root/ADiConnectConfigObjects/ADIMethods/TestMethodCollection/ *<Method2 Name>*, click on the ApprovedMethod aspect, under the analyzer radio button expand analyzer setting icon and click on stream.

5. Confirm that the Background Age is set to *<Background Age Method2>*.

6. Click on the sample radio button and record the configuration of the *<Method2 Name>* in Table 4-12.

Table 4-12: Test method to verify background validity: Method 2 configuration

Parameter	Value
Velocity Type	VELOCITY_TGS_6329
Resolution Type	RESOLUTION_4
Low Frequency	650
High Frequency	4000
Numbers of Scans	4
Allowable Out of Range Scans	1

Gain Type	GAIN_1
Aperture	150
Aperture 380	100
Apodization Type	APODIZATION_BEER_STRONG
Phase Method Type	PHASEMETHOD_MERTZ
Auto Set Filter	FALSE
Low Pass Filter Type	LOWPASS_11000
High Pass Filter Type	HIGHPASS_20
Interferogram Maximum	9.8
Interferogram Minimum	2
Peak Minimum	10
Absorbance Peak Minimum	-0.50
Absorbance Peak Maximum	3.00

Results

Once the analyser settings are confirmed on both the methods, Method 1 is then initiated through ECS server. A valid background is taken using Method 1 on Location1 making sure the new background is valid for 30 minutes as set in the initial configurations with no further sample measurements taken.

New set of measurements are then initiated using Method 2 on Location 1. A valid background is taken with expiry time of 30 minutes. The measurements are then aborted and aperture setting on analyser for sample and background measurements is changed from 100 to 80. On initiating a measurement using Method 2 on Location 1 it could be seen that the background was now invalid as expected. To further verify if the changes on analyser setting for Method 2 affected Method 1 a new set of measurements was initiated using Method 1. It was observed that the background on Method 1 which was set to expire after 30 minutes was

still valid. This test confirmed that a separate background for analyser with same settings is not required.

4.4.5 ADI Alarms

Alarms and Events are used to inform the operator of the status of the system. Events give information regarding changes in the process and other operational occurrences. Events do not normally need immediate operator attention. Alarms are messages that alert an operator to an abnormal process or system state. Alarms generally need operator attention or actions. All alarms are events but not all events are alarms.

Alarms can be split in two groups 'model alarms' and 'process alarms'. Model alarms are configured and triggered (during method execution) within the prediction model and specifically reflect model deviations. Process alarms are configured and triggered within the ADI server and are used to reflect process deviations. Both model and process alarms are generated and time stamped within the ADI server.

4.4.5.1 Prediction Alarms

Prediction alarms are triggered in the run time model when the measurement values for the predicted properties and statistics are outside the Alarm and Warning Limits that are configured within the model. It is also possible to apply a separate process alarm to the predicted value which is configured within the method. This would be used where the process requires different limits of alarm from the model. In both cases the following limits are available as shown in Table 4-13.

Table 4-13: Alarm Descriptions

No.	Alarm	Model Alarm Description	Process Alarm Description
1	Warning Lo, Warning Hi, Alarm Lo, Alarm Hi	A model tag with a defined alarm will go into alarm when predicted properties or statistics drift outside of the operating parameters for the model.	A process tag with a defined alarm will go into alarm when the value of the tag moves outside the limits defined for process operation.

4.4.5.2 Sample Diagnostic Alarms

Sample diagnostic alarms are triggered in the run time model when the measurement values for the Predicted Properties and Statistics are outside the Alarm & Warning Limits configured in the current model instance.

4.4.5.3 Background Spectral Diagnostics Alarms

Background spectral diagnostic alarms are generated in the ADI server when the background diagnostic checks are outside the Alarm and Warning Limits configured in the current model instance.

4.4.5.4 ADI Alarm Test

In order to be able to carry out alarm test models were developed such that would be able to demonstrate the functionality of the model alarms. It was also required to configure csv. files in a manner that various levels of alarms would be generated in order to successfully complete the defined test.

Purpose of Test

To verify both model and process alarms are generated and displayed correctly with an alarm list in the ADI server.

Test Method

Table 4-14: ADI Alarm Test set

Test Configuration	Value
Method	Property 7
Location	Location 2
CSV file	HM_TestforAllAlarms

For the purpose of this test ABBH.csv file was loaded into the measurement Location 2. On executing the selected method each simulated row in the csv file was arranged so that it would generate model alarms as designed within the model. To generate the process alarms the method was configured by setting the alarm limits under Input tab. The alarm limits were defined in a Method and depending on the predicted value the alarm would be classified as

either Warning Low (<10), Alarm Low (<0), Warning High (>30<40) or Alarm High (>40) The resulting process values have been recorded in Table 4-15.

Table 4-15: Process Alarm

Alarm Type	Predicted Value	Alarm Setting
Warning Low	4.08	<10
Alarm Low	-17.84	<0
Warning High	38.61	>30 <40
Alarm High	60.8	> 40

4.4.6 Model Data Export

As previously explained in Section 4.3.1.6 the MDE is an important tool for chemometricians to access the spectral data along with relevant predicted values and lab values. Being able to access the right data at the right time is crucial to continuously improve and optimise the process operations. The FAT's were planned such that it would cover the various combinations of data export as per customer requirements. The following screen shots (Figure 4-10 to Figure 4-17) give an overview into the various filters that can be applied while exporting the data. The names of the models, methods, locations and property have been blanked out for confidentiality reasons.

The first step as seen in Figure 4-10 ensures that only the authenticated users may use the data exporter.

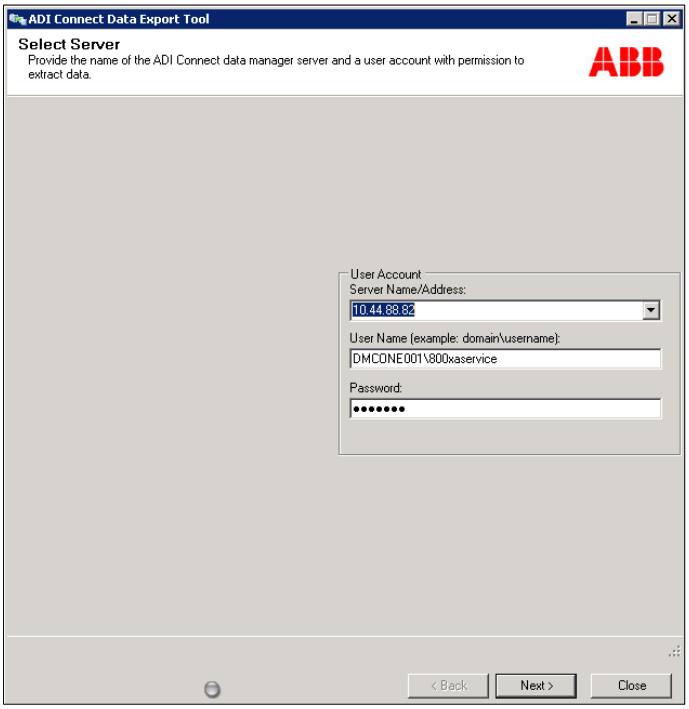

Figure 4-10: Data export screenshot – User Authentication

The next step allows the user to either load a saved query or start a new query to begin the data filtering process.

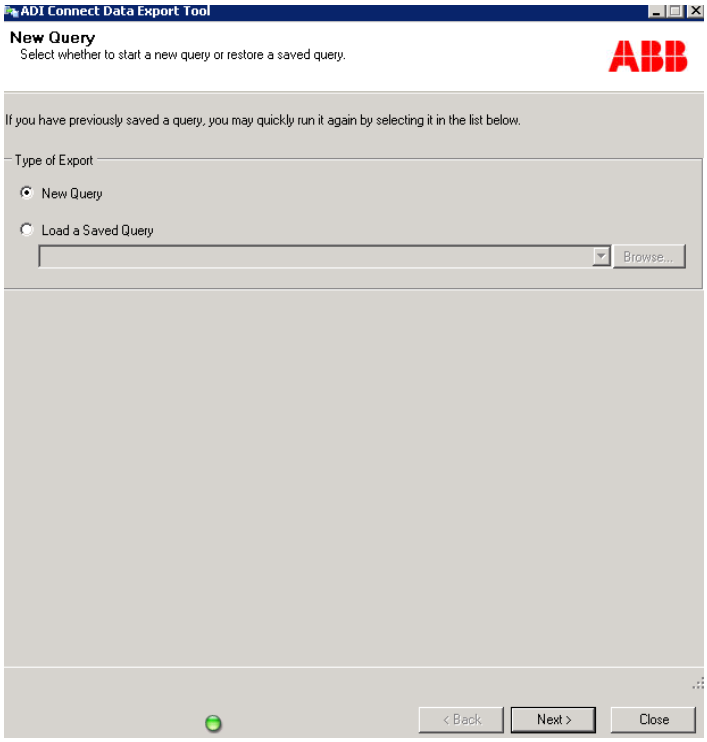

Figure 4-11: Data export screenshot – New Query or Saved Query

Selecting start a new query brings up the first filter which allows the user to select a model.

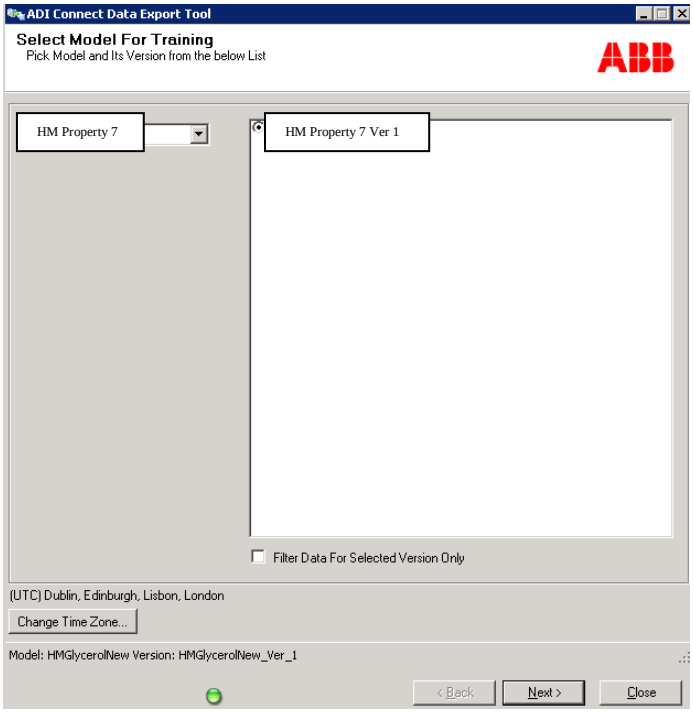

Figure 4-12: Data export screenshot – Model Selection

The next screen shows all the available methods and method versions. The user can select one or more methods or method versions that have used the selected model.

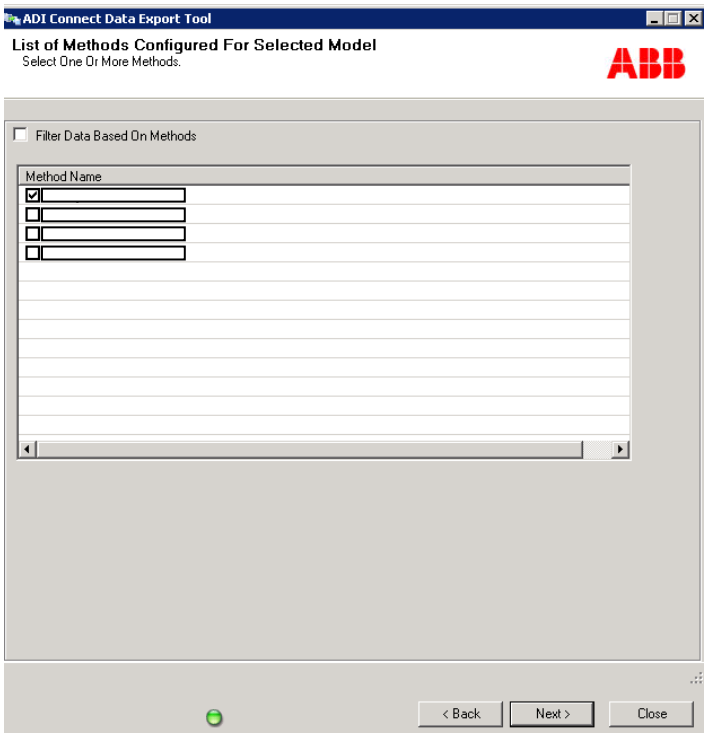

Figure 4-13: Data export screenshot – Method Selection

The next two screens have various options to apply filters on number of variables such as location, time window, alarms, lab values etc.

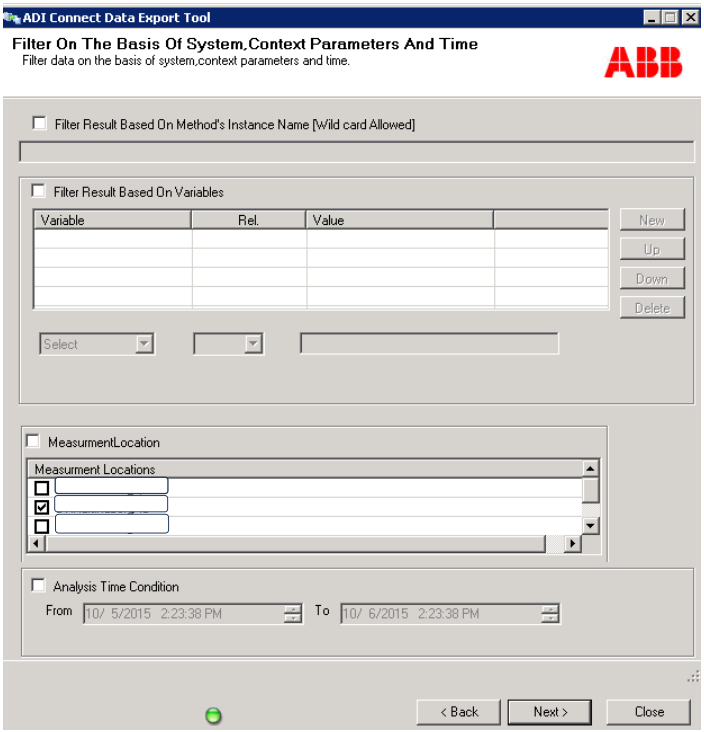

Figure 4-14: Data export screenshot – Variable Filter

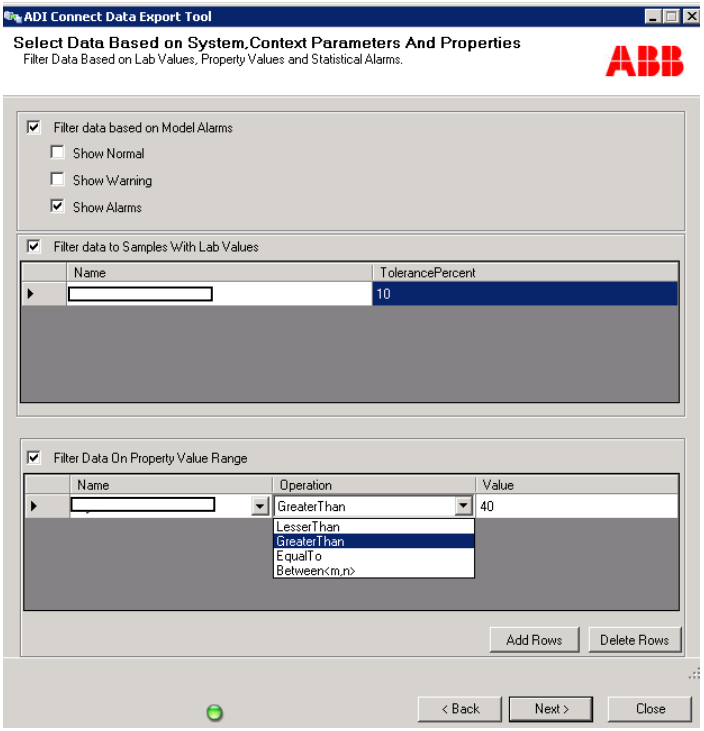

Figure 4-15: Data export screenshot – Variable Filters

The final filters are then applied on selecting templates for data export and also for generating the data either based on the predicted values or spectral data.

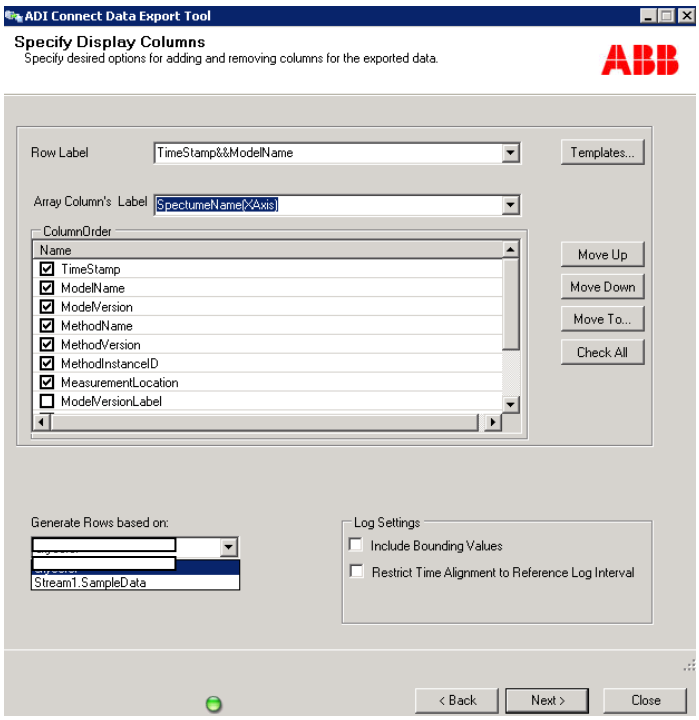

Figure 4-16: Data export screenshot – Template Selection

The last screen of the data export tool indicates if data has been successfully exported.

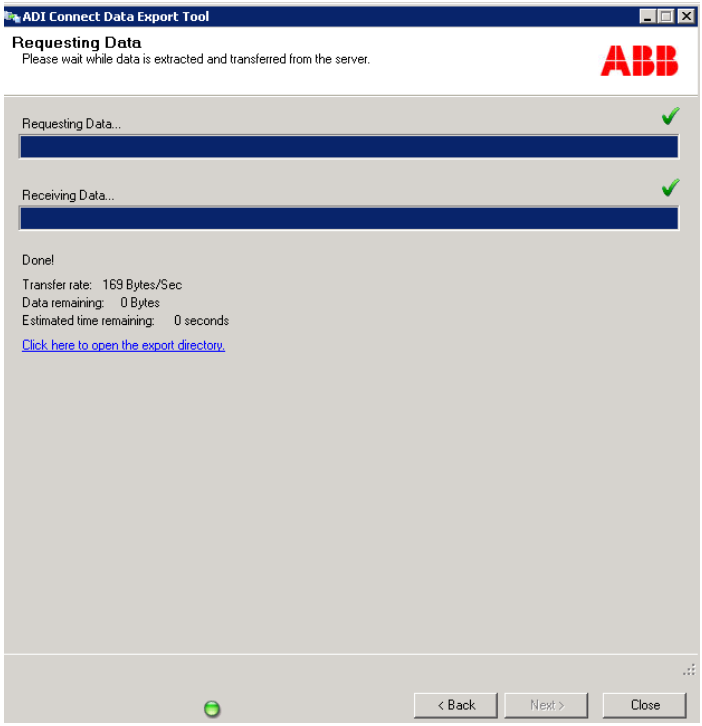

Figure 4-17: Data export screenshot – Data Export Outcome

3.1.1.1 FAT for Data Export

A number of tests were carried out to test the robustness of the MDE and optimise the tool for customer use. A number of issues were identified at the testing stages associated with model selection, method association, time filter, template layout etc. that were addressed and resolved prior to FAT.

Table 4-16 lists the various tests performed and the purpose of various test scenarios. The first two tests were standard export of sample data and background data while the latter case studies looked at various combinations of test sets developed to verify the robustness and capability of the MDE for various data export scenarios.

Table 4-16: Data Export Test List

Type of Test	Purpose of Test	Result
Data export of Sample Measurement	To verify that method data can be exported from ADI server into Unscrambler 10.2 by use of filters	Data successfully exported using filters on Models, Method, Method Instance ID and Measurement locations.
Export Background Data	To verify Background data can be exported	Data successfully exported for filter applied on the background models.

Export Lab Data	To verify that the lab data can be exported in Unscrambler. Confirm that lab data can be assigned to a spectrum Confirm that data can be extracted and sorted according to presence or absence of lab data.	Data successfully exported to Excel file and Unscrambler.
Case Study 1	To verify that it is possible to export only data with lab results. To confirm if it possible to filter data according to an interval of sample number and that the filtered data may be exported from the ADI server into Unscrambler 10.2 by use of filters against test data supplied by the customer.	Data was successfully exported for the interval sample number and lab results as supplied by the customer

Cast Study 2	To verify that model property position in ADI has no influence on the filtering. Verify that only lab results can be exported Verify that the measurements are tagged correctly according to ECS and that it is possible to filter data according to an interval of sample number. Verify that data may be exported from ADI server to Unscrambler 10.2 using filters against test data supplied by the customer	The property position in every model is verified followed by taking new set of measurements. For the data export the filters were applied to sample number, lab data which successfully exported as per the test data supplied by the customer.

Case Study 3	To verify that it is possible to filter data on different parameters such as material number, UserID and batch numbers and that the data may be exported from the ADI server into Unscrambler 10.2 by use of the filters against test data supplied by the customer. The purpose is also to verify that only data with lab results are exported and that these are tagged correctly according to following parameters: • Sample Number (ECS) • Material number (ECS) • UserID (ECS) • Date/time (replicates) (ADI) • Instrument (Measurement Location) (ADI) • Model versions (ADI) • MethodID's (ADI)	Data successfully exported applying various filters as per customer requirements.

Case Study 4	To verify that method data may be exported from the ADI server into Unscrambler 10.2 by use of filters against test data supplied by the customer The purpose of this test is to verify that stat alarms/warnings are correctly regenerated in Unscrambler (by the HM Predictor) after export of spectra with alarm summaries of alarm or warning and compare with specific values in ECS.	The data was successfully exported applying the relevant filters and the generation of alarms was verified by comparing using the HM Predict functionality in Unscrambler 10.2 for HM Property 7 model
Case Study 5	To verify that it is possible to use saved queries.	For this test a new query was created in the data exported with selection on various filters such sample number, Location, Model alarms and Lab data. This Query was saved and exporter tool closed. To confirm the functionality of the saved query a new instance of data export was started by loading the saved query and exporting the data within Unscrambler. The data was successfully exported using the saved query.

Case Study 6	To verify that it is possible to trace instrument settings (e.g. numbers of scans, aperture size, gain, accessory ID etc.)	The purpose of this test was to be able to trace instrument settings using the exported data. The data was exported using the filters on the model and the model version as well. The subsequent methods used with those models were then selected with further filters applied on the data export. As the data is exported with the method version it can be used to trace the instrument settings configured in the ADI server.

4.5 Chapter Summary

This chapter looked at the PAT solution that was developed by ABB as per the customer requirements. With various components of the software such as the third party analysers and the Chemometric packages having OPC interface this product can be considered as the next step towards the successful implementation of the PAT concept within the pharmaceutical, bio-pharmaceutical, chemical and food & beverages industry.

The chapter also discusses the in-depth scenarios of the various tests that were specifically designed by the author as a part of this project for the successful completion of the FAT's. Since the product was novel in the market the system was configured and test sets were developed to showcase the working functionality of the product. The authors contribution to the ADI software involved integrating the existing chemometric models with the ADI system, generating the alarms within the ADI server using a simulated csv file and further being able to export the data to be accessible for continuous improvement and process optimisation. This not only required the understanding of chemometric model development but also the functioning of the ADI software. The work discussed in the thesis focussed on developing solutions to a number of issues that were resolved prior to conducting the FAT.

This included:

- addressing the method compatibility issue,
- validation of background for various settings on the analysers and methods,
- identifying the problem associated with the size of the model and suggesting modification of the imported models into ADI by stripping out passive data not required for property prediction,
- generating required alarms within the system by manipulating the csv files and
- ensuring successful export of data using the MDE.

After completing the FAT's the product was successfully commissioned on site and has been taking regular at-line measurements from various parts of the world from a central location.

Project 2: Using Multivariate Analysis Techniques to Identify Crystallisation Problem in Polymerisation Reactor

5 Chapter 5 – Background information to the crystallisation problem

5.1 Chapter Overview

This chapter gives an overview of the project undertaken with an ABB customer who manufacture speciality polymers and have been facing an unknown crystallisation problem in their polymerisation reactors.

Section 5.2 elaborates about the crystallisation problem addressed by the customer and possible multivariate solution that could be used to solve the issue.

Section 5.3 further introduces to the problem in detail.

Section 5.4 looks at the measurement and control for the polymerisation reactors reported in literature and looks at the actual control of polymer reactors at the customer site.

Section 5.5 discusses how multivariate analysis could be used to solve the problem.

Section 5.6 is the chapter summary.

5.2 Introduction

This project looks into a real industry problem in polymerisation process due to the occurrence of unexpected crystallisation in the reactor. Polymerisation is a process where polymers (plastics) are formed when monomer molecules react together to form linear chains or three dimensional network of polymer chains. There are two main kinds of polymerisation reactions: 1) addition polymerisation where monomer molecules add on to a growing chain one at a time with the process taking place over three steps which are initiation, propagation and termination and 2) condensation polymerisation where single monomer molecules are joined together to form polymer chains while forming by-product such as water. Polymers such as thermoplastics are formed through addition polymerisation.

In addition polymerisation monomer x gets converted to polymer x with no by product formed during the reaction. In condensation polymerisation two or more monomers form a polymer plus a by-product is formed which is usually water. Condensation polymerisation can either form thermosetting polymers or thermoplastics. Thermoplastics are class of plastic that can be re-shaped and remoulded a number of times by heating and consecutive cooling of the polymers. Thermosets on the other hand cannot be reheated or reshaped once they are formed (McArthur and Spalding, 2004).

The customer is the world's leading manufacturer of high performance Poly-ether-ether-ketone (PEEK) polymer, a type of Thermoplastic that is used in aerospace, automotive, electronics, energy, medical and semiconductor industry. According to Zalaznik et al. (2015) the high performance polymer is becoming increasingly popular due to its excellent mechanical and thermal properties and are replacing metal in number of industries.

At the customer site polymerisation takes place in a vessel containing 4 monomers and a solvent. The weights/proportions of the monomers are critical to the viscosity of the final polymer produced. With processing temperature of PEEK polymers extremely critical on the tribological properties of PEEK material Zalaznik et al. (2015) has reported significant impact of varying temperatures on the final polymer product and identified the importance of temperature on the final polymer product. The temperature control of the reactor is crucial to the process and in this case is controlled by means of a cascade control loop with contents temperature as the master loop and jacket temperature as the slave loop. The vessel contents are heated following a temperature profile which allows for reaction gases to be released in a controlled manner. The temperature profile is critical to the process as the product can crystallise/precipitate if sufficient heat is not applied. The temperature profile contains two holding points where the temperature is maintained at a constant level to allow gas evolution whilst mitigating the risk of crystallisation. The reaction is ultimately controlled by agitator power, the aim being to produce a polymer of a particular Molecular Weight (MW).

The variability in MW gives them product quality issues, as does crystallisation (the engineers believe this could be leading to an increase in black spec as the crystals can scour off residue from the polymer vessel wall) and increase in level of foaming (leading to contamination from residue off the top of the vessel).

The customer currently record the following online process values: jacket temperature, contents temperature, agitator power, agitator power rise, gas evolution (on some vessels), agitator speed and level (for foaming) and the following lab measurements: contamcount, moisture, precipitation, and molecular weight. Additional single point measurements taken for OEE purposes are: DPS, Hold 1, Hold2, Heat_250, DPSCharge, BDFDIFF, HQDIFF, SCDIFF, PolymerConcentration, H2LZ, Heat_Poly and BxDelay. The detailed explanation of the measured variables can be seen in Chapter 6 Section 6.2.

Currently a crystallised batch is identified by a 'kink' in the contents temperature profile with no other quality variable able to further verify this unusual behaviour that may have an effect on the final product. However the work reported in the thesis established the fact that the

batches classified as crystallised (identified by monitoring the kink) did have an underlying unusual process behaviour that was different to a normal batch. Thus, although currently no quality variable is co-related to the crystallised batch, the unknown process deviation that manifests itself as a kink in the contents temperature profile is associated with the crystallisation problem reported by the customer.

The primary objective of MSPC that would be performed out on the dataset would be to provide early warning of changes in process behaviour which might be leading to crystallisation during polymerisation. The customer has attempted to identify the crystallisation issue by monitoring variables independently. However univariate SPC technique (the relationship between temperature profiles and crystallisation in this case) only considers the deviations from the target value and does not take into account the inter-relationships that occur between variables. Valuable process information concerning the behaviour of the process may thus be ignored. Analysing the data using PCA method enabled to perform a multivariate analysis of the variables that will take into account the correlation between variables.

Initially PCA models were developed from existing data gathered from 'good' batches – to understand why a good batch is successful. The developed models will then be superimposed onto the 'bad' batches - the purpose of this was to identify correlation between variables and the differences between good and bad batches.

If the verification stage is successful and the root cause for crystallisation is identified using the PCA technique, the models can be further developed to handle a range of polymer reactors, products, grades or recipes. This would be an extension to PCA which would allow a number of grades to be monitored through a single generic model.

Upon the successful creation of a prediction model capable of predicting 'bad' batches across all grades and reactors the project could further aim to propose an online solution which will provide process operators with early warning of batch deviation so that action can be taken to save the batch either manually or if possible by automated process control. Figure 5-1 is a workflow of the possible automated solution that would be suggested to predict crystallisation in the reactors.

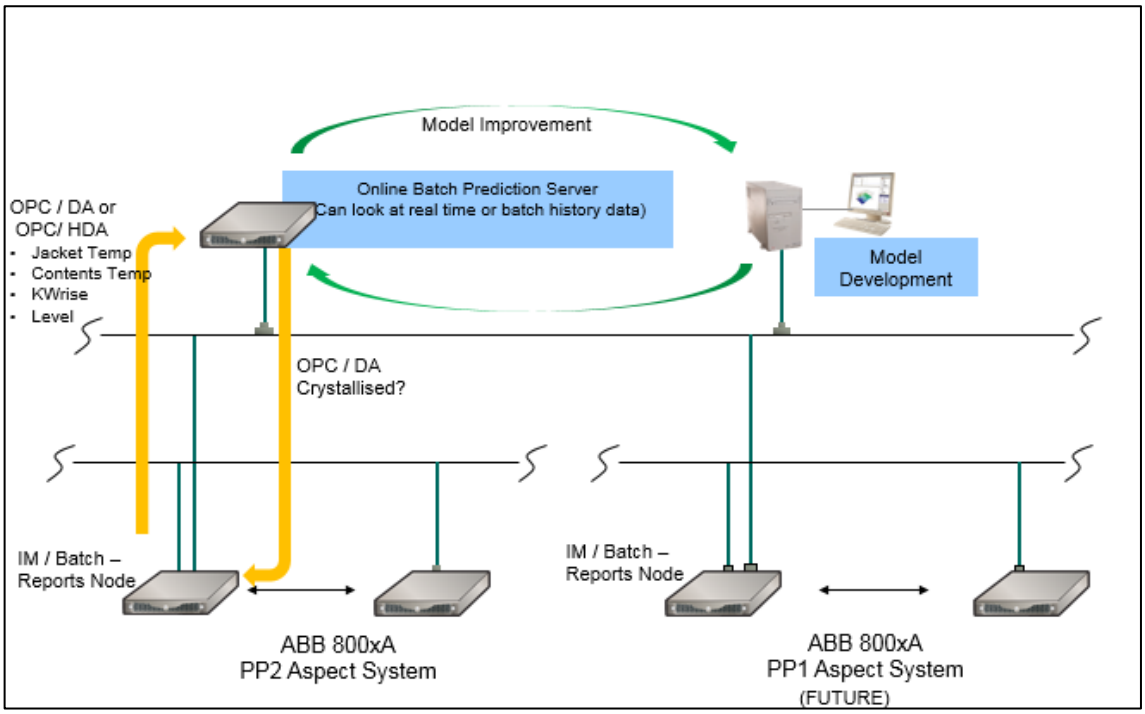

Figure 5-1: Automated solution to predict crystallisation

5.3 The problem

The customer is the leading global manufacturer of high performance polyaryletherketones, including the versatile PEEK polymer. They have two operational powder plants with a total of six polymer reactors. A third plant is currently being built which will bring the total number of polymer reactors to ten.

For some time now polymer reactors periodically exhibit an unexplained behaviour where process upset during the polymerization reaction leads to a phenomenon described as ‘crystallisation’. This manifests itself as a ‘kink’ in the temperature profile after hold point 2 during the heating cycle of the polymer reactors. If this kink occurs it is widely believed that crystals will be formed which will directly affect the quality of some product grades. This leads to the down grading or in the worst case total loss of a batch which in turn leads to loss of revenue.

The customer has identified a number of theories as to why ‘crystallisation’ occurs during polymerisation. It has however been difficult to verify these theories as it takes a considerable amount of time to align the relevant data in order to look at the problem and even when this has been done there is no obvious single cause that has yet been identified.

Polymerisation takes place in a vessel containing 4 monomers and a solvent. Weights and proportions of the monomers are critical to the final product quality. Figure 5-2 shows a typical sequence for the addition of monomers to the reactors. Hydroquinone (HQ) and

Diflurobenzophenone (BDF) that are the raw materials are first blended in Pre-Melt Vessel 1 (PMV 1) and Na_2CO_3 followed by K_2CO_3 is then added to the Pre-Melt Vessel 2 (PMV 2). PMV 1, PMV 2 along with Diphenyl Sulphone (DPS) is then added to the polymerisation reactor. There is a cascade control loop around the jacket where the contents temperature controller is the master loop while the jacket temperature controller is the slave loop. The heat up profile, raw material quality and proportions are considered to be critical to crystallisation.

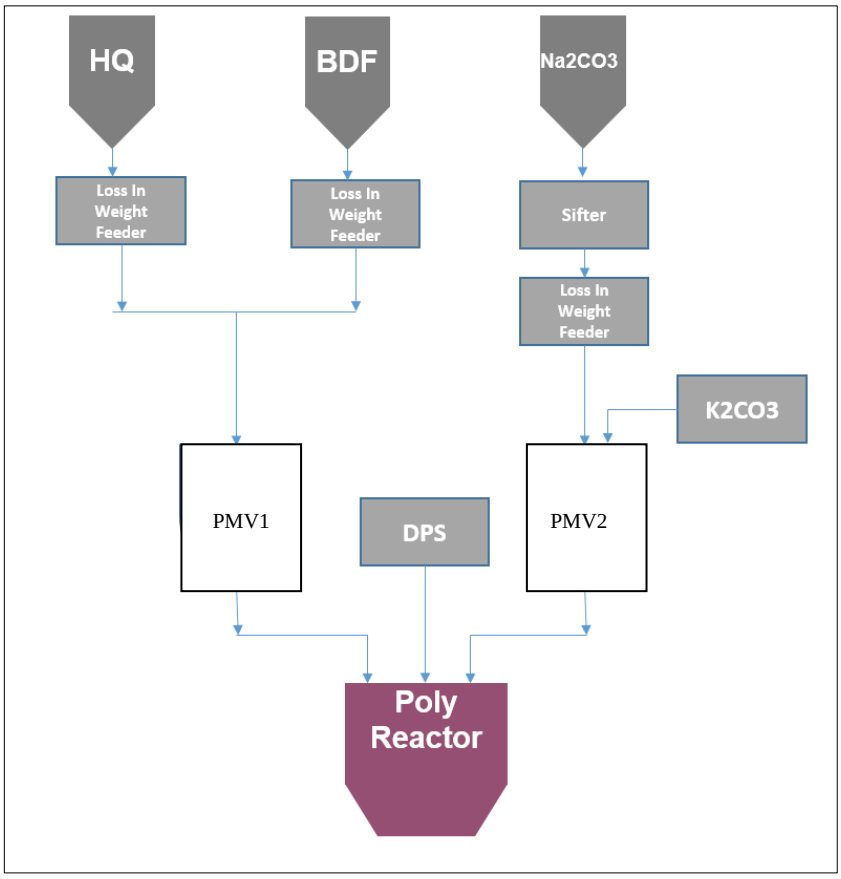

Figure 5-2: Polymer Reactor

Possible crystallisation reasons that have already been identified by the customer are:

- Insufficient DPS
- Faster than normal reaction rate
  - when incorrect polymerisation temperature profile is used (hold 1 temp to high)
  - Na_2CO_3 with a smaller particle size distribution
- Insufficient heat input to the contents of poly vessel.
- Holding polymerisation at contents temperature greater than 160°C
- Sudden increase in KW rise at the end of hold 2

- Sudden increase in contents temperature during heat up to polymerisation (i.e. hold 2 to end of polymerisation)

The customer has been unable to identify one particular variable which defines crystallisation but have instead defined a batch as likely to be crystallised if there is a kink in the temperature profile after hold 2 in the contents temperature. This study will focus on the identifying possible variable deviation which could lead to the 'kink' in the profile. The heat up profile of typical batch can be seen in Table 5-1.

Table 5-1: Typical heat up profile of polymerisation reactor

Process	Temperature	Duration
Poly charge	135°C	40 minutes
Heat to Hold one	135°C to 180°C	30 minutes
Hold one	180°C	100 minutes
Hold two	180°C to 200°C	40 minutes
Heat to Poly	200°C to 302°C	120 minutes
Polymerisation	305°C	120 minutes + 'cooking'
Endcap	305°C	30 minutes

Figure 5-3 shows the heat up profile of good and a bad batch. The process usually starts at around 150 °C after which it is heated to hold 1 up to 180°C where it is held for about 100 minutes. It is then further heated to hold 2 till 200°C and held there for another 40 min following heat up to polymerisation temperature of 302°C and final endcap up to 305°C. As in the figure unlike a good batch a bad or crystallised batch has a kink just after hold 2 temperature of 200°C. This distinctive temperature rise in the profile has been observed in every batch that has been classified as crystallised batch.

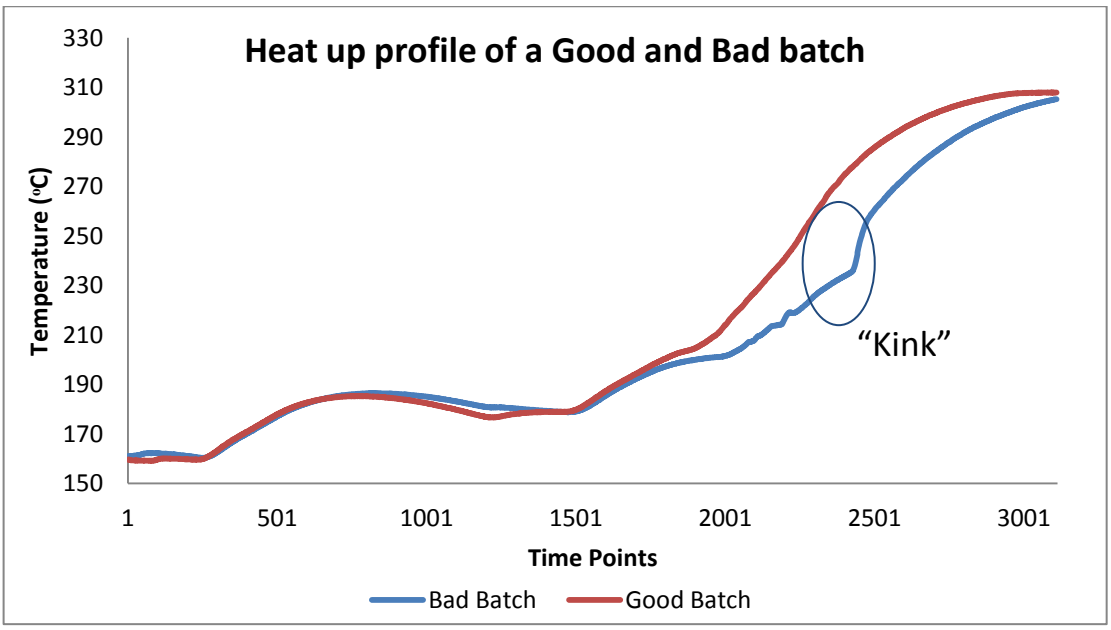

Figure 5-3: Heat up profile of a typical Good and Bad batch

Chapter 6 further discussed in the thesis looks in the application of multivariate methods to identify the cause of the problem using historical data that was provided by the customer.

5.4 Measurement and control of polymerisation reactors

Batch reactors are more commonly used in polymerisation processes as they allow flexibility to process multiple products according to its own recipe with variable operating conditions. These reactors are also appropriate for low volume product and products with multiple grades which is commonly observed in speciality polymer production. Thus with variability between batches process control of the batch reactor is an important factor in order to be able to produce polymer with consistent properties in its final shape and form, for the intended application. Fundamental control models play an important role for the control of polymerisation reactors due to a number of reasons such as the lack of available online sensors, the complex polymerisation process and the non-linear operating space of batch and semi-batch reactors (MacGregor, 1988). Richards and Congalidis (2006) have discussed the measurement and control techniques based on their hierarchical approach framework as seen in Figure 5-4.

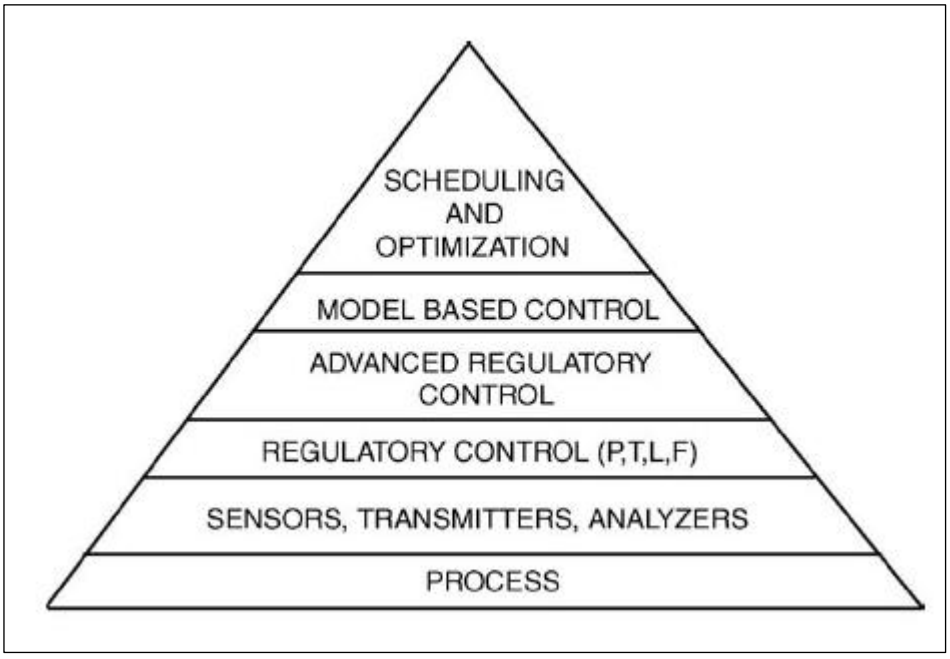

Figure 5-4: The process control hierarchy (Richards and Congalidis, 2006)

Traditionally in the polymer industries the variables measured in order to ensure robust control are process variables such as pressure, temperature, level, flow (PTLF), density, viscosity and quality variables such as composition of raw materials and final products, surface tension, molecular weight distribution and particle size distribution. The PTLF measurements are the fundamental measurements required for regulatory and advanced control strategies. In the case study discussed further in this report the temperature control around the reactor is maintained using a cascade control loop. A number of chemical

processes with non-linear behaviour have adapted the cascade control loop since it addresses the drawback of feedback control where deviation of controlled variable from a setpoint is the only time a control action occurs. With polymer properties being very sensitive to temperature changes cascade control resolved the hindrance in feedback and feed forward control by the addition of temperature on the reactor feed. Figure 5-5 illustrates the cascade control loop that is typically used in industry. The actual control for the reactor at the customer has been further discussed in Section 5.1.1.

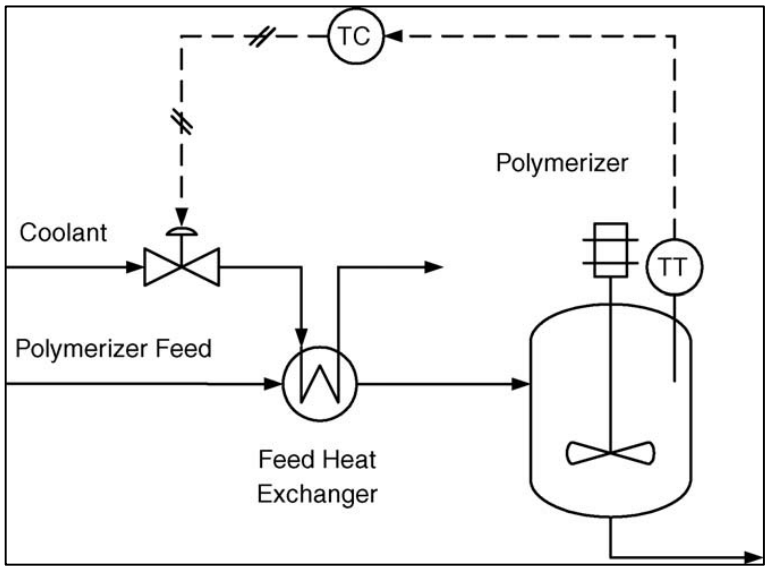

Figure 5-5: Conventional cascade control of polymerisation reactor (Richards and Congalidis, 2006)

For composition measurement of raw material and finished product samples are normally send to the lab for analysis. However sampling and analysis in labs is time consuming and by the time the results arrive it is too late to make any control changes to optimise the process. Recently a number of advances have been made in sensor technology to allow online measurement of quality variables. Fibre optic linked devices such as Raman and NIR along with advancement in chemometrics have expanded the capabilities of these devices. Ohshima and Tanigaki (2000) have also looked into the modelling of inferential systems for polymerisation processes which can predict the quality of variables such as molecular weight, conversion, melt index, density etc. from process variables such as temperature and concentration in the reactors. The models developed have been categorised into three groups which is phenomenological model based on the models developed from first principles, empirical models derived from laboratory data and multivariate statistical models such as the PCA and PLS.

Final polymer quality depends on the molecular weight averages of the product during the polymerisation process. Gel permeation chromatography (GPC) has been used traditionally to monitor this variable.

Particle size of raw material is another important variable which is a critical parameter for process performance and final product quality. A number of particle size measurement techniques are used in the industry such as optical imaging, electron imaging, optical diffraction and scattering, electrical resistance changes, sieving, sedimentation and ultrasonic attenuation. Measuring only the average particle is not always enough with the presence of different sizes in the population that could result in multimodal distribution. For the purpose of identifying the crystallisation problem the customer have monitored the raw material particle size of Na_2CO_3 for a short period of time using three different sizes of sieve. The purpose of this experiment was to be able to identify if variation in particle size of the raw material affects crystallisation in the polymerisation reactors. The particle size data was used to analyse the crystallisation problem and the results have been discussed in Chapter 6.

5.1.1 Polymer reactor control at customer site

The customer uses a cascade control loop to maintain the polymer reactors at the required temperatures. The reactor is under jacket control which is the slave loop during the initial monomer addition at the start of the polymerisation reaction. Once the set point temperature of 145°C is achieved control is then switched to the contents control which is the master loop. The contents temperature control is a split range control. Once the control is switched to contents temperature the set point ramps up to hold point 1 with rise in a degree every minute. The temperature probes used to monitor the temperature of the reactor have an accuracy of 0.1% and are calibrated once a month.

Figure 5-6 looks at the Hot Oil Service for polymer reactor 5 and polymer reactor 6 that is used to control the jacket temperature as well the contents temperature. The hot oil services supply the hot oil to the jacket of the reactor. Depending on the set point the hot oil either passes through the cooler before entering the jacket or by-passes the cooler completely before being recycled into the reactor. Figure 5-7 is the screenshot of the faceplate used to control the reactor temperature for Poly 5 (Faceplate no – TIC40606) and Poly 6 (Faceplate no – TIC 40607).

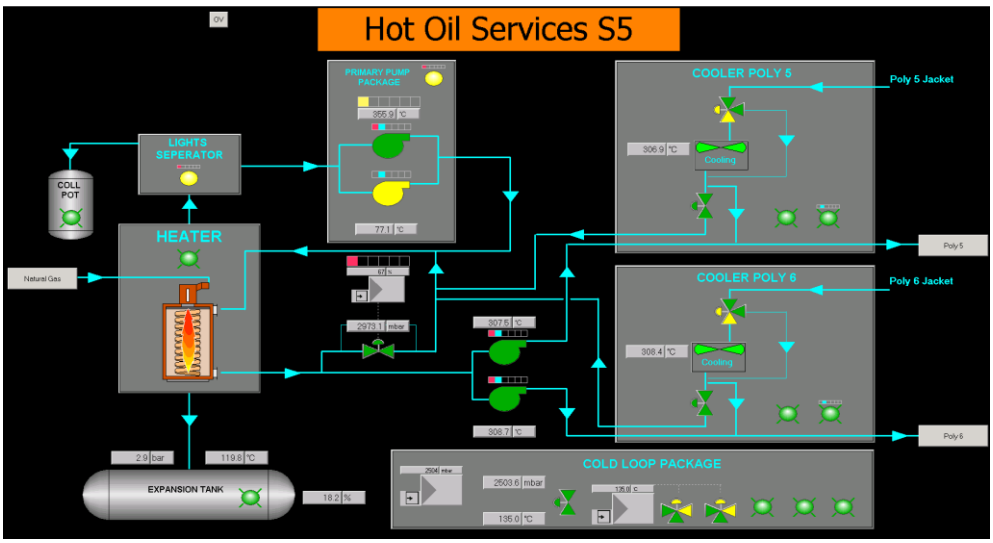

Figure 5-6: Screenshot for hot oil services

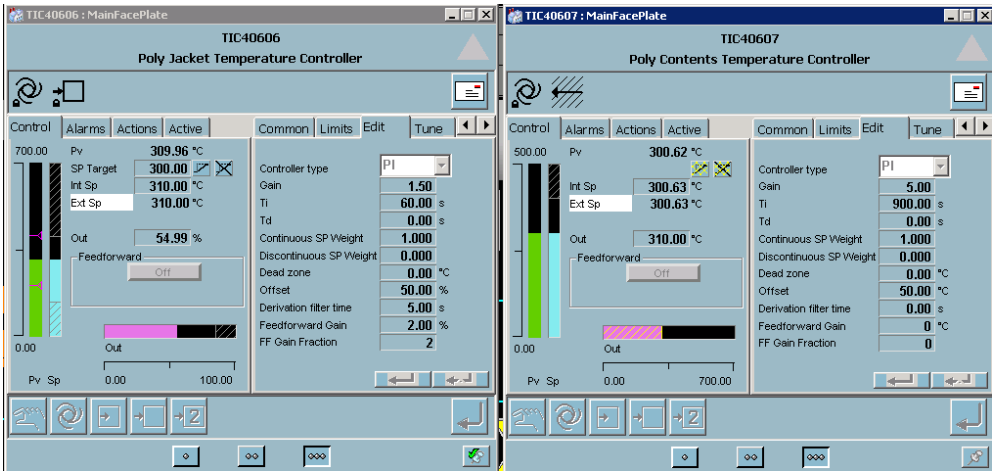

Figure 5-7: Screenshot faceplate of the cascade control

5.5 Current Applications of MSPC Methods Reported in Literature

There are a number of techniques to monitor/control the processes in real time: however this report will explore MSPC techniques applied in industry. In recent years SPC has gained importance in manufacturing industries with the need to monitor and control the processes in real time. As mentioned previously traditional SPC techniques which use SPC charts such as Shewart, CUSUM and EWMA to monitor and control the process variables are no longer satisfactory in modern process industries where large amounts of process data are collected containing variables with complex relationships. SPC which aims to identify cause for process deviation is different from a typical feedback control loop which functions by compensating for the process disturbances. Traditional SPC was based on monitoring only a few variables using univariate control charts. With the advent of computers and advanced measurement techniques, large amount of process and product quality data are being collected for reporting and continuous improvement of processes. Process measurements such as pressure and temperature are measured almost every second while quality data such as polymer molecular weight or viscosity are measured less frequently. However monitoring this data univariately or independently might be of little use since a number of variables in the process are correlated and need to be interpreted relative to each other. Multivariate methods analyse the variables simultaneously and also make it easy to detect an event otherwise unnoticed due to low signal to noise ratio in each variable.

Monitoring a process using univariate control charts further developed into multivariate process control using the traditional monitoring charts such as Shewhart, CUSUM and EWMA. The use of these charts was practical in situations where less number of process and quality variables were monitored. These days most of the process operations measure and store large number of variables to monitor their processes. However with the highly correlated data and low signal to noise ratio the historical data is not utilised to its full potential to identify the key issues in a process. Multivariate statistical projection techniques such as PCA and PLS previously discussed in Chapter 2 tackle these issues by reducing the dimensionality of the variables in the principal component space. While PCA develops a predictive model for the process data (i.e. in the X data set) PLS is used to model two matrices such as process data (X) and the corresponding quality data (i.e. in the Y data set). Thus PLS analysis enables the prediction of quality attributes of the final product in future batches based on the model developed using historical batches. Any unusual event that would affect the final product quality would have its fingerprints in the process behaviour. In order to diagnose the occurrence of a special event multivariate control charts such as SPE and Hotelling's T^2

provide a good indication of an out of control batch process. By further investigating the PCA model, the contribution plot for the scores and the monitoring statistics such as the SPE and Hotelling's T^2 multivariate analysis enables the diagnosis of the underlying problem in the process.

A number of examples explaining the monitoring of polymerisation using multivariate projection techniques have been reported in literature. MacGregor and Kourti (1995) have described an example situation where the application of MSPC method to monitor a polymerisation process has been investigated. In the work published the data was unfolded batch wise and with the available process and quality data a reference model was developed. This calibrated or reference model was then used to classify a new batch as good or bad by monitoring the Hotelling's T^2 and SPE at each time interval. In another example of polymerisation control Ohshima and Tanigaki (2000) have proposed an integrated solution that focuses on on-line sensing and optimal grade changeover control. One of the major issues discussed in the paper is the plant wide control of polymer quality not only in the reactors but also in the blending and extruding operations. Lack of on-line sensors has been one of the major obstacles to ensure the quality control of polymer properties. Developing an inferential model using the process data to determine the polymer quality has been suggested in a number of publications. For example Khatibisepehr et al. (2013) have discussed the application of inferential models using three design procedures: 1) knowledge driven model based on first principle model, 2) data driven model also known as black box model and 3) gray box model that is a combination of first principle model and black box model. Fevotte et al. (1996) have highlighted the application of calorimetric sensors to perform an inferential estimation of unpredicted variation in a batch polymerisation process and Sharmin et al. (2006) have demonstrated the application of inferential PLS models to predict polymer conversion using measured process variables. Not only would these models enable online monitoring of product quality but also allows continuous process improvement. The application of the first principle models however can be complex due to the simultaneous estimation of kinetic parameters. At the same time calorimetric sensors can be prone to error due to noise in the temperature measurement and variations to heat transfer estimation parameters which are dependent on polymer conversion rate in the reactor. For the successful implementation of inferential models detailed knowledge of process variables would be helpful to design control systems to monitor and improve process operations. Figure 5-8 describes the typical inferential model that can be used to determine the polymer quality based in the process data. Thus every time quality variable is available from the lab it is

updated by taking the difference between the predicted and the actual quality value. The updated model is then used to predict to the quality of the variable on-line until the next measurement is available from the laboratory. Faggian et al. (2009) has reported the application of MSPC technique to not only predict the product quality in batch process but also estimate the batch length in real time thus allowing the scheduling of manual intervention, optimisation of manpower and the forecasting of production time which in turn enables the optimal utilisation of plant equipment.

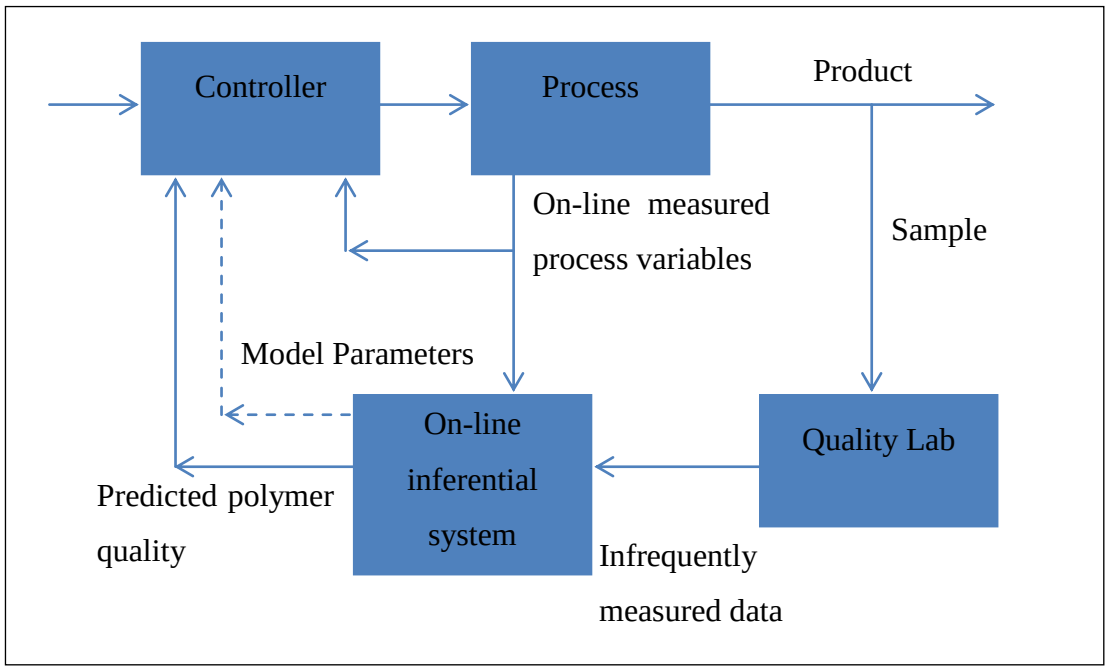

Figure 5-8: Inferential model to predict product quality

5.6 Chapter Summary

Extensive work has been reported in literature studying the application of MSPC techniques to monitor and control industrial polymerisation reactors. This project aims to identify an unusual crystallisation issue in the polymerisation reactors using the techniques reported in literature. What makes this problem unique is the missing link between process variables and quality variable or laboratory measurement to identify the root cause of crystallised batches.

Exploratory data analysis has been performed on the historical process and quality data provided by the customer that have been further explored in Chapters 6. Crystallised or bad batches have been projected on the calibrated model to examine how the bad batches behave differently from the good batches and if crystallised batches cluster out separately as outliers. The next chapter investigates the various types of data (process and quality) that are used to identify the problem in the reactors using multivariate methods.

6 Chapter 6 – Application of Multivariate Analysis Methods to Identify the Crystallisation issue

6.1 Chapter Overview

This chapter carries out an analysis on the data provided to determine a relationship between the crystallised batches and the measured variables. The aim of the study was to identify the root cause of crystallisation using the MSPC.

Section 6.2 explains how the crystallisation problem will be addressed using MSPC techniques within the scope of this project.

Section 6.3 analyses the Quality and 'Static' process variables by developing a single PCA model for two reactors.

Section 6.4 investigates only the quality data for Polymerisation Reactor 5 and Polymerisation Reactor 6.

Section 6.5 further looks into the analysing the problem using the process data. This required unfolding the three dimensional process data to two dimensional data prior to applying the PCA model.

Section 6.6 investigates the effect of particle size on the crystallised batches.

Section 6.7 further investigates the additional process data provided by the customer and using the knowledge obtained from the previous process data analysis this section outlines the fingerprinting technique for identifying a typical bad batch.

Section 6.8 summarises the chapter.

6.2 Introduction

The problem will initially be looked at by applying PCA on the quality data, process data and particle size data. The quality data was available from the years 2008 to 2014 on all the six reactors for all the grades. With a typical batch time of approximately 8 hrs each reactor processed around 2 batches per day. Initially the process data was provided by the customer for Polymerisation Reactor 5 (Poly 5 or P5) and Polymerisation Reactor (Poly 6 or P6) from December 2012 to February 2014. For the process data analysis Section 6.5 uses the data that was provided for the first part of the project. Section 6.7 analyses the additional process data provided on Poly 5 reactor along with the data that has been used in section 6.4. With the initial analysis consisting of 19 quality variables and 9 process variables it is expected that these variables would be highly correlated with each other.

Process data at the start of the project on Poly 5 and Poly 6 was obtained for around 1804 batches which consisted of 8 crystallised batches on Poly 5 and 12 crystallised batches on Poly 6. For the analysis a total of 70 good and bad batches on Poly 5 and Poly 6 were selected. This included 50 randomly chosen good batches (24 good batches on Poly 6 and 26 good batches on P5) and 20 crystallised batches (12 bad batches on Poly 6 and 8 bad batches on Poly 5). Also it was reported by the customer that batches producing Grade 450 crystallised more than the other remaining Grades such as Grade150 and Grade 380. Thus it must be noted that only Grade 450 batches were selected for the analysis on both the reactors.

Table 6-1 explains more in detail the genealogy of all the batches used in the analysis.

Table 6-1: Genealogy of batches used for the analysis

Batch Number	Good	Crystallised	Poly 5	Poly 6
B1-B12		✓		✓
B13-B20		✓	✓	
B21-B44	✓			✓
B45-B70	✓		✓	

Data analysis is divided in four parts depending on the type of data used in the study. In the first instance static process points at critical stages in the temperature profile were selected which were then combined with the quality data. A single PCA model was developed for both

Poly 5 and Poly 6 as seen in Section 6.3. Studying the results obtained using the static process data and quality data, individual models for each reactor were developed using only the quality data [Refer section 6.4]. Static process data can be described as measurements of process variables at specific instant during batch progression. The selection of static process points has been explained further in Section 6.3. Process data was investigated for Poly 6 and results have been discussed in Section 6.5. The three dimensional process data (Batches x Variables x Time) was unfolded using the batch-wise unfolding approach to enable batch to batch comparison. Additionally particle size data of Na_2CO_3 which is one of the raw materials added to the reactors was also provided with the corresponding quality data for the reactors. The particle size data however was provided only for Poly1, Poly 2, Poly3 and Poly 4. Section 6.5 in this report looks at exploratory data analysis that was performed on the particle size data provided and the results have been discussed further in this chapter.

Table 6-2 lists the quality variables, Table 6-3 lists the process variables and Table 6-4 lists the static process variables that have been used for investigating the defined problem.

In the following table DPS, HQ, BDF and SC are the raw materials used for the polymerisation process.

Table 6-2: Quality variables description

Variable name	Type of Data	Description
Contamcount	Quality	Amount of blackspec measured in laboratory at the end of a batch
Precip	Quality	Amount of precipitation in the product at the end of batch
Moisture	Quality	Amount of moisture in the product at the end of the batch
MW	Quality	Molecular weight
DPS	Quality	Amount of DPS in the product at the end of the batch
Hold 1	OEE (Quality)	The time for which a batch is held at 180°C in the temperature profile

Hold 2	OEE (Quality)	The time for which a batch is held at 200°C in the contents temperature profile.
Heat_250	OEE(Quality)	The time required for a batch to reach 250°C from hold point 2
Heat_poly	OEE (Quality)	The time required for a batch to reach polymerisation temperature from heat_poly
BxDelay	OEE (Quality)	Overall delay in a batch
DPSChrg	OEE (Quality)	DPS input start of the batch
BDFDIFF	Quality	BDF input start of the batch
HQDIFF	Quality	HQ input start of the batch
SCDIFF	Quality	SC input start of the batch
H2LZ	OEE (Quality)	High level alarm
Polymer concentration	Quality	Polymer concentration

Table 6-3: List of process variables

Variable	Description
Jacket Temperature	Temperature of jacket around the reactor
Contents temperature	Temperature inside the reactor
KwRise	Energy input into the agitator
Gas Flow	Flow rate of gas
Oil Temperature	Temperature of oil into the jacket of the reactor

Pressure	Pressure of the reactor vessel
Condenser	Temperature of the condenser
Level	Contents level in the reactor
Agitator	Speed of the agitator

Table 6-4: List of 'Static' Process Variables

Process Variables	Description
Initial Content Temperature	Starting contents temp
Initial Jacket Temperature	Starting jacket temperature
Temp 1	Contents Temperature at Hold point 1
Jack 1	Contents Temperature at Hold point 1
KwRise 1	KwRise at Hold point 1
Gas Flow 1	Gas Flow rate at Hold point 1
Pressure 1	Vessel Pressure at Hold point 1
Condenser 1	Condenser temperature at Hold point 1
Level 1	Level in the reactor at Hold point 1
Temp 2	Contents Temperature at Hold point 2
Jack 2	Contents Temperature at Hold point 2

KwRise 2	KwRise at Hold point 2
Gas Flow 2	Gas Flow rate at Hold point 2
Pressure 2	Vessel Pressure at Hold point 2
Condenser 2	Condenser temperature at Hold point 2
Level 2	Level in the reactor at Hold point 2
Temp 3	End cap contents temperature
Jack 2	End Cap Jacket Temperature
KwRise 3	Endcap KwRise
Gas Flow 3	Endcap gas flow
Pressure 3	Endcap vessel pressure
Condenser 3	Endcap condenser Temperature
Level 3	Endcap Level in the reactor

6.3 Quality + Static Process data analysis

6.3.1 Materials and Methods

In order to examine the process and quality data together an initial analysis was performed by merging these two sets of data. Four significant temperature points in the polymerisation process which are the Start Temperature of the reactor, Hold 1, Hold 2 and the Endcap Temperature were chosen. The relevant values for the variables were selected for the analysis at these four temperature points in the process.

'Static' points for process variables were selected for every batch which was then combined together with the quality variables. The main purpose of carrying out the analysis in this unique way was to be able to merge the process behaviour of batches with the product quality variables and carry out the analysis. This would also enable identifying the impact of process variables on the product quality variables and thus indicate possible reasons behind crystallisation in the reactor. PCA was performed on this data and the results of which have been discussed in Section 6.3.2.

PLS regression method would have been an ideal way to establish the relationship between quality and process variables. However since no quality variable had a direct correlation to the crystallisation problem, it restricted the application of PLS regression technique.

The heat map or correlation map in Figure 6-1 highlights inter-relationship amongst the various variables related to the process. An expected positive correlation is observed along the diagonal axis on the correlation map. Level 3 which is the level in the reactor at time point 3 is negatively correlated with DPS and the Start Temperature in the reactor. A negative correlation is observed in KwRise 2 and H2LZ which would indicate that high KwRise 2 at time point 2 would result in low level in the reactor at hold 2. PCA on the data set will further investigate the underlying behaviour of these variables influencing the process operation.

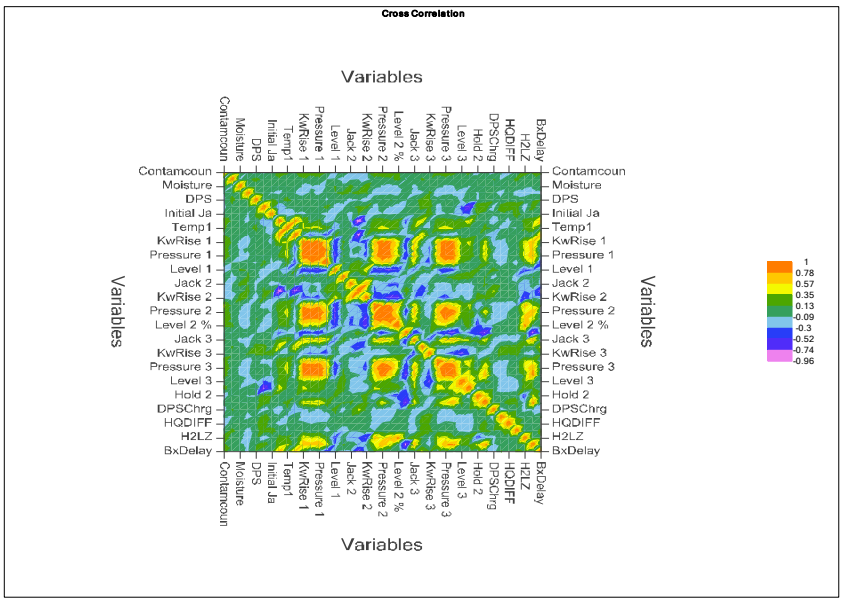

Figure 6-1: Correlation map for Static Process and Quality Variables

6.3.2 Results and Analysis

PCA was performed on the good batches with a total of 41 variables using the PLS_Toolbox in the Matlab software. The scores plot for PC1 vs. PC2 is shown in Figure 6-2. PC1 only explains 28.72% variation in the data set and PC2 explains 12.78%. The lower percentage of

variance captured by the first few PCs may not only be due to the large number of variables used in the analysis but also due to the lack of strong cross-correlation in the dataset. An important observation detected in the scores plot is a clear separation between reactors Poly 5 and Poly 6 is observed in PC1 [Figure 6-2]. This separation indicated a difference in variance for the two reactors suggesting that each reactor is controlled differently for the same product.

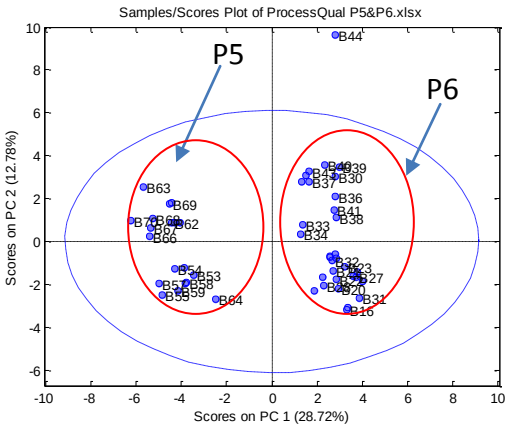

Figure 6-2: Scores plot of P5 and P6 for quality and static process variables – PC1 vs. PC2

This was further verified by looking at the univariate scores plot for PC1 as seen in Figure 6-3. The batches produced by Poly 6 have a positive orientation while the ones on Poly 5 have a negative orientation indicating an obvious difference in the operation of the two reactors. In order to identify the variables that may be responsible for the orientation of the scores it was necessary to analyse the loadings plot as seen in Figure 6-4 and Figure 6-5. However due a random distribution of the variables in the loadings plot it was not possible to establish a consistent relationship between the batches and the variables

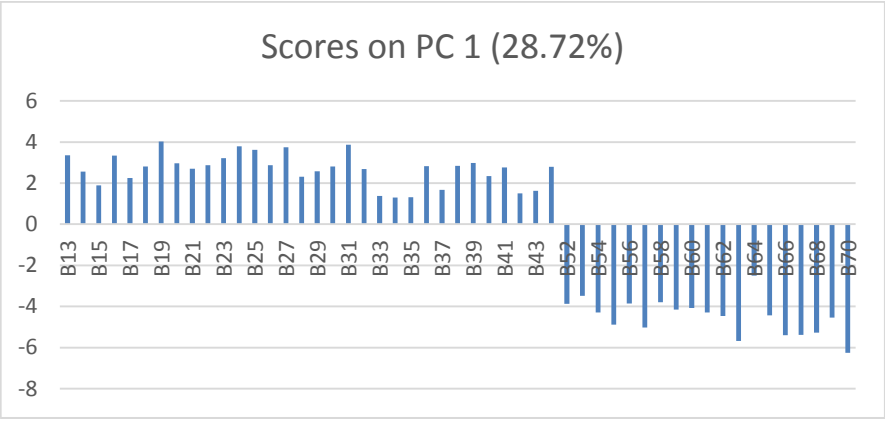

Figure 6-3: Univariate scores plot for PC1

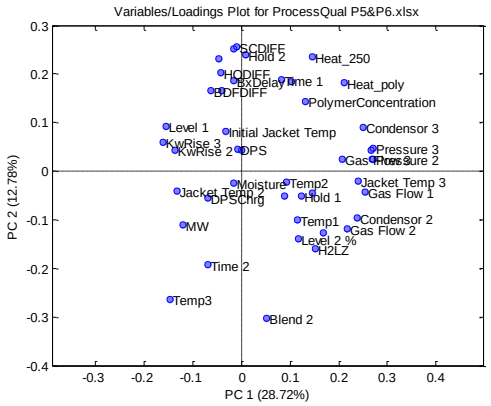

Figure 6-4: Loadings plot of P5 and P6 for quality and static process variables PC1 vs. PC2

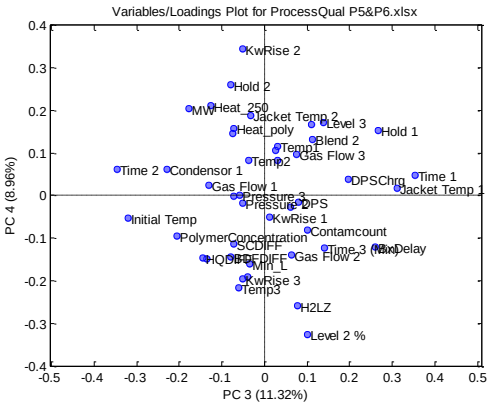

Figure 6-5: Loadings plot of P5 and P6 for quality and static process variables PC3 vs. PC4

Following this analysis B44 that was good batch was identified as an outlier in PC1 vs. PC2 scores plot. This batch was removed from the analysis and a new model was developed. Bad batches on both the reactors were projected on the new PCA model. Except for B64 all the bad batches clustered along with the respective reactor with all the batches lying within 95% confidence bounds as seen in Figure 6-6.

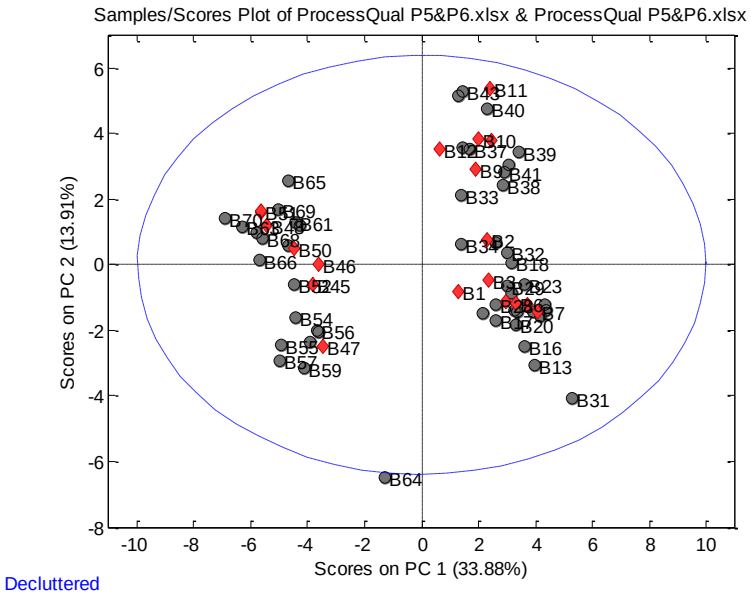

Figure 6-6: Projecting bad batches on PCA model

6.3.3 Conclusions

The results indicated that both the reactors behave differently and therefore it would be ideal to model each reactor separately. On projecting the bad batches on the calibrated models, none of the crystallised batches separated out as outliers but instead clustered along with the good batches with the respective reactors. It can therefore be concluded that analysing the data by merging the quality and static process data was not able to identify the source of crystallisation or identify any correlation between variables. Modelling the quality data and process data separately for each reactor was therefore considered as more viable option to analyse the root cause of crystallisation.

6.4 Analysis of Quality Data

This section of the thesis discusses the Materials and Methods used to analyse Quality data on both of the Polymer reactors. Section 6.4.1.1 discusses the results obtained for Poly 5 and section 6.4.1.2 discusses the results for Poly 6.

A total of 19 quality variables listed in Table 6-2 were used for this analysis. Most of these variables were quality measurements acquired in the lab at the end of the process except for Hold 1, Hold 2, Heat_poly and Heat_250 which are time measuring variables collected for the purpose of Overall Equipment Effectiveness (OEE). OEE is a best practice tool that is used within industry that enables monitoring and improving the efficiency of manufacturing processes. OEE is calculated by multiplying three data sources from machines which are Availability, Performance and Quality. OEE variables have been included in the analysis as they were eventually used to monitor the quality of the product. The central objective of doing a PCA on the quality data was to look for relationships within the quality variables and investigate if any of these variables might have a direct correlation to crystallised batches.

A correlation map or heat map was generated using The Unscrambler software V.10.3 which gives an overview of existing relationship between quality variables in the raw data. Figure 6-7 indicates a positive correlation between Hold1, Hold 2 and Heat_250 and a negative correlation between molecular weight & BDFDiff. Also the high level alarm (H2LZ) was negatively correlated to the time variables such as Hold1, Hold 2 and Heat_250. These correlations will be further investigated to identify the effect these variables have on the product quality.

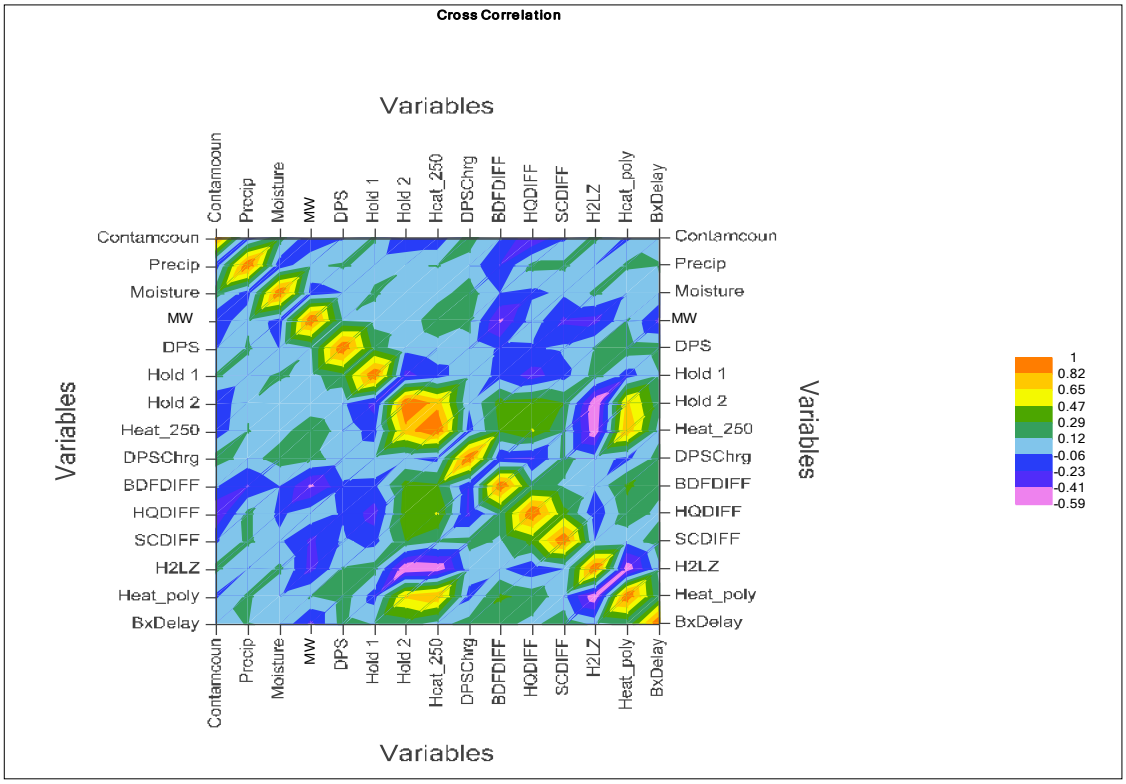

Figure 6-7: Correlation map quality variable

6.4.1 PCA on Quality data

6.4.1.1 Poly 6

PCA was performed initially on the good batches and bad batches were later projected on the PCA model developed for all of the 15 quality variables on Poly 6. The data was primarily auto-scaled to zero mean unit variance. By performing a PCA on quality data it was expected to identify if crystallised batches had correlation with measured quality variables. This would be evident in the scores plot if the bad batches clustered out separately from the good batches or if they are detected as outliers in the monitoring statistics. Table 6.5 details the variance captured by each PC and also the cumulative variance for retained PCs. The model explains 71.88% of the variance contained in the data set.

Table 6-5: PCA model for Quality data on P6

PC Number	Eigenvalue of Cov(X)	% variance captured by this PC	% cumulative variance
1	5.93	31.19	31.19
2	2.70	14.23	45.42
3	1.99	10.47	55.89
4	1.63	8.57	64.46
5	1.41	7.42	71.88

From the bivariate scores plot for PC1 vs. PC2 in Figure 6-8 B30 was identified as an outlier. The remaining good batches lie within the 95% confidence bounds. However further investigating the model by analysing the influence plot in Figure 6-9 it was observed that along with Batch 30 being an outlier in Hotelling's T^2, Batch 19 and Batch 44 were outliers in Q-residual.

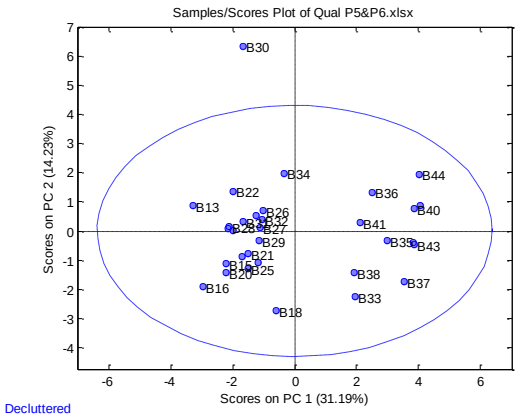

Figure 6-8: PC1 vs. PC2 P6 Quality data

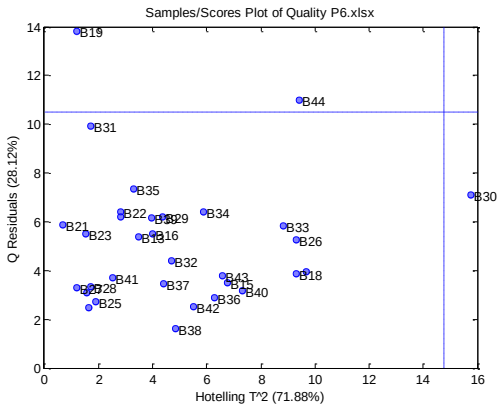

Figure 6-9: Influence plot PCA model for P6

To further look into the variables that may be responsible for the outlying batches the Q-residual contributions for Batch 19 and Batch 44 and the Hotelling's T^2 contributions for Batch 30 were investigated. Figure 6-10 indicates that the contamcount variable in Batch 19 has the highest contribution to the Q-residuals making it an outlier. Comparing the actual value of this variable with the remaining batches as seen in Figure 6-11 it further verifies the high contamcount contribution to the outlying behaviour of Batch 19.

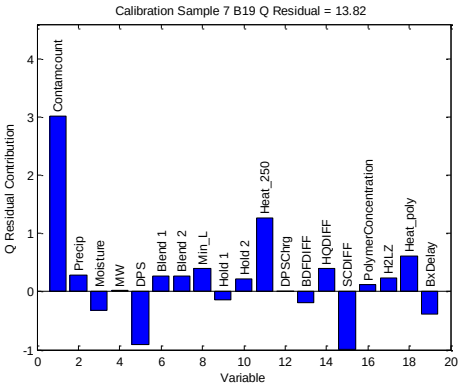

Figure 6-10: Batch 19 variable contribution to Q-residual contribution

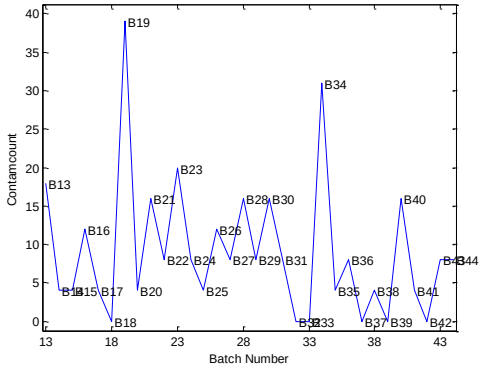

Figure 6-11: Contamcount comparison of all the batches

Examining Batch 30 which indicated a high value in the Hotelling's T^2 contribution it can be seen that MW and BxDelay have a higher contribution to this batch [Figure 6-12]. Comparing the actual values of MW and BxDelay with the remaining batches in Figure 6-13 and Figure 6-14 it can be seen that MW for Batch 30 is the lowest while the BxDelay is the highest as compared to the remaining good batches.

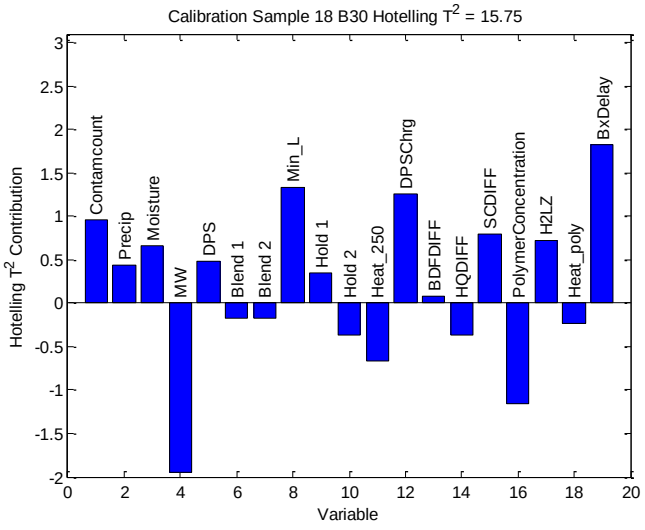

Figure 6-12: Batch 30 Hotelling's T^2 contribution

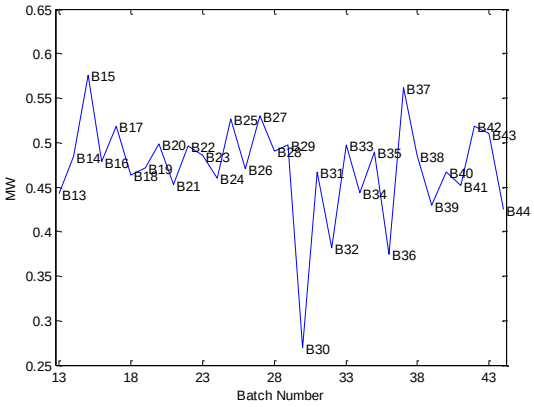

Figure 6-13: MW comparison for all the batches

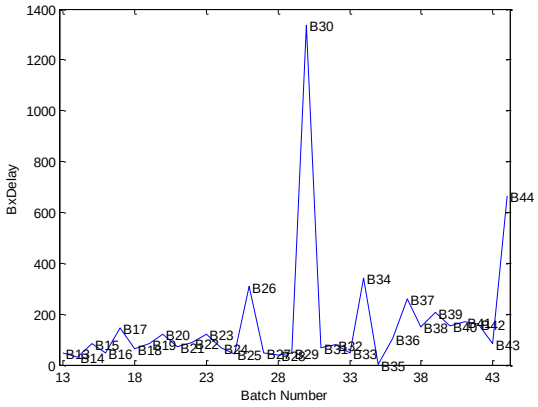

Figure 6-14: BxDelay comparison of all batches

The influence plot in Figure 6-9 also indicated that B44 does have higher Q-residual statistic compared to the other batches. Analysing the Q-residual contribution for B44 in Figure 6-15, Hold1, DPS, Min_L and BxDelay were high contributors to the Q-residual statistic with Hold 1 having the highest Q-residual value. Comparing the Hold 1 values for Batch 44 with the remaining batches that were used to develop the calibration model it can be seen that the actual value of Hold 1 for B44 is extremely high (Figure 6-16).

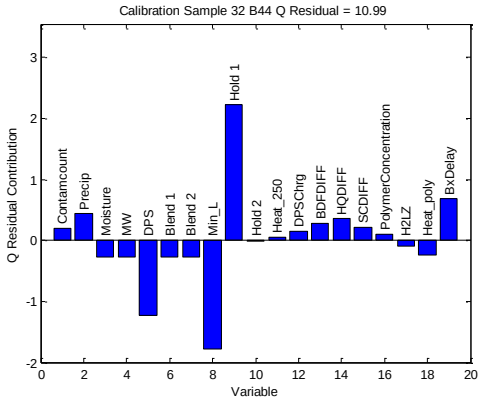

Figure 6-15: Q-residual contribution for B44

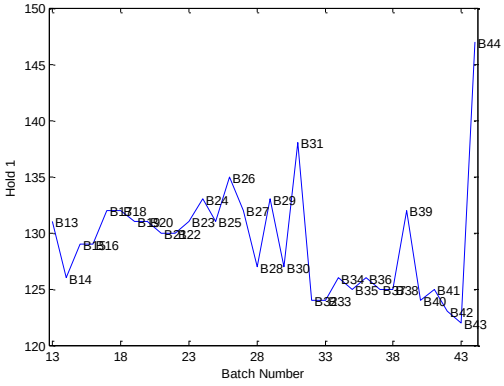

Figure 6-16: Comparing Hold 1 values

After examining the outlying good batches (B19, B30 and B44) it was decided to remove these batches from the calibration data and re-develop the model.

Figure 6-17 is the scores plot for PC1 and PC2 and Figure 6-18 is the influence plot for PCA model redeveloped without B19, B30 and B44. The redeveloped model is a good calibrated model for the selected batches with all the batches lying within the 95% confidence interval. The crystallised/bad batches were then projected on this calibrated model. The scores plot for PC1 against PC2 in Figure 6-19 indicated that most of bad batches seem to fit the calibrated

model well. However B9 and B12 have been detected as outliers in the influence plot with high Q-residual statistic. Interrogating the contribution plots for these two bad batches a high value for BxDelay for B9 (Figure 6-21) and high value of moisture content was observed for B12 (Figure 6-22). The remaining bad batches remained within control with regards to quality variables.

Also another important observation in the scores plot (Figure 6-17) indicated that all the batches were separating in two distinguishable groups. The first group roughly rests in the 2nd and 3rd quadrant while the second group in the 1st and 4th quadrant. Tracing the batches back to the raw data indicated that the reason for the distinct separation in the batches was determined by the time of the year these batches were processed. The first group on the left hand side of the scores plot are the batches that were processed in the month of October/November 2013 while the ones on the right hand side were processed in the month of January 2014. The projected bad batches (Figure 6-19) which lie within the 95% confidence interval also distinguish themselves into the respective groups depending on when these batches were processed.

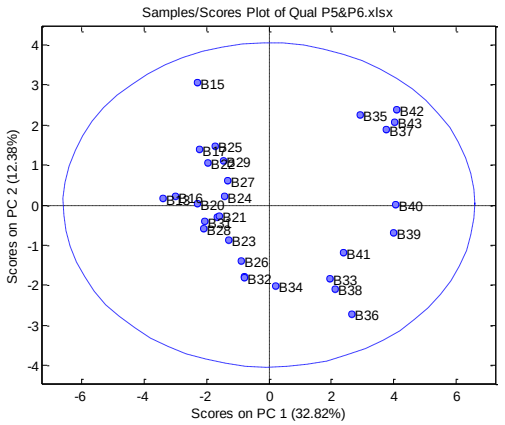

Figure 6-17: PCA model without B19, 30 and 44

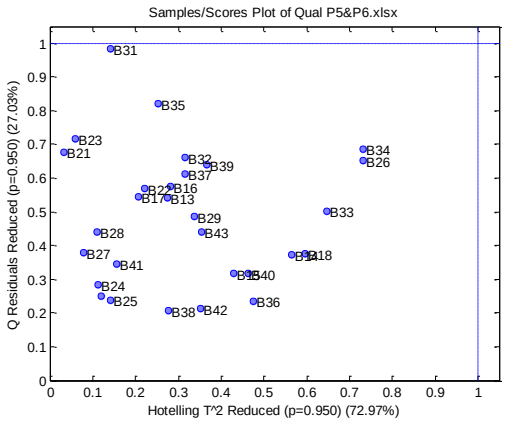

Figure 6-18: Influence plot for PCA model without B19, 30 and 44

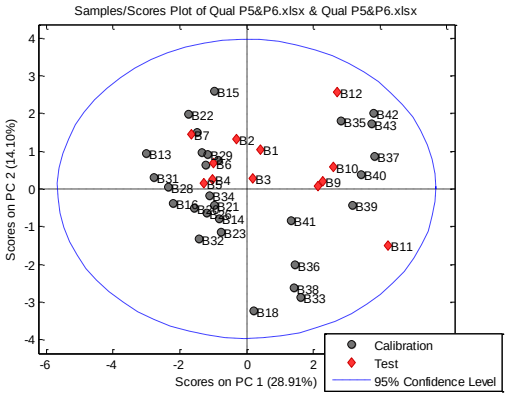

Figure 6-19: Projected bad batches on PCA model

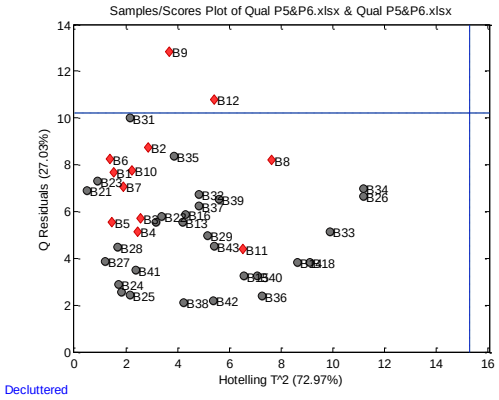

Figure 6-20: Influence plot with the projected bad batches

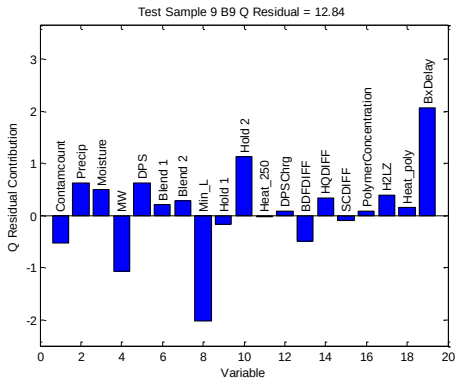

Figure 6-21: B9 Q-residual contribution

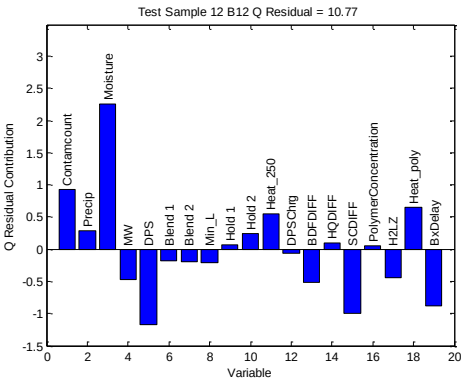

Figure 6-22: B12 Q-residual contribution

6.4.1.2 Poly 5

PCA was performed for randomly selected good batches on Poly 5. A PCA model with 26 good batches that included 19 quality variables was developed retaining 5 PCs explaining a total of 72.56% variance in the data set as seen in Table 6-6.

Table 6-6: PCA model for Quality data on Poly 5

PC Number	Eigenvalue of Cov(X)	% variance captured by this PC	% cumulative variance
1	4.91	25.82	25.82
2	3.18	16.76	42.58
3	2.36	16.76	55
4	1.84	9.67	64.67
5	1.50	7.89	72.56

The scores plot for the first 2 PCs and the influence plot indicated that there are no outliers in the calibration set. After developing the calibration model 9 crystallised batches were projected on it. The scores plot for PC1 vs. PC2 (Figure 6-23) does not suggest any bad batches as outlying or extreme with respect to the calibration model. However studying the Influence plot in Figure 6-24, B49, B47 and B50 have Q-residuals higher than the remaining bad batches and lie outside the 95% confidence bounds.

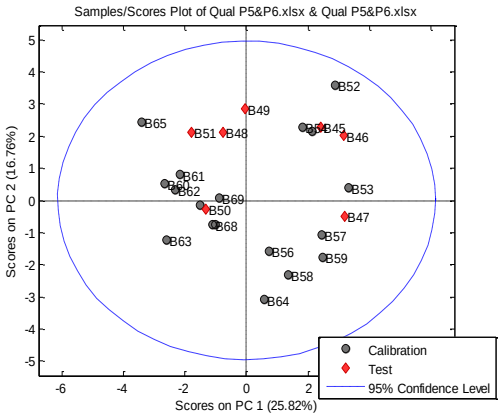

Figure 6-23: PC1 vs. PC2 for P5 PCA model

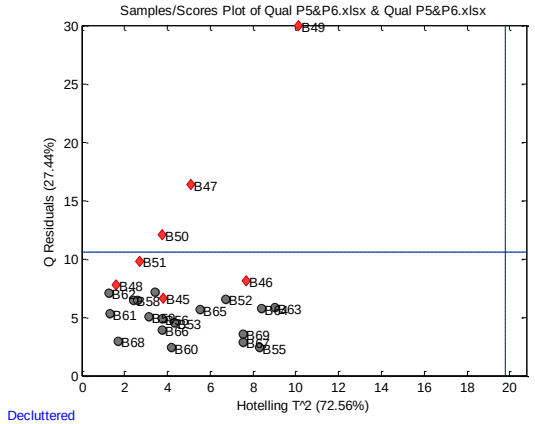

Figure 6-24: Influence plot P5

Investigating the contribution plot for B49 as seen Figure 6-25 it was noted that Heat_250 variable is the highest contributor followed by Hold 2, Contamcount and DPSCharge. On

comparing the actual values of these variables with the other bad batches it was observed that Heat_250 and Contamcount were the main contributors to the outlying batch.

B47 was an outlier due to low value of HQDiff when compared to the other batches as seen in Figure 6-30 and Figure 6-31. An unusual high value for SCDiff and H2LZ seems to be the reason for making B50 an outlier in the influence plot.

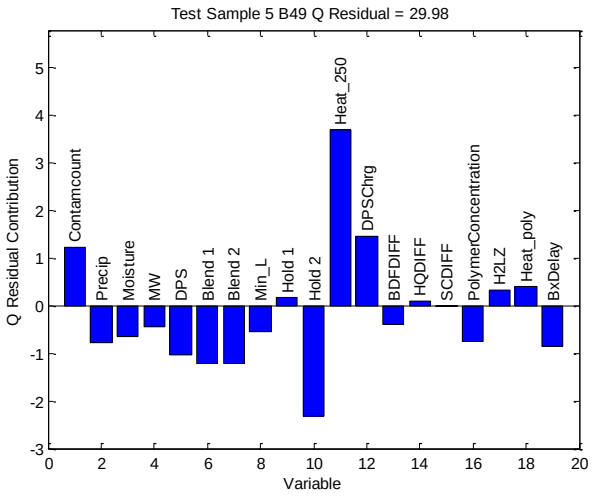

Figure 6-25: B49 Q-residual contribution plot

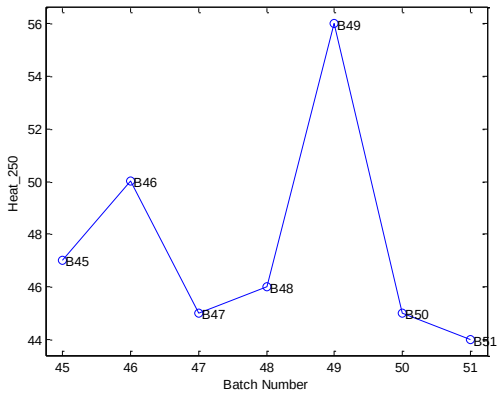

Figure 6-26: Comparing the Heat_250 for bad batches on P5

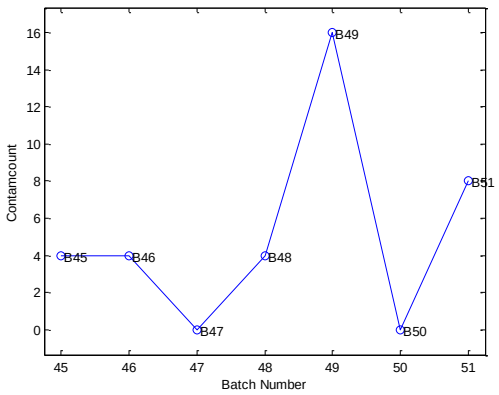

Figure 6-27: Comparing contamcount for bad batches on P5

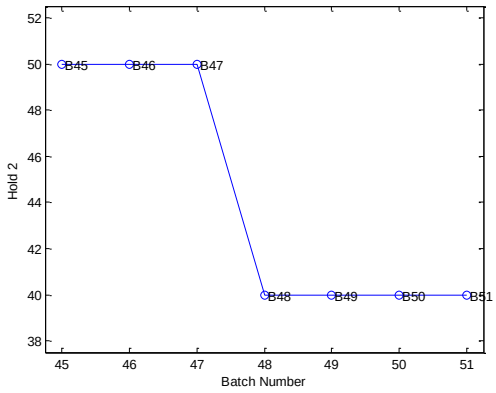

Figure 6-28 - Comparing the Hold 2 for bad batches on P5

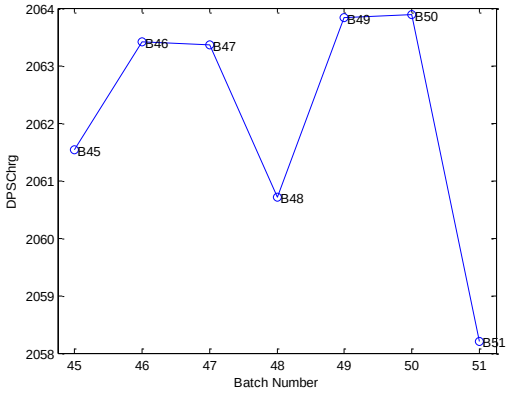

Figure 6-29 - Comparing DPSCharge for bad batches on P5

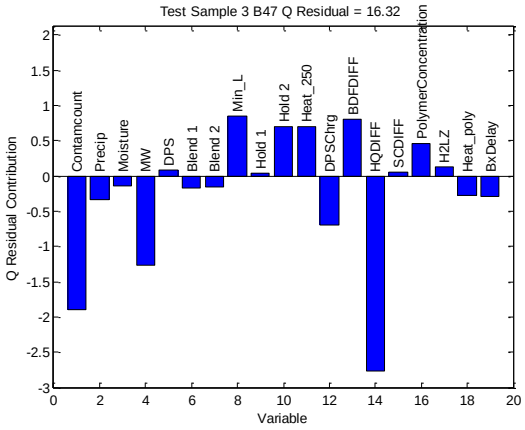

Figure 6-30: Q-residual contribution plot for B47

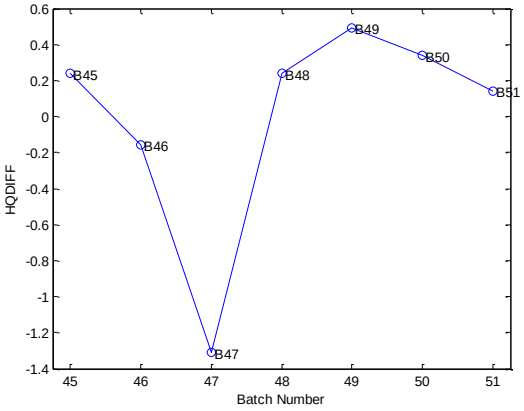

Figure 6-31: Comparing the HQDiff for bad batches on P5

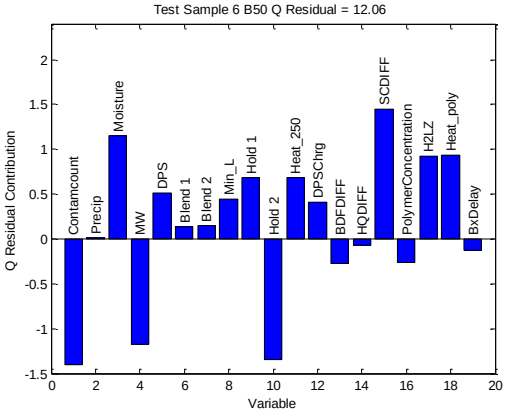

Figure 6-32: B50 Q-residual contribution

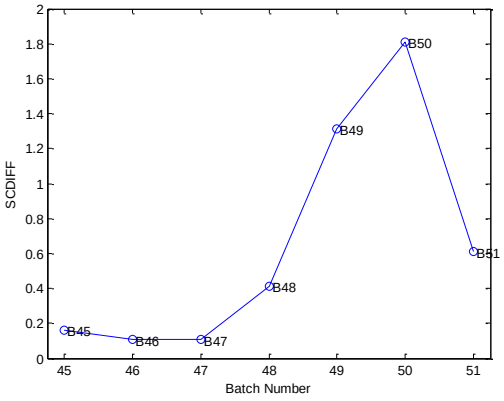

Figure 6-33: Comparing SCDiff for bad batches on P5

6.4.2 Conclusion

As mentioned previously in the introduction the customer has not been able to identify a quality variable that would indicate a direct correlation to a crystallised batch. They have contemplated a possible direct correlation of the contamcount variable with the crystallised batches. This is because as crystallisation occurs in the reactors it scours off the inside of the reactor which possibly increases the amount black spec in the reactors. The amount of black spec in the final product is measured in the contamcount variable and customer speculates that this variable might have a direct correlation to the crystallised batch. One of the purposes of carrying out multivariate analysis was to identify one or more variables that would establish a direct relationship to crystallised batches.

The quality data analysis on both the reactors has explored the option of establishing this relationship between contamcount and crystallised batch. An unusually high contamcount was observed on B19, a good batch used as test set to develop the calibration model. However none of the other bad batches on Poly 6 exhibited an out of specification value for the contamcount variable. B49 which is a bad batch on Poly 5 did showcase a high contribution for the contamcount variable and could be a possible cause of making it an outlier in the Influence plot. However this observation was only seen in B49 and was not consistent with any of the other crystallised batches. The two other outlying bad batches on Poly 6 were out of limit due to high moisture content and a longer batch delay. On the other hand variables such as HQDiff, SCDiff and H2LZ (High level alarm at Hold 2) were responsible for the outlying bad batches on Poly 5. Looking at the results from the MPCA analysis it can be concluded that none of quality variables used in the investigation were able to identify the root cause of crystallisation.

6.5 Polymerisation Reactor 6 - Process Data Analysis

Following the analysis of quality data it was decided to carry on investigating the problem using only the process data. Initially for the process data analysis the model was developed only for Poly 6 since it consisted of greater number of crystallised batches than Poly 5. The results obtained were presented to the customer following which additional process data was provided to further carry out the investigation discussed in Section 6.7.

6.5.1 Materials and Methods

The same set of 36 batches used in the quality data analysis that included 12 bad batches and 24 good batches on Poly 6 were selected for the current investigation. The data was unfolded from three dimensions to two dimensions such that the batch direction would be preserved and allow batch to batch comparison.

6.5.1.1 Pre-processing

Prior to unfolding the bad batches they were all plotted together to examine if they are time aligned and if the kinks followed a similar pattern. As can be seen in Figure 6-34 all the batches seem to start and end at the same time. This is because the data provided by the customer had same number of total time points for every batch which was averaged over six seconds for the entire batch. Also the pattern of the kinks seems to be following a similar trend but occurring at different time points as well as at varying temperatures. As seen in Figure 6-35 Batch 12 kinks around time points 2555 which is at a temperature of 241°C and Batch 4 kinks at 225 ° C which is at time point 2273

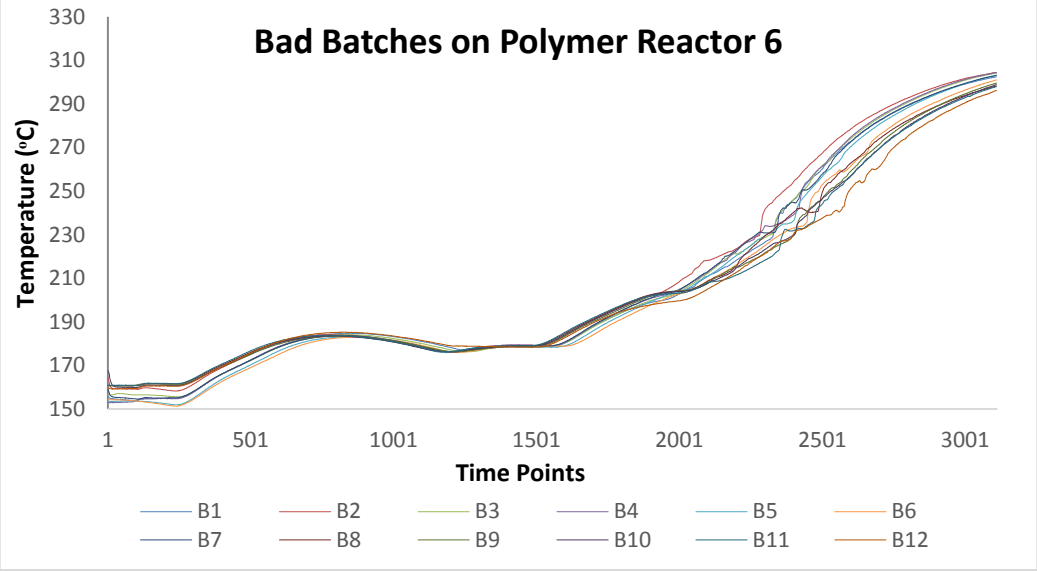

Figure 6-34: Heat up profile of bad batches on P6

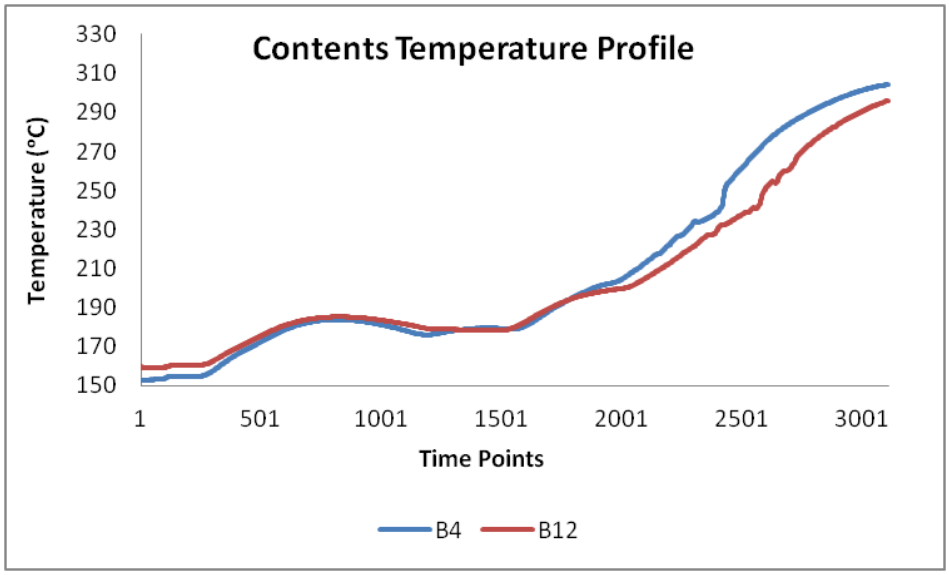

Figure 6-35: Contents temperature profile for B4 and B12

To perform MPCA using batch-wise unfolding method a total of 1081 time points were selected for each batch to ensure equal length of batches. The data was selected so as to incorporate the kink in the contents temperature profile. Since the position of kink varied between batches as seen previously in Figure 6-35 these data points were selected after the hold point 2 which was +/- 1700 to the endcap temperature time point of +/-2800 (Figure 6-36).

The data was unfolded in Matlab version R2012a by applying reshape and concatenate commands. Once the data was unfolded (for the 36 batches) it was pre-treated by auto scaling the matrix to mean zero unit variance. Auto-scaling the matrix removed the mean trajectory in the data set.

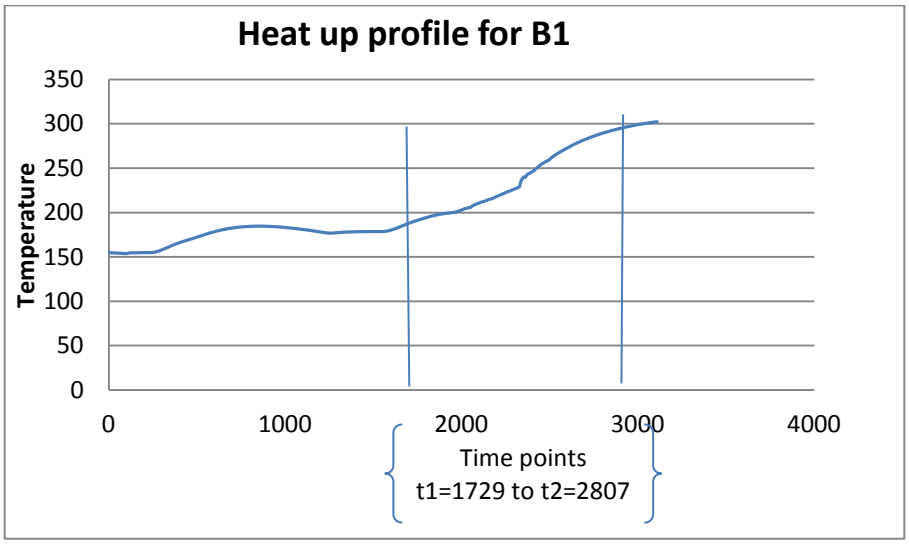

Figure 6-36: Time point selection for MPCA

6.5.2 Results and Analysis

The unfolded data (36 x 9729) for 1081 time points and 9 variables was auto-scaled followed by cross validation using the random subsets with 10 splits and 5 iterations. PCA model was developed retaining 6 principal components explaining 71.10% cumulative variance. B30 and B44 were identified as outliers in the Hotelling's T^2 while B31 and B39 were high in residuals in the influence plot seen in Figure 6-37.

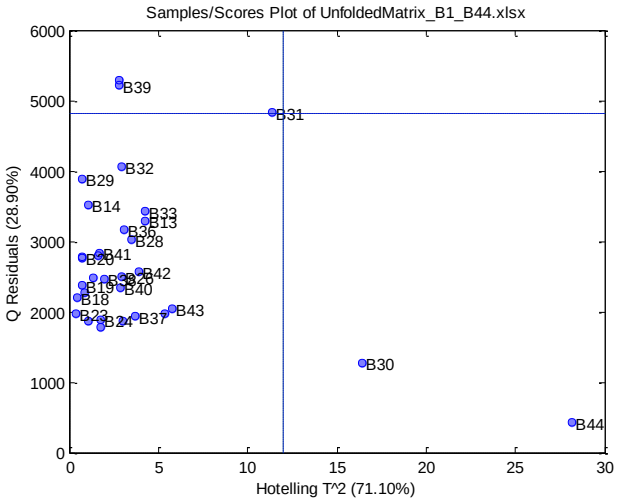

Figure 6-37: PCA model Process Data

The PCA model was re-developed without the outlying batches and retaining 6 PCs. The bad batches were then projected on this model. Table 6-7 shows the final PCA model developed for the process data on Poly 6.

Table 6-7: PCA model for Process data on Poly 6

PC Number	Eigenvalue of Cov(X)	% variance captured by this PC	% Cumulative Variance
1	3.86 e003	39.72	39.72
2	1.54 e003	15.85	55.57
3	7.89e002	8.11	63.68
4	5.17e002	5.31	69
5	4.67e002	4.80	73.80
6	3.81e002	3.92	77.72

Whilst the scores plot illustrates the relationship between different batches only one batch is detected as an outlier in the PC1 vs. PC2 scores plot shown in Figure 6-38. The influence plot however indicates three bad batches which are B6, B9 and B12 as extreme outliers while B5 and B8 are just outside the 95% confidence bounds for Q-residuals.

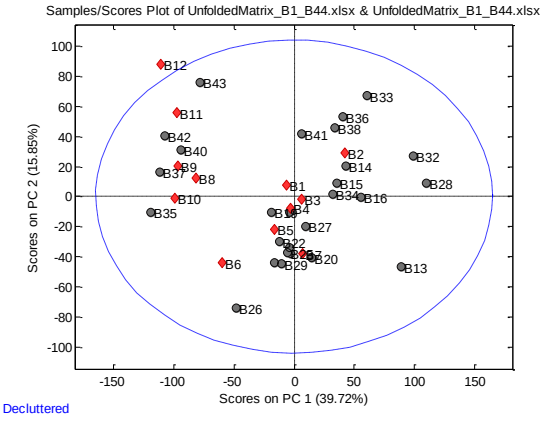

Figure 6-38: PC1 vs. PC2 PCA model for process data on P6

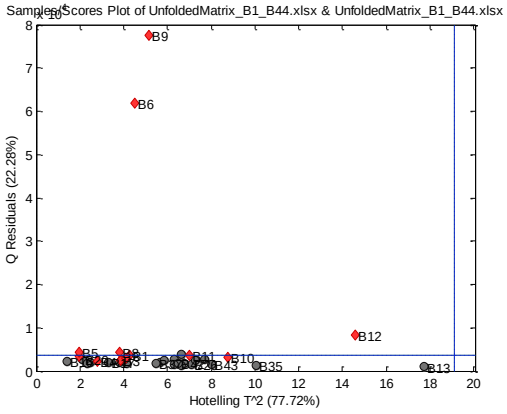

Figure 6-39: Influence plot for process data model

Investigation of the Q-residual contribution plot for B6 and B9 indicated that the agitator speed and KwRise were the problem variables for the outlying batches. The raw data showed that Agitator speed and KwRise for both these batches was lower than all the other batches at time point 2208 as captured in Figure 6-40 and Figure 6-41. B12 on the other hand is an outlier due to high level in KwRise as seen in Figure 6-42. However unlike B6 and B9 the KwRise for B12 is much higher than for the other batches at a time point 39 which is after hold 2.

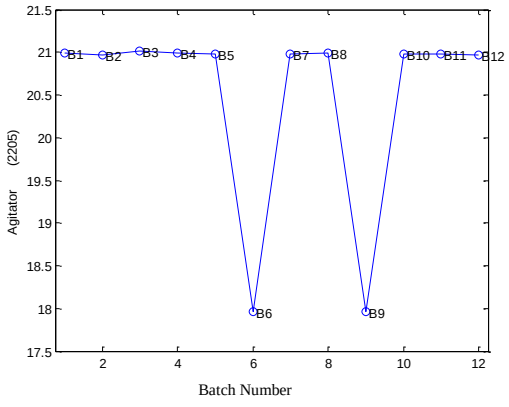

Figure 6-40: Comparing agitator rise at time point 2205

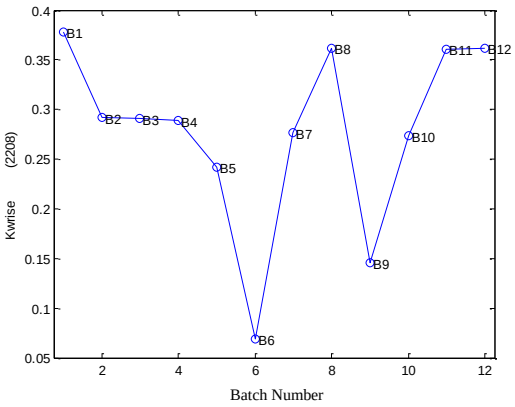

Figure 6-41: Comparing KwRise at time point 2208

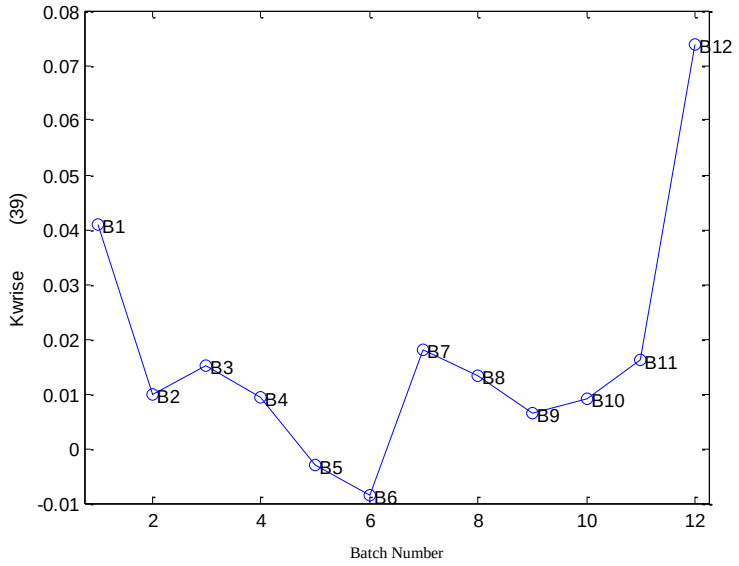

Figure 6-42: Comparing KwRise at time point 39

6.5.3 Conclusion

The process data was examined to see if a distinct deviation in the process variables for the bad batches is observed as a batch progressed. This would then enable to establish a link with the crystallised batches and identify variables leading to the unusual behaviour in the contents temperature in the reactor. In the analysis it was observed that only three out of the twelve bad batches on Poly 6 showcased an unusual behaviour as compared to the good batches which was clearly seen in Figure 6-39. Also B9 and B12 were outliers in the process data analysis as well the quality data analysis discussed previously in Section 6.4. High level of moisture content for B12 in the quality data analysis and high level of KwRise in the process data analysis was the root cause of this outlying batch. In the case of B9 a high value of BxDelay and lower value of KwRise were responsible to make it an outlier.

From this analysis it was concluded that there was no clear link between crystallised batches and any particular process variables that might account for the unusual behaviour in the process. It was therefore decided to analyse additional data with more number of crystallised batches investigated further in Section 6.7.

6.6 Particle size analysis

The particle size of Sodium Carbonate (Na_2CO_3) is crucial for the rate of reaction in the polymerisation reactors. However with much smaller quantities of the Na_2CO_3 required for the polymerisation reaction the customer is unable to obtain the exact particle size required for their process and have to accept the raw material size as provided by the suppliers. To analyse if particle size might be having an impact on crystallisation the customer monitored the particle size of raw material on reactors 1, 2, 3, and 4 for a limited period of time. These measurements for raw material particle size were carried out using three different sizes. This section of the report analyses the effect of particle size of Na_2CO_3 on the quality data.

6.6.1 Materials and Methods

The core objective of doing PCA with quality and particle size data was to characterise any correlation between the crystallised batches and particle size. The particle size data was provided for Poly 1, Poly 2, Poly 3 and Poly 4 from 30th January 2014 to 9th February 2014. This is the time period when the experimental trials to measure to particle size were carried out on site. The particle size data measured for the experiments was 1) total charge weight, particles size 2) >106, 3) between 106-53 and 4) <63. In addition a total of 14 quality variables were also included in the analysis. The genealogy of the batches used in the analysis can be seen in Table 6-8. The good batches were randomly selected around a similar time to the bad batches.

Table 6-8: Genealogy of Batches used Particle size analysis

Batch Number	Reactor	Type
B1-B4	Poly 1	Bad
B5	Poly 2	Bad
B6-B8	Poly 3	Bad
B7-B12	Poly 4	Bad
B13-B21	Poly 1	Good
B22-B30	Poly 2	Good
B31-B42	Poly 3	Good
B44-B49	Poly 4	Good

6.6.2 Results and Analysis

A PCA model was developed for good batches, and the bad batches were projected on the model. The model was developed retaining 5 PCs explaining total of 72.01% cumulative variance as seen in Table 6-9.

Table 6-9: PCA model for Particle Size Analysis

PC Number	Eigenvalue of Cov(X)	% variance captured by this PC	% cumulative variance
1	7.16	26.53	26.53
2	4.85	17.98	44.51
3	3.03	11.24	55.75
4	2.34	8.67	64.41
5	2.05	7.60	72.01

Figure 6-43 that looks at scores plot indicates that the batches on Poly 4 have clustered out on the right hand side of the plot as compared to the batches from the remaining reactors. This observation was consistent for both the calibrated and the projected data.

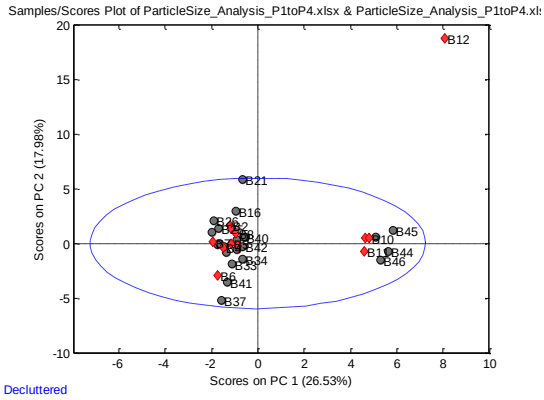

Figure 6-43: Scores plot PC1 vs. PC2: Particle size data

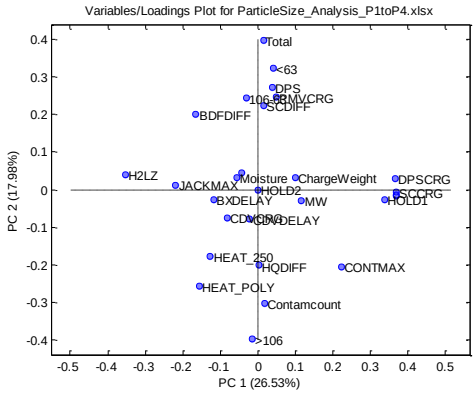

Figure 6-44: Loadings plot PC1 vs. PC2: Particle size data

Investigating the loadings plot in Figure 6-44 it can be seen that DPSCharge, SCCharge and Hold 1 are positively correlated to the batches on Poly 4. Analysing the raw material data for DPSCharge (for the good and bad batches) as seen inFigure 6-45 and Figure 6-46 indicate the addition of DPSChrg is noticeably higher in the batches on Poly 4 compared to the other batches. Similar behaviour is also observed in BDFCharge, HQCharge, SCCharge and Hold1

for this reactor and this can be attributed to the fact that the size of Poly 4 is larger than the rest of the three reactors thus requiring larger amounts of raw material.

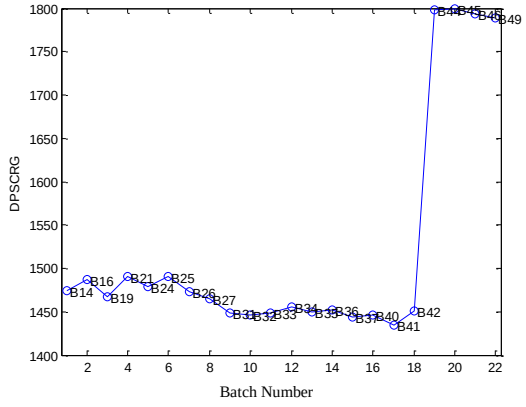

Figure 6-45: DPSCharge Good Batches

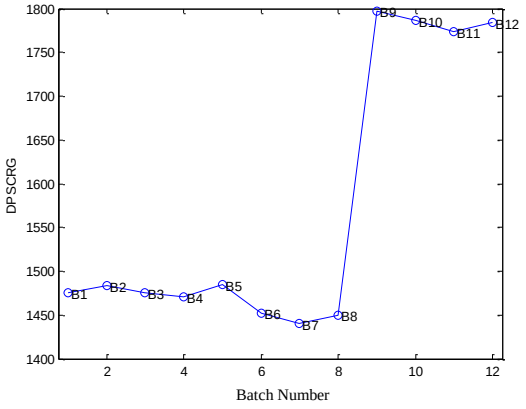

Figure 6-46: DPSCharge Bad Batches

Further investigation of the influence plot for the good and bad batches shows that Batch 12 is an extreme outlier in terms of Hotelling's T^2 and Q-residuals (Figure 6-47). The Q-residual contribution for this batch indicates unusually high PMVCharge as the reason for the batch being an outlier. Zooming in on the remaining batches in the influence plot (Figure 6-49) indicated five other batches with high Q-residual indicating the presence of a new behaviour present in the abnormal batches.

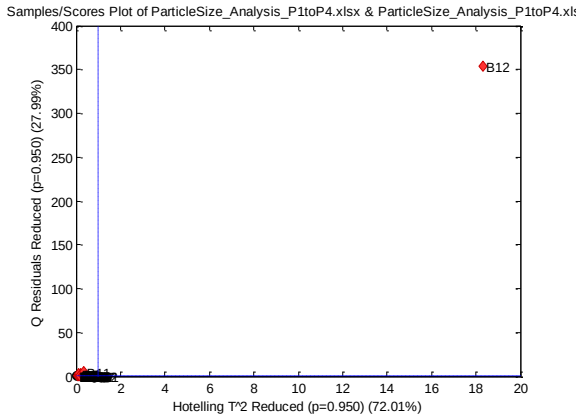

Figure 6-47: Influence plot Particle size data

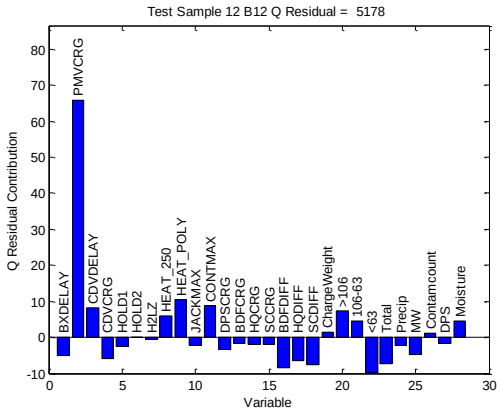

Figure 6-48: Q-residual contribution of B12 Particle size data

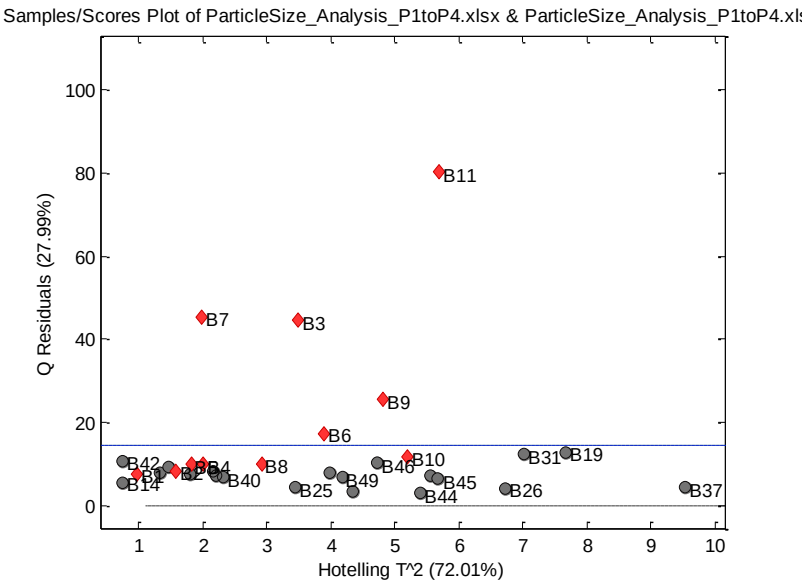

Figure 6-49: Influence plot for Particle size data zoom in

Investigating the contribution plot for these batches indicated a number of different variables responsible for the outlying batches. For example Batch 11 was an outlier due to high contribution of particle size 106-63. While for batch 7 there were two variables 1) Jackmax and 2)106-63 that were higher than the remaining batches. A high contribution in CDV charge and CDV delay made Batch 3 an outlier in the Q-residuals plot and a high contamcount made Batch 9 an outlier in the influence plot. B6 was just outside the 95% confidence bounds for the Q-residuals because of high value of Heat_250 and moisture as compared to the remaining bad and good batches. Thus it can be said that 50% of the bad batches did cluster out from the calibrated model however only couple were identified as outliers due to the particle size. Also no particular variable was identified to be responsible for the outlying batches in the Q-residuals.

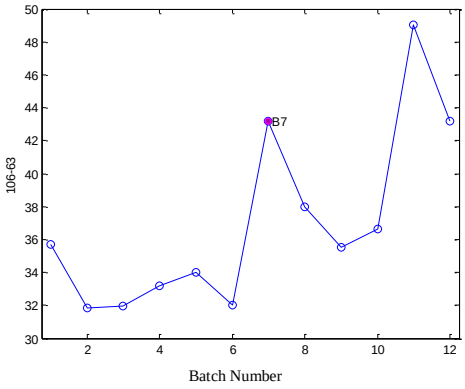

Figure 6-50: Batch 7 106-63 particle size contribution

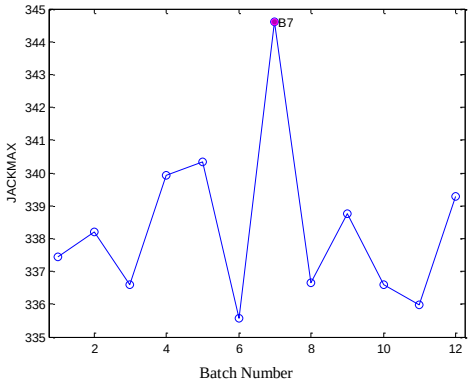

Figure 6-51: Batch 7 Jackmax contribution plot

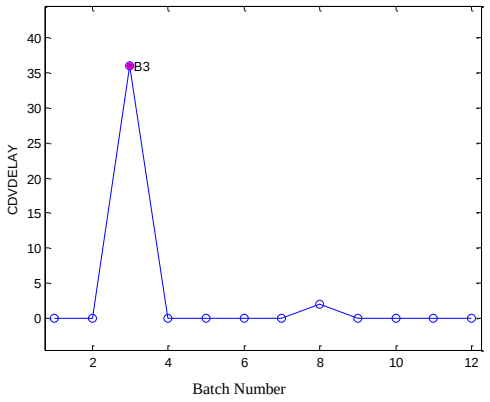

Figure 6-52: Batch 3 CDVDelay particle size analysis

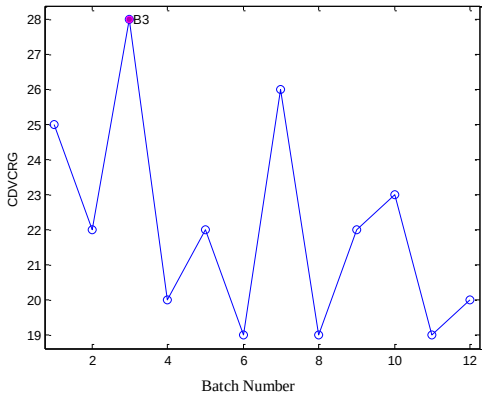

Figure 6-53: Batch 3 CDVCharge particle size analysis

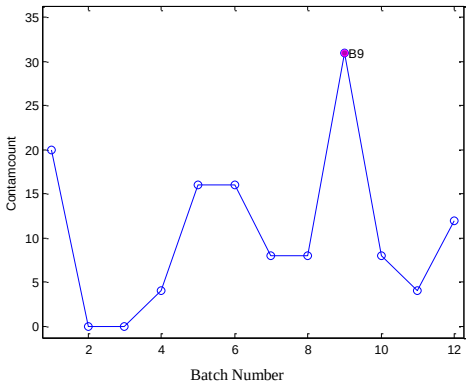

Figure 6-54: Batch 9 Contamcount Particle size analysis

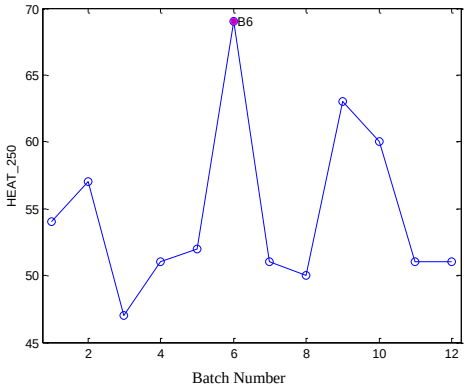

Figure 6-55: Batch 6 Heat_250 Particle size analysis

6.6.3 Conclusion

An initial analysis on the data did not indicate any significant correlation between the crystallised batches and the particle size of Na_2CO_3. Although two of the bad batches (B7 and B11) behaved differently due to particle size, this behaviour was not consistent across the majority of crystallised batches. Batches on Poly 4 clustered out from the batches on remaining reactors due to larger quantities of raw material added to this reactor. This was due to larger size of Poly 4 in comparison to the other reactors.

6.7 Polymerisation Reactor 5 – Process Data Analysis

Following an initial analysis on the Process and Quality data for reactors additional process data was provided from April 2014 to June 2014. The data was made available for the Poly 5 reactor with additional crystallised batches identified on the new data set. For the analysis currently studied in this section the first set of data provided at the beginning of the project and the new data set were merged together. Unlike process data analysis for Poly 6 discussed in Section 6.5 where data was unfolded using only the batch-wise unfolding method, this section analyses the process data for Poly 5 using batch-wise unfolding method as well as variable-wise unfolding method. Through the application of the both these unfolding methods combined with the analysis of other statistics such as Hotelling's T^2 and Residuals, it was possible to develop the fingerprinting method discussed further in Section 6.7.4.

For Poly 5 a total of 69 batches were used for the investigation that included 23 bad batches. An exploratory data analysis was carried out using the MPCA technique with the aim to observe 1) if bad batches clustered out from the good due to unexplained behaviour in the process data indicating the reason for crystallisation in the reactor and 2) if it is possible to fingerprint a bad batch using the process data.

With the crystallisation occurring after hold point 2 the customer believed that analysing the data in the process after hold point 1 through to the endcap temperature would be important. A total of 1000 time points were selected for every variable based on the contents temperature (in order to include the kink) starting from time points 1500 (which was generally after hold point 1) to around samples 2500 which is just before the polymerisation temperature as seen in Figure 6-56.

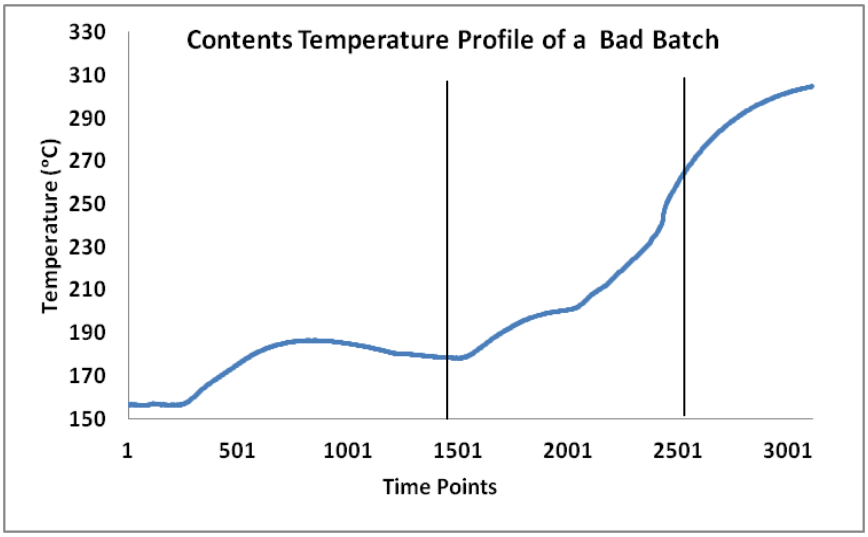

Figure 6-56: Variable Selection based on Contents Temperature Profile

6.7.1 Unfolding, Pre-treatment and Analysis

Following the process data analysis discussed in Section 6.5 it was decided to include only four variables out of nine that were initially utilised. Thus, jacket temperature (Variable 1 or V1), contents temperature (Variable 2 or V2), KwRise (Variable 3 or V3) and level (Variable 4 orV4) were the four variables included in this analysis. The results from the previous analysis indicated that Contents temperature and KwRise were crucial to the unusual behaviour of crystallised batches which justified their inclusion in the analysis. Also since Jacket temperature controls the contents temperature it was important to include this variable as well. The level control was included in the analysis since crystallisation in the reactor results in foaming which increases the level in the reactor. The data was unfolded in two ways 1) batch wise and 2) variable wise using the Matlab version R2012a software and the models developed have been discussed further. The Matlab code used to unfold the matrices has been attached in Appendix B. The batch-wise unfolding method allows batch to batch comparison while variable-wise unfolding unfolding analyses variables across the batch time.

6.7.2 Material and Methods

Table 6-10 shows a list of the batches on Polymer reactor 5 selected to perform the analysis. A total of 69 batches that included 23 bad batches and randomly selected good batches are used in this thesis.

Table 6-10: P5 Batches used for analysis

Batches	Type	Time period
B1-B7	Bad	Oct 2013-Jan 2014
B8-B23	Bad	April 2014 – June2014
B24-B40	Good	Oct 2013 – Jan 2014
B41-B69	Good	April 2014 – Jan 2014

This sub-section looks into MPCA model that has been developed for Poly 5. PCA models were developed using two different unfolding methods. The results obtained from both methods were compared along with control charts and raw process data to identify the root cause of crystallisation.

6.7.2.1 Nomikos and Macgregor Approach

An MPCA model was developed using only the good batches and retaining 6 principal components to explain around 85% variation in the data set [Table 6-11]. Figure 6-57 and Figure 6-58 looks at the influence plot for PCA model with 95% confidence bound and 99% confidence bound respectively. Both the figures indicate B35 as an extreme batch with high values for both the statistics. B30 and B50 were detected as outliers with a higher Hotelling's T^2 statistic. Although B38 and B40 were outliers in the Q-residuals they were still retained in the model to avoid over fitting the model by removing too many calibration samples. It was therefore decided to exclude B30, B35, and B50 and re-model the data. The bad batches were then validated against this model to identify how well they fitted the calibrated model. The obtained results have been further discussed in detail in this section.

Table 6-11: PCA model Poly 5 Batch wise unfolded

PC number	Eigenvalue of Cov(X)	% variance captured by this PC	% cumulative variance
1	1.84e+03	45.88	45.88
2	6.03e+02	15.07	60.95
3	4.04e+02	10.10	71.06
4	2.39e+02	5.97	77.02
5	1.67e+02	4.17	81.19
6	1.21e+02	3.02	84.21

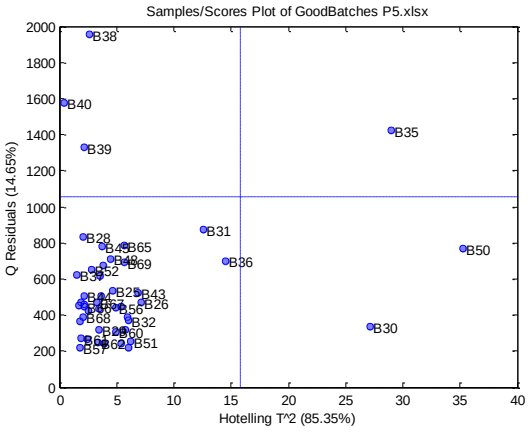

Figure 6-57: PCA model on P5 Good batches with 95% confidence bounds

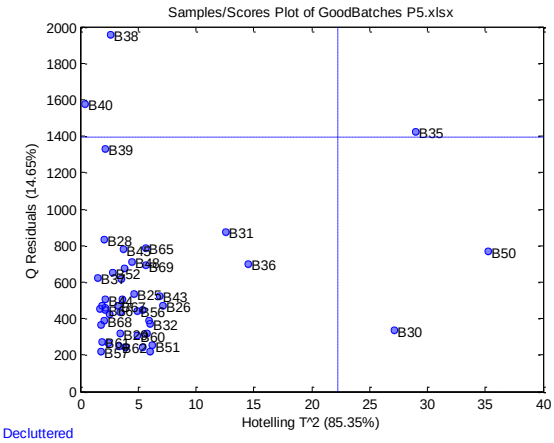

Figure 6-58: PCA model on P5 Good batches with 99% confidence bounds

The Hotelling's T^2 analysis as seen in Figure 6-59 which looks at the good and bad batches and Figure 6-60 which only looks at the bad batches distinctly identifies the abnormal crystallised batches. The Q-residuals plot in Figure 6-61 and Figure 6-62 indicates most of the bad batches as being abnormal. This plot indicates a new behaviour in the bad batches not explained in the PCA model developed using the good batches.

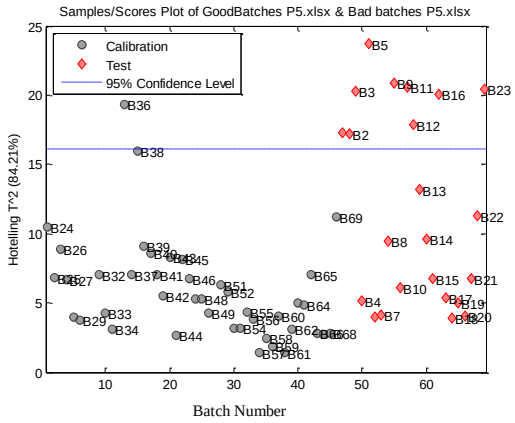

Figure 6-59: Hotelling's T2 Good and bad batches

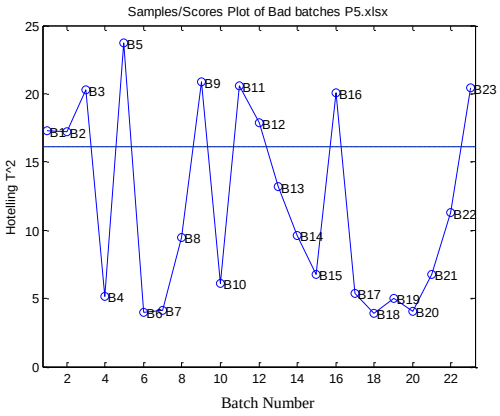

Figure 6-60: Hotelling's T^2 Bad batches on P5

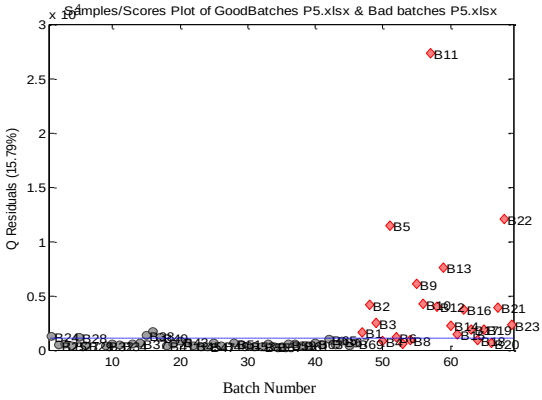

Figure 6-61: Q-Residuals Good and Bad batches

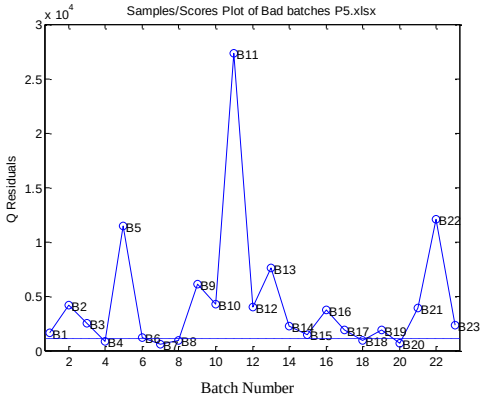

Figure 6-62: Q-residuals Bad batches on P5

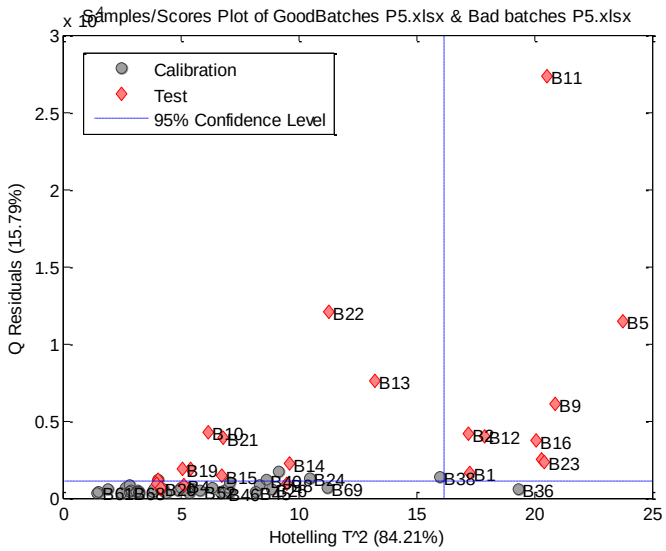

Figure 6-63: P5 Influence plot Good and Bad batches with 95% confidence bounds

Further analysing the outlying batches in the influence plot [Figure 6-63] it can be seen that B11 is extremely located as compared to the remaining bad batches. B1, B5, B9, B10, B12, B16 and B23 are also outside the 95% confidence limits in Q-residuals as well as the Hotelling's T^2 statistic. Figure 6-64 looks at the influence plot for good and bad batches but with 99% confidence bounds. The plot indicates B5 as the only abnormal bad batch. In order to take into consideration the worst case scenario influence plot with 95% confidence bounds has been used for the analysis.

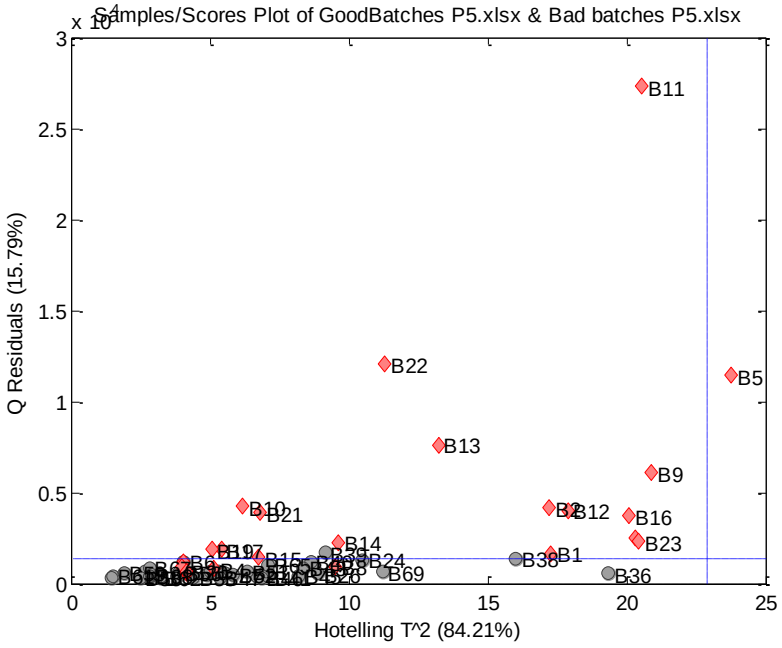

Figure 6-64: P5 Influence Plot – Good and Bad batches with 99% confidence bounds

In order to further examine the abnormal batches and identify the variables leading to the variation in these batches contribution of individual variables were studied which are discussed further in Section 6.7.3.

6.7.2.2 Variable-Wise Unfolding Approach

For the data set unfolded using the variable-wise unfolding approach, a PCA model was developed retaining 2 PCs explaining about 89% variance in the data set. It is observed that the percentage variance explained in the model by retaining just 2 PCs is significantly higher. This is because of the high correlation between the variables when the data is monitored using the variable-wise unfolding approach than when it is analysed using the batch-wise unfolding approach.

Table 6-12: PCA model Poly 5 Variable wise unfolded data

PC Number	Eigenvalue of Cov(X)	% variance captured by this PC	% cumulative variance
1	2.37	59.35	59.35
2	1.20	30.00	89.35

The model was developed using the same set of batches as used in the batch-wise unfolding approach. The scores plot for the model developed using the good batches is seen in Figure 6-65.

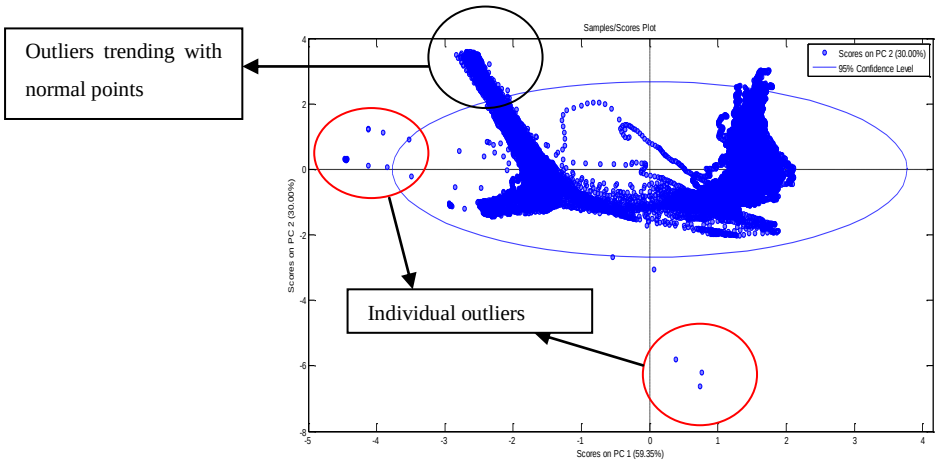

Figure 6-65- Calibration model developed using 46 batches measured across 1000 time points

The outlying points in the scores plot that followed the trend with the normal points were retained however the individual outliers from the above model were removed and model was redeveloped. The bad batches were then projected to develop a predictive model. In order to analyse the scores plot the scores matrix was re-arranged as previously explained in Chapter 2, Section 2.7.2, so as to explain the trajectory of the variables as the batch evolves. The re-arranged scores matrix for every batch has been compared with the contribution plots obtained from batch-wise unfolding analysis to identify the crystallisation problem in the next section.

6.7.3 Results and Analysis

This part of the report looks into the results obtained from unfolding the data using the batch-wise unfolding approach and the variable-wise unfolding approach followed by the analysis of various other plots to identify the crystallisation issue. Unfolding the data set using the batch-wise unfolding approach looks into batch to batch variation in the process data while the variable-wise unfolding approach tracks the behaviour of variables across time. By comparing the results from both methods followed by analysing other statistical and raw data plots the report aims to narrow down the reasons for crystallization solely based on anomaly in the process data. The influence plot in Figure 6-63 emphasises the bad batches as outliers in either Hotelling's T^2 or Q-residuals or both. Batches that are outliers in both of the statistics can be considered to be more extreme than the rest. The following analysis starts with the discussion around the most extreme batch on the influence plot (Batch 11), followed by the analysis of the remaining batches lying outside the confidence bounds.

B11 is the most extreme batch in both the monitoring statistics. In order to identify the problem variables the contribution plots for the Q-residuals were analysed as seen in Figure 6-66. The extreme behaviour in B11 can be attributed to high contribution of variables at the end of the plot. This behaviour of Q-residual contributions is different from a typical good batch as seen in Figure 6-67 and Figure 6-68. Also both the good batches have a lower total value of Q-contributions at 490.8 and 308.1 as compared to B11. Figure 6-69 and Figure 6-70 look at the contribution of variable 1 which is the jacket temperature of all the bad batches at time points 3883 and 3845 (With 1000 points analysed for the four variables retained in this analysis, the x-axis has a total of 4000 variables in the contribution plot). It can be clearly observed that B11 has the lowest value for the jacket temperature as compared to the remaining bad batches.

The variable-wise unfolding approach analysis of the batch as seen in Figure 6-71 indicates a normal trajectory within the defined statistical limits of +/- 2 and 3 standard deviation for most of the batch. A small blip in the profile at the end of batch can be attributed to the low value of jacket temperature as observed in the batch-wise analysis. The location of the drop in jacket temperature coincides with the kink in the temperature profile (Figure 6-72). No other process deviation can be observed across the time and between variables for the batch.

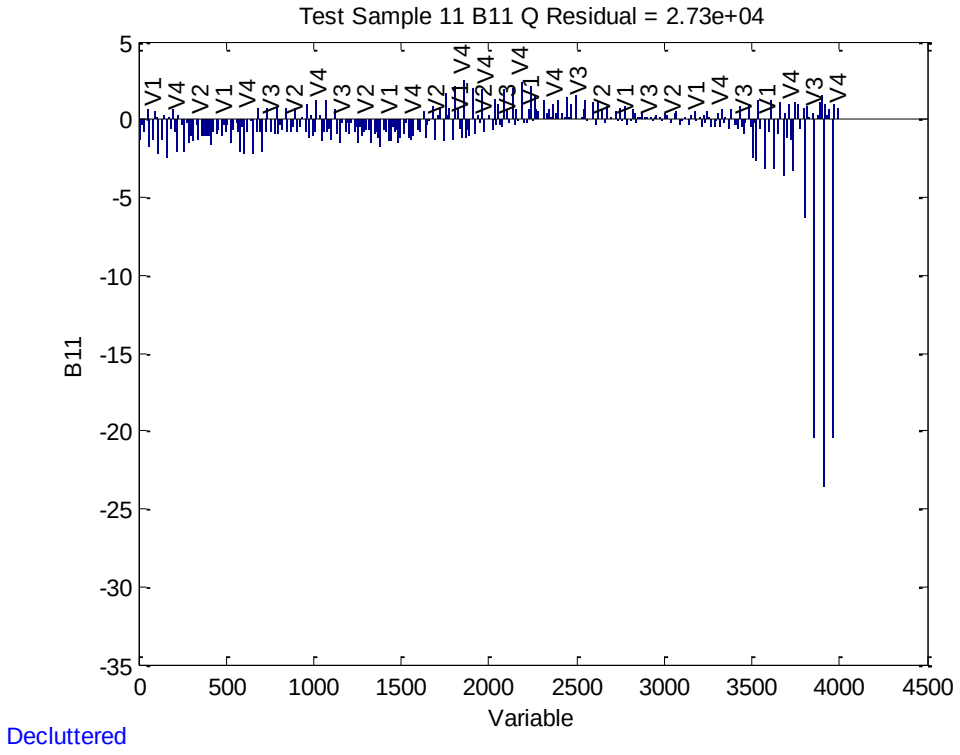

Figure 6-66: Q-residual contribution plot - Batch 11

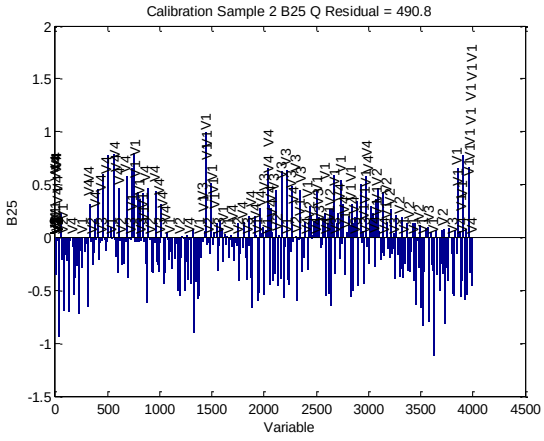

Figure 6-67: Q-residual contribution of good batch - B25

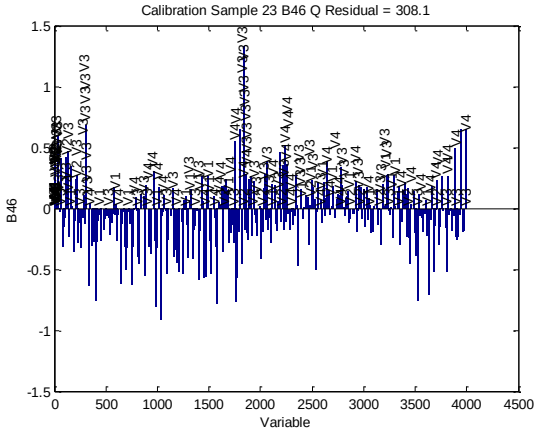

Figure 6-68: Q-residual contribution of good batch B46

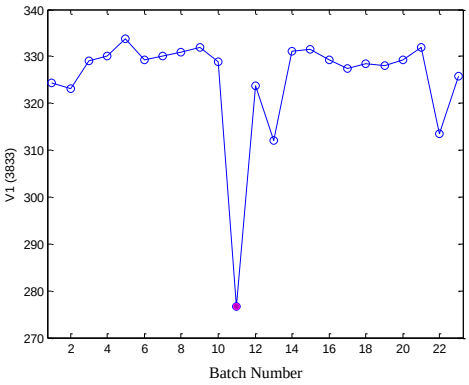

Figure 6-69: Batch 11 Q-residual contribution at time point 3833 for Variable 1

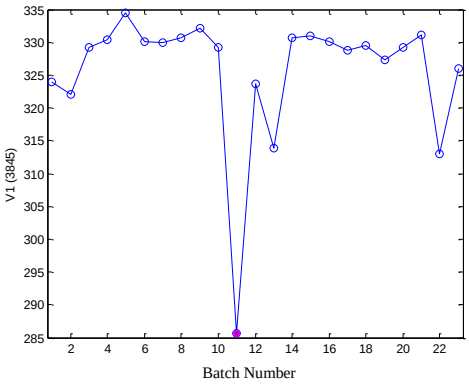

Figure 6-70: P5 B11 Q-residual contribution at time point 3845 for Variable 1

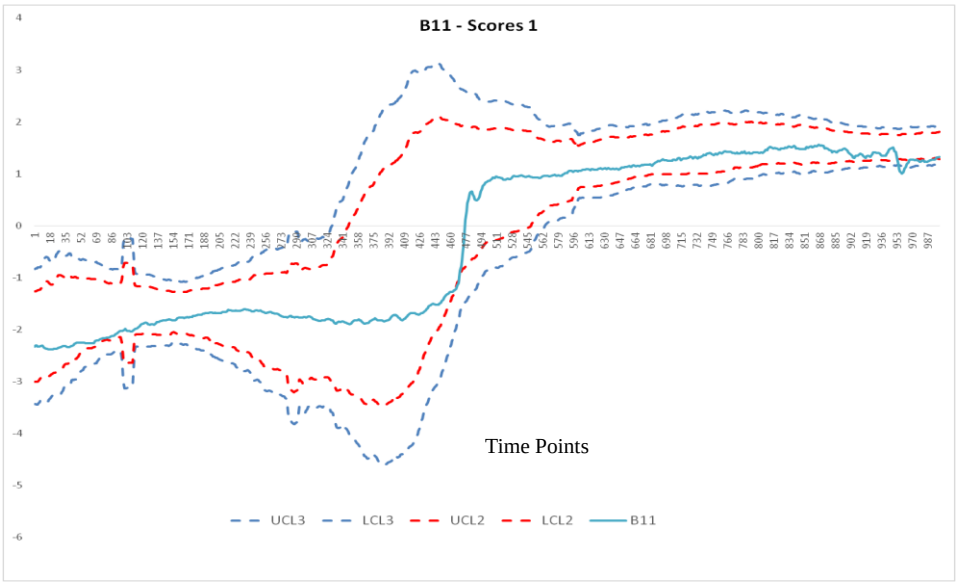

Figure 6-71: Wolds approach unfolding for Batch 11 – Scores 1
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

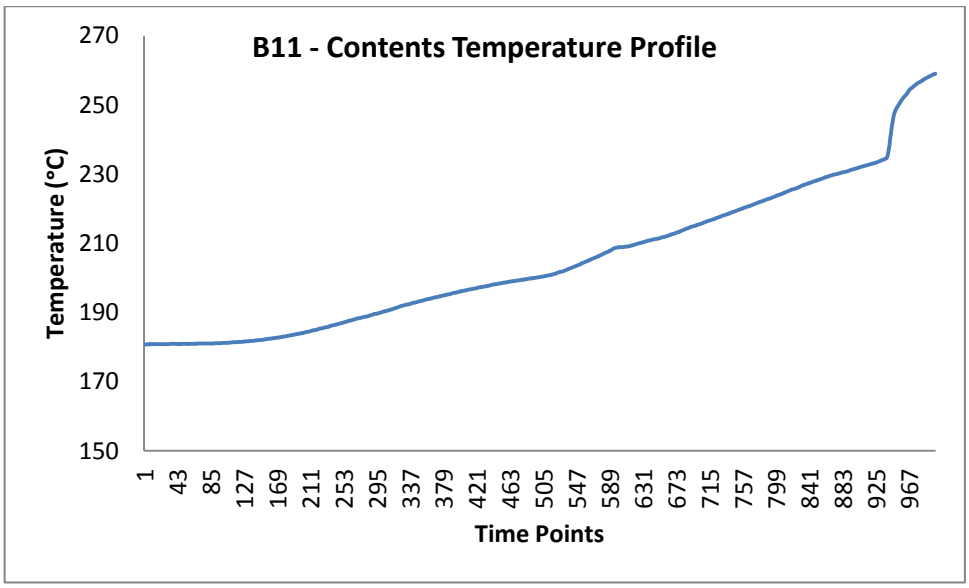

Figure 6-72: B11- V2 Contents Temperature profile

Most of the other bad batches further inspected in this thesis either indicate an unusual increase in the KwRise (V3) and/or uncontrolled level (V4) before the occurrence of kink in the temperature profile. Also for batches outside the 95% confidence limit in the influence plot a high contribution of Q-residuals with either contents temperature (V2) or KwRise (V3) or both has also been observed. An unusual high value for the Jacket Temperature (V1) was observed for B16 and B23.

B1, B2 and B3 although not very extreme in the influence plot have Hotelling's T^2 and Q-residual statistic outside the confidence bounds as seen in Figure 6-60 and Figure 6-62. (All the other relevant figures further discussed for batches B1, B2 and B3 are attached in Appendix B2, Fig 8.7 to Figure 8.26) B1 and B3 do not have a significant kink in the temperature profile either as compared to B2 and could possibly qualify as a good batches if the characterisation is solely based on the 'kink' in the profile. With a total Hotelling's T^2 value of about +/- 18, B1, B2 and B3 have a high contribution of variable 3 which is KwRise in the Hotelling's T^2 plot just after hold point 1. Also for the contribution plots for Q-residuals a high contribution of variable 3 midway through the batch was identified. Looking at the raw data scatter plot for variable 2 vs variable 3 it was seen that the KwRise for all these three crystallised batches reaches above 0.4 at around 200°C. At the same time the KwRise for a normal batch usually varies around 0.35 at the same temperature of 200°C. The scatter plot for a good batch (B42) and bad batch (B1) is seen in Figure 6-73 and Figure 6-74.

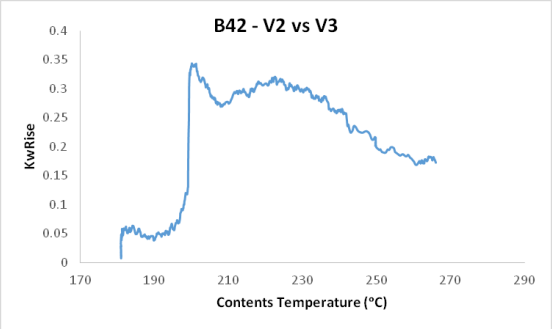

Figure 6-73: B42 Good Batch Contents Temperature vs KwRise

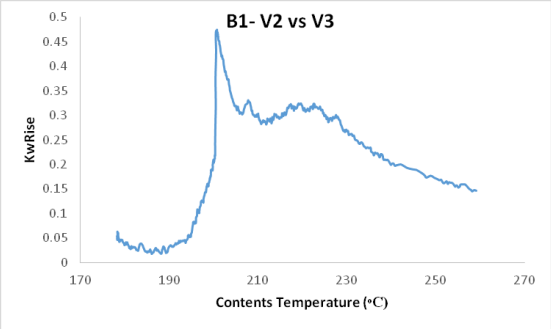

Figure 6-74: B1 Bad Batch Contents Temperature vs KwRise

It must also be noted that B35, B36 and B38 which were outlying good batches were processed around the same time as B1, B2 and B3. Thus although batches 35, 36 and 38 were not classified as bad batches they do suggest a behaviour more close to the crystallised batches.

Looking at the variable wise unfolded scores data for B35, B36 and B38 it could be clearly seen that the batches were outside the normal limits for scores 1 and scores 2 plots in Figure 6-75 and Figure 6-76.

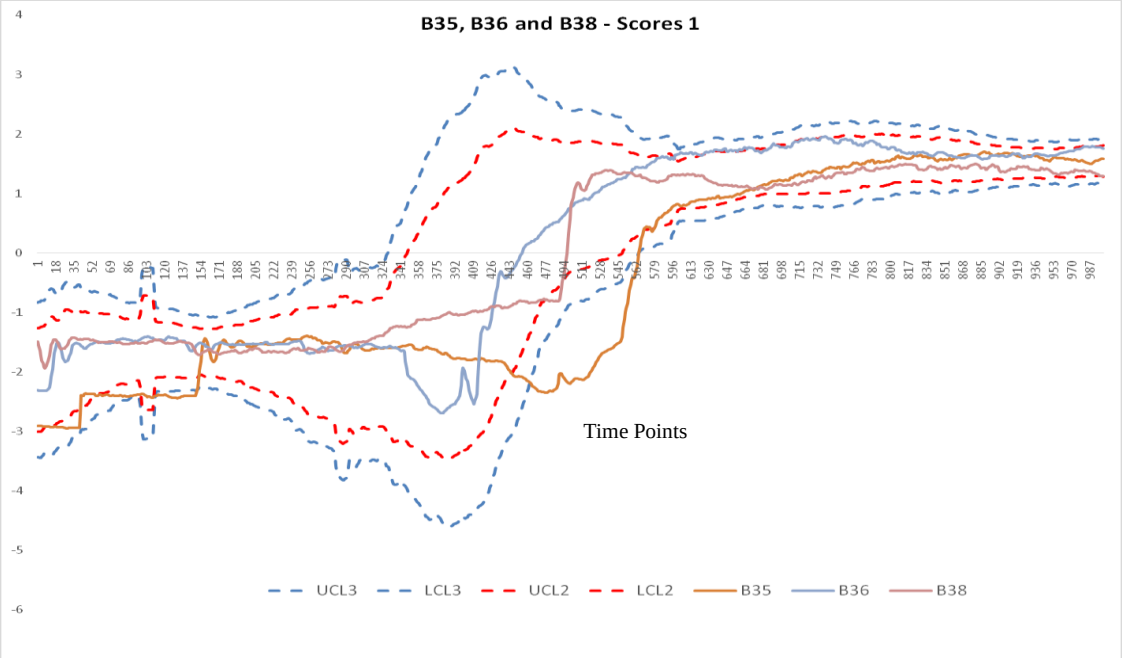

Figure 6-75: Variable-Wise analysis for t1 B35, B36 and B38
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

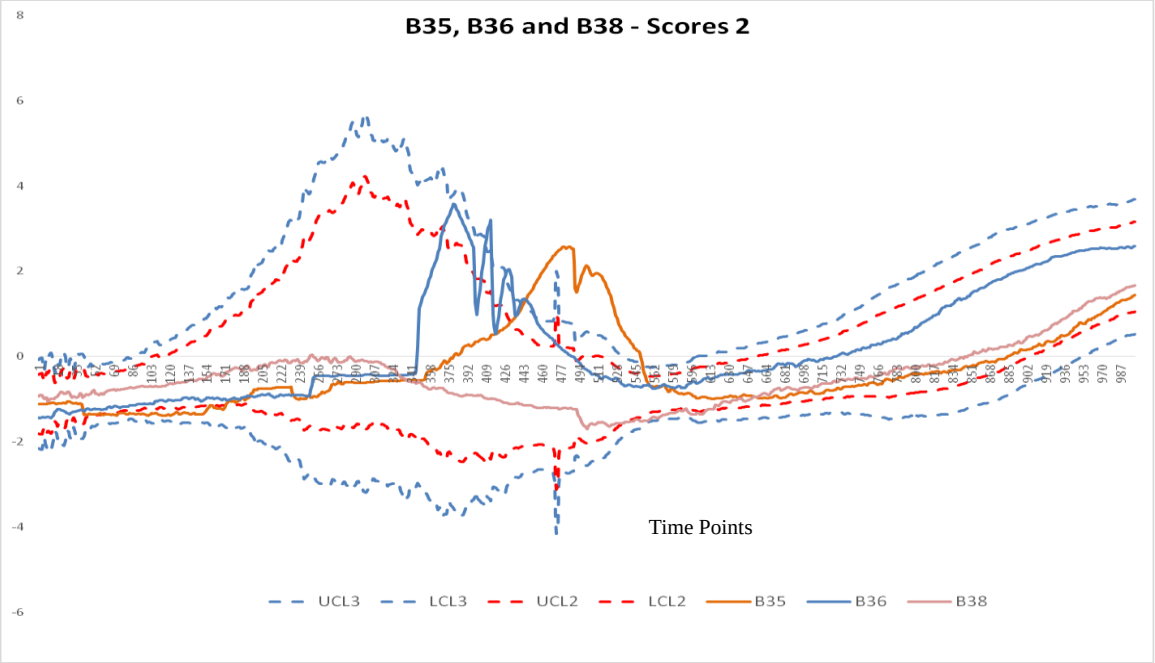

Figure 6-76: Variable-Wise analysis for t2 B35, B36 and B38
Control limits: 1) UCL3 and LCL3 are upper and lower control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the upper and lower control limits of + and – 2 standard deviation

Batches 4, 5, 6 and 7 which have been characterised as bad batches and processed in the month of October and November 2013 have been discussed further (Refer Appendix 2 Figure 8.27 to Figure 8.33 and Figure 8.42 to Figure 8.53). Looking at the results predicted by the PCA model for Batch 4, 6 and 7, these batches seem to resemble behaviour similar to a good batch. They have a lower Q-residual and Hotelling's T^2 values as seen Figure 6-60 and Figure 6-62 and do not highlight any unusual behaviour in terms of other process variables. Also studying the contents temperature profile of these batches it was observed that they do not have significant kink in the temperature profile either.

Batch 5 on the other hand is a definite abnormal batch with a high value for Q-residual and Hotelling's T^2 statistic. Further studying the contribution plot for Q-residuals (Figure 6-77 and Figure 6-78) it was observed that contents temperature and KwRise have high values as compared to the remaining bad batches. The raw data scatter plot for contents temperature against the KwRise in Figure 6-79 also reveals the unusual behaviour observed in the monitoring statistics with the KwRise increasing above 0.5 at around 200°C. The variable wise analysis plot for score 1 in Figure 6-80 also indicates an unusual behaviour between time points 300 and 450. As seen in the figure the batch is out of the upper control limits from time point 300 until it is back within the control limits around time point 450.

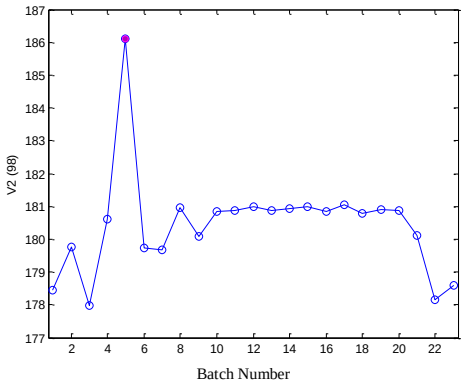

Figure 6-77: Batch 5 - V2 contribution to Q-residuals at time point 99

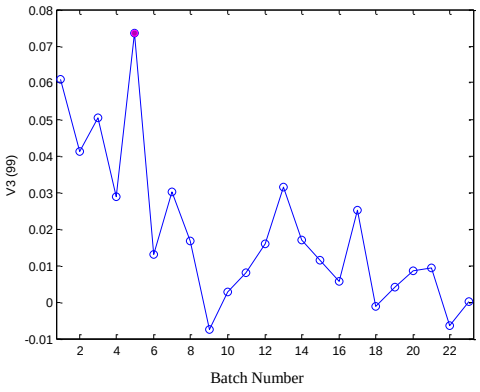

Figure 6-78 - Batch 5: V3 contribution to Q-residuals at time point 99

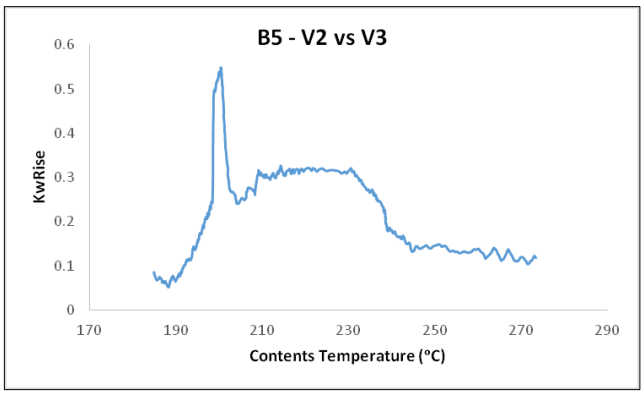

Figure 6-79: Batch 5 variable 2 vs. Variable 3 Scatter Plot

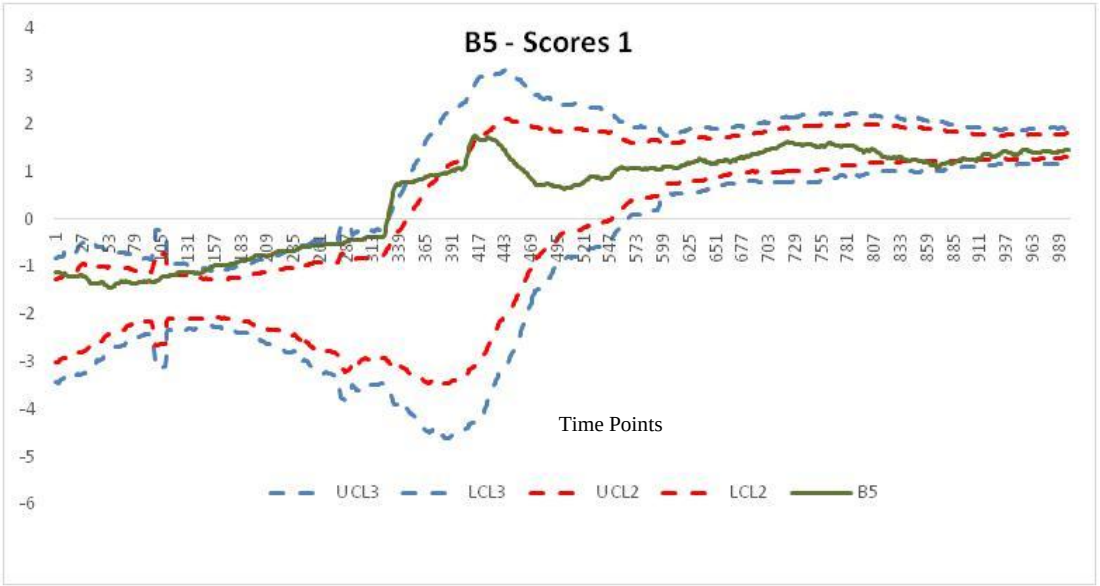

Figure 6-80: Batch 5 Variable-Wise unfolding approach unfolding analysis Scores 1
Control limits: 1) UCL3 and LCL3 are upper and lower control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the upper and lower control limits of + and – 2 standard deviation

Batch 8 although an outlier only in the Q-residuals plot does seem to show all the symptoms of a typical bad batch based on the contribution plot, KwRise or the kink in the temperature profile (Refer Appendix B2 Figure 8.54 to Figure 8.60). Batch 9 does not have a significant kink in the contents temperature profile however unusual behaviour in KwRise is observed after Hold point 1. Also examining the variable-wise unfolding approach unfolded data indicated out of control behaviour of the B9 between time point 350 and 500 in Scores 1 and Scores 2 plot (Refer Appendix B2 Figure 8.61 to Figure 8.67). Batch 10 on the other hand showcases behaviour closer to a typical good batch (Refer Appendix B2 Figure 8.68 to Figure 8.75). Batches 12, 13 and 14 are well within control in the variable-wise unfolding analysis plots and do not have a significant kink in the contents temperature profile as well. However all the three batches have a high value of KwRise after hold point 1 along with high contribution of variable 4 i.e. level control in the Hotelling's T^2 statistic (Refer Appendix B2 Figure 8.83 to Figure 8.102).

Batches 15 to 23 that have been processed in April and May 2014 show a different trend for the level control in the reactor as compared to the remaining bad batches processed in 2013 (Refer Appendix B2 Figure 8.103 to Figure 8.155). This trend is observed in all the batches that were processed during the similar period of time. Scores 2 for the variable-wise unfolding approach unfolded data resemble the behaviour in level control. It can be seen from Figure 6-82 and Figure 6-84 that this trend in the scores plot is within control for B57 which is good batch but is out of the limits for B15 i.e. bad batch

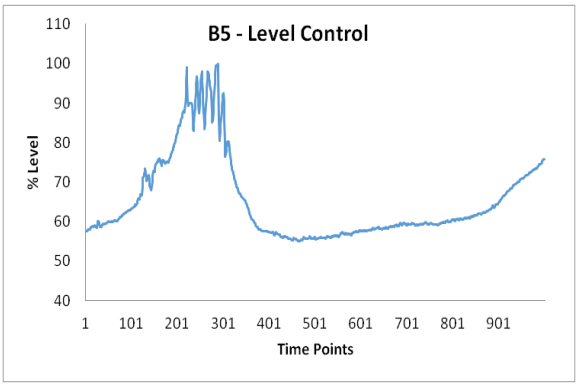

Figure 6-81: B15 Level Control

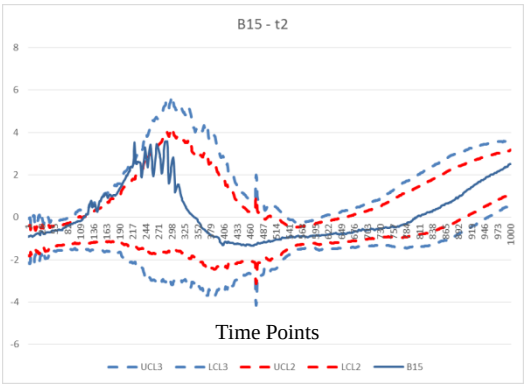

Figure 6-82: B15 Bad Batch Variable Wise unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper and lower control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the upper and lower control limits of + and – 2 standard deviation

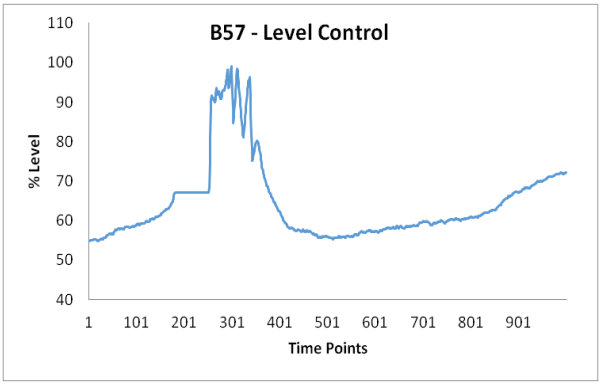

Figure 6-83: B57 Level Control

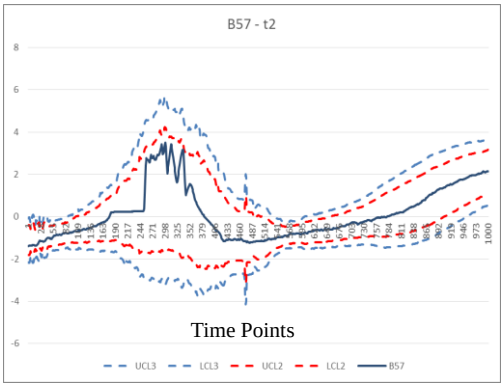

Figure 6-84: B57 Good Batch Variable Wise Unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper and lower control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the upper and lower control limits of + and – 2 standard deviation

Thus for most of the bad batches an increase in KwRise above 0.35 at 200°C is evident along with the kink in the temperature profile. For the batches produced in 2014, the trajectory of the KwRise and Level is different to the ones produced in 2013. Also for the 2014 batches along with the unusual increase in KwRise an out of control level in the Scores 2 plot is also identified mid-way through the batch.

6.7.4 Fingerprinting a crystallised batch and Conclusion

The PCA model which performs an exploratory data analysis on the process data for polymerisation reactor has been able to fingerprint the bad batches by:

1) Examining influence plot for the model developed using the batch-wise unfolding approach
2) Investigating the contribution of variables to the Q-residuals and Hotelling'sT^2 statistic.
3) Identifying one of two possible scenarios:
    a. High contribution of KwRise in Q-residual after Hold point 1 in addition to the 'kink' in the contents temperature profile
    b. High contribution of KwRise in Q-residual after Hold point 1 in addition to high contribution of Level after hold point 1
4) Monitoring the batch evolution for unfolding the data using variable-wise unfolding approach and identifying time points outside the limits.
5) Relating the observed statistics to possible evidence in raw data

Once the data has been pre-treated examining the influence plot for the batch-wise unfolded data is the first step that is clearly indicative of a 'bad' or 'crystallised' batch. Thus in addition to the current method of univariately identifying a bad batch for e.g.: monitoring the kink in the temperature profile, exploratory data analysis takes into consideration the effect of multiple variables on a batch that could lead to a particular batch being an outlier i.e. making it a bad or crystallised batch.

Extreme batches having high Q-residual and high Hotelling'sT^2 statistics are easily identified from the influence plot [Figure 6-63]. Investigating the Q-residuals and Hotelling's T^2 plot for the bad batches it was seen that most of the crystallised batches have high residuals and some of these batches also have a high Hotelling's T^2. Out of limit residuals indicate that these batches have an underlying behaviour not observed in the good batches used to develop the calibration model. While out of limit Hotelling's T^2 would entail that these batches fit the model well but have variables that are influential in determining the orientation of the model.

Except for batches 4, 7, 18 and 20 (which also are close to the confidence bounds) all the other bad batches have value high for the total Q-residual statistic. For batches with high value of Q-residuals the KwRise variable has a high contribution to this statistic either after

hold point 1 or mid-way through the batch. While the investigation of contribution plots for Hotelling's T^2 indicates a high value in either contents temperature or jacket temperature or both after hold point 1.

The next step would consist of monitoring how a batch develops across time by unfolding the data variable wise for all the retained variables. With this analysis one can follow how a batch is evolved with time and also detect the time point at which the deviation occurred.

Following the identification of unusual variables using the monitoring statistics and examining the batch trajectory using variable-wise unfolding approach the batches are further investigated by analysing the raw data. Consistent with the investigation of the contribution plots the raw data analysis for a bad batch indicated that the KwRise was above 0.35 at 200°C. For the batches produced in 2014 this unexpected increase in KwRise also coincided with uncontrolled level in the reactor. Typically the KwRise of a normal batch is below 0.35 at 200°C. By relating the results from the MSPC method to the evidence in raw data the customer was easily able to understand the working of multivariate analysis.

Out of the 23 bad batches classified as crystallised based on the kink in the temperature profile, some of these batches have demonstrated a behaviour more close to good batches. For example batches 4, 6 and 7 although classified as bad batches have a low value for Hotelling's T^2 and Q-residual statistic. Also the KwRise for these batches is not exceptionally high in the contribution plot for Q-statistics and Hotelling's T^2 and when monitored variable wise the batches seem to be within the control limits. Batch 38 on the other hand is a good batch however based on the MPCA model developed it was identified an outlying batch for the Hotelling's T^2 statistics. The scatter plot for contents temperature against KwRise indicated a higher than usual KwRise at 200°C, a characteristics observed in a typical crystallised batch. The contents temperature profile for B38 indicates a normal behaviour except for tiny 'kink' after hold 2 that would have made it difficult for the operators to identify this batch as an outlier.

Thus it is observed that MSPC is a very powerful method of clustering that takes into account the correlation between the variables. Although the root cause of the problem has not been clearly identified in this thesis the technique enables the operator/engineer to easily categorise a crystallised batch rather than simply identify the kink in the contents temperature profile. The investigation has also indicated that the problem may lie just after hold point 1 due to high contents temperature or KwRise for some batches and high jacket temperature for the

others. An observation that is consistent for most of the bad batches is the increase in KwRise value above 0.35 at 200°C which occurs before the kink appears in the contents temperature profile before the end cap temperature in the reactor.

6.8 Chapter Summary

The aim of this study is to identify the crystallisation issue by monitoring unexpected deviation in the quality and process data. This unusual phenomenon of crystallisation is observed on all of the six polymer reactors currently operating at the customer site. However for this study data was provided for two of the reactors to demonstrate the proof of concept of MSPC to the customer.

The first attempt to merge the quality and static process variable points indicated a distinct difference between the performances of the two reactors. Since the variance on each reactor was different the analysis suggested modelling each reactor separately. Two separate models were created for reactors Poly 5 and Poly 6 to perform the quality data analysis. The results from the investigation did not indicate a strong correlation of any particular quality variable to the crystallised batches. Also most of the bad batches exhibited behaviour close to the good batches making it difficult to narrow down the root cause of crystallisation by studying the quality variables only. Process data analysis was carried out on Poly 6 by unfolding the three dimensional data to lower dimensions and developing a MPCA model. The analysis clearly identified three bad batches as outliers in the model. A more than normal increase in KwRise and agitator speed was the main reason for these outlying batches. The analysis proposed that KwRise and Level must have an impact that could lead to process deviation in the reactor. With contents temperature being critical to the reaction which is then controlled by the jacket temperature, it was concluded that for further analysis it might be ideal to include contents temperature, jacket temperature, KwRise and level in the analysis instead of including all the nine variables measured during the process operation.

The particle size analysis on four other polymer reactors (Poly 1 to Poly 4) was also unable to establish a relationship between the particle size of the raw material and the crystallised batches. However few of the bad batches did cluster out separately but due to unusual observation in variables other than the particle size. For this analysis it might have been beneficial if the particle size data was also provided for the same reactors that were used for the quality and process data analysis. This would have allowed a direct comparison between the quality data, process data and particle size data.

Overall the initial study indicated that there is hidden information in the data with a few crystallised batches branching out from the cluster of good batches. This indicated that the unusual behaviour occurring in the crystallised batches eventually affects the final product quality. The customer was keen to identify the issue further and also look for a possible online

solution. This would enable the customer to identify the problem while the batch is being processed and allow the batch to be controlled before it crystallises.

The final section looked at additional process data for Poly 5 along with the process data used that was initially provided by the customer to fingerprint a crystallised batch. The fingerprinting technique discussed previously in the thesis provides a temporary solution to identify a crystallised batch and remove operator dependability for identifying the kink in the reactor.

7 Conclusions and Recommendations for Future Works

7.1 Conclusions

Project 1

ADI delivers a unique solution for business improvement by integrating the business systems such as SAP and the manufacturing systems. It was developed in line with the OPC UA standards providing an open platform and ensuring interoperability between different systems. The work presented in the thesis focussed on various problems that were addressed during product development stage of the software and the exclusive solutions developed by the author to allow successful model integration, method deployment and optimising the data exporter tool. The work discussed in the thesis facilitated the development of robust and flexible software. The method and model deployment allowed successful acquisition of at-line measurements. The process and model alarms generated during the testing would maintain a process within control and deviation with the quality of product would be monitored rapidly through at-line measurement. While optimising the MDE plug-in to include additional selection methods would facilitate the access and visualisation of measurement and meta data thus enabling process optimisation and improvement through new model development and calibration.

Project 2

Performing multivariate analysis to identify the crystallisation issue was extremely beneficial to the customer. Prior to the work discussed in this thesis there was lack of understanding of the relationships that exist between multiple variables as the data was only analysed univariately. Multivariate analysis carried out in this thesis enabled the customer to further enhance their knowledge of unknown process variations and identify possible co-relations between variables.

Analysis of the quality data concluded that contamcount variable as well as insufficient DPS did not have a correlation to a crystallised batch as previously believed by the customer. Also with the limited amount of data provided for the particle size analysis it could be concluded that raw material did not have a direct effect on the rogue batches. From the process data analysis, unfolding the batches using batch-wise unfolding approach and variable-wise unfolding approach indicated a different behaviour of contents temperature and/or level and/or KwRise just after hold point 1 especially with the crystallised batches. This unusual

behaviour of variables after hold point 1 could be the root cause of crystallisation that occurs after hold point 2.

A method was developed to fingerprint a typical crystallised batch. Previously the abnormal batch identification was solely dependent on the operator observing a ‘kink’ in the contents temperature profile with no definite method for identifying the crystallised batches. The fingerprinting method developed will provide the customer with a statistical based method of segregating the rogue batches and remove operator dependency on the ‘kink’.

7.2 Future work

Project 1

The next step for ADI will be the implementation of an online solution. An online solution would enable the customer to monitor the process in real time, improve manufacturing efficiency and reduce product waste. Managing the quantity of data produced while interfacing to batch packages and associating measurement data with the relevant batch would be some of the major challenges to be addressed. The software will also in the future accommodate up to 15 statistics that could be used by chemometricians to monitor process operations. Also while currently the software has built interface for NIR spectrometer further work will also be carried out to include Raman spectroscopic measurements within the software.

Project 2

Following the results of the analysis the customer are keen to roll out the fingerprinting method across all of their reactors and have expressed an interest to monitor the process variation online. They would also like to develop a method to achieve targeted molecular weight for each grade. This will involve some basic data analysis to model development followed by automation of the models.

In order to get to the above the following data needs to be provided by the customer:

1. Known reasons of why molecular weight varies from batch to batch.
2. Is there one grade that varies more than others?
3. Can the customer provide a list of ‘good’ batches by grade and Poly? (where good means MW close to or on target)
4. Can the customer provide a list of ‘bad’ batches by grade and reactor?

The following steps would need to be worked through to provide the solution:

1. Identify and collect data for a specific grade and reactors
2. Pre-process the data (Unfolding, auto scaling etc)
3. Perform Regression Analysis (Determine the main contributors to fluctuating MW)
4. Build a model
5. Automate the model
6. Roll out to other reactors this will require validating other reactor data against the model.

In summary once a model has been automated for an online solution any process deviation that would affect the final polymer quality i.e. molecular weight can be identified as the process is going on. This would allow the flexibility to manipulate certain process variables so that the target molecular weight is achieved in spite of the process deviations.

7.3 Industry benefit

This thesis focused on the application of multivariate analysis techniques within industry to improve and optimise industrial processes. For the first project the author acted as bridge between the software developers and customer while also developing a robust product for the end user. Along with model development and deployment the author also contributed towards method configuration and optimising MDE tool thus enabling successful implementation of the software. As a part of a bigger team the author was responsible for successfully carrying out the FAT's which further facilitated carrying out the SAT's. The work presented at various conferences attracted a number of potential future customers for the further application of the novel product in industry. The prospect of the application/utilisation of the software with various customers will allow for further research and development of the ADI software.

For the second project the author was responsible for introducing multivariate analysis techniques to the customer to identify the root cause of problem batches that were originally identified by monitoring a kink in the temperature profile. Application of multivariate analysis enabled the customer to look at various co-relations between variables and establish relationships that were previously unknown such as the increase in KwRise just after hold point 1 was observed for most of bad batches. The developed method has allowed the customer to have a well defined method to identify the crystallised batches.

References

BAKEEV, K. A. 2010. Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. 2nd ed.: John Wiley & sons Ltd.

BARNES, R. J., DHANOA, M. S. & LISTER, S. J. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy,* 43**,** 772-777.

BAUGHMAN, E. 2005. Process Analytical Chemistry: Introduction and Historical Perspective. *In:* BAKEEV, K. A. (ed.) *Process Analytical Technology*. Blackwell Oublishing Ltd.

BOX, G. E. P. 1954. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. 290-302.

BOUDREAU, M. & MCMILLAN, G. 2007. New Directions in Bioprocess Modelling and Control: Maximising Process Analytical Technology Benefits. 1st ed.: ISA - The Instrumentation System and Association Society

BRO, R. 1997. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems,* 38**,** 149-171.

CAMO. 2013. *Hierarchical Modelling Solutions* [Online]. Available: http://ll1.workcast.net/10447/1164406724395478/Ics/Hierarchical_Model_brochure_2013.pdf.

CANDOLFI, A., DE MAESSCHALCK, R., JOUAN-RIMBAUD, D., HAILEY, P. A. & MASSART, D. L. 1999. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *Journal of Pharmaceutical and Biomedical Analysis,* 21**,** 115-132.

CÁRDENAS, V., CORDOBÉS, M., BLANCO, M. & ALCALÀ, M. 2015. Strategy for design NIR calibration sets based on process spectrum and model space: An innovative approach for process analytical technology. *Journal of Pharmaceutical and Biomedical Analysis,* 114**,** 28-33.

CARROLL, J. D. & CHANG, J.-J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika,* 35**,** 283-319.

CDER. 2013. *OPS Process Analytical Technology - (PAT) Initiative* [Online]. Available: http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm088828.htm.

CHAVEZ, P.-F., SACRÉ, P.-Y., DE BLEYE, C., NETCHACOVITCH, L., MANTANUS, J., MOTTE, H., SCHUBERT, M., HUBERT, P. & ZIEMONS, E. 2015. Active content determination of pharmaceutical tablets using near infrared spectroscopy as process analytical technology tool. *Talanta***,** Article TALD1501660.

DAVIES, A. M. C. 1995. The data analysis handbook, I. E. Frank and R. Todeschini, Elsevier, Amsterdam, 1994, ISBN–444–81659–3,386 pp., Dfl325, US$185. *Journal of Chemometrics,* 9**,** 431-432.

DE MAESSCHALCK, R., JOUAN-RIMBAUD, D. & MASSART, D. L. 2000. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems,* 50**,** 1-18.

DONG, D. & MCAVOY, T. J. 1996. Batch tracking via nonlinear principal component analysis. *AIChE Journal,* 42**,** 2199-2208.

ENGEL, J., GERRETZEN, J., SZYMAŃSKA, E., JANSEN, J. J., DOWNEY, G., BLANCHET, L. & BUYDENS, L. M. C. 2013. Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry,* 50**,** 96-106.

ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., J.TRYGG, C.WIKSTROM & WOLD, S. 2006. *Multi- and Megavariate Data Analysis,* Umetrics AB.

FAGGIAN, A., FACCO, P., DOPLICHER, F., BEZZO, F. & BAROLO, M. 2009. Multivariate statistical real-time monitoring of an industrial fed-batch process for the production of specialty chemicals. *Chemical Engineering Research and Design,* 87**,** 325-334.

FDA, 2004. *Guidance for Industry PAT- A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance* [Online]. Available: http://www.fda.gov/downloads/Drugs/Guidances/ucm070305.pdf.

FENG, S. C. 2000. Manufacturing planning and execution software interfaces. *Journal of Manufacturing Systems,* 19**,** 1-17.

FEVOTTE, G., BARUDIO, I. & GUILLOT, J. 1996. Reaction CalorimetryAn adaptive inferential measurement strategy for on-line monitoring of conversion in polymerization processes. *Thermochimica Acta,* 289**,** 223-242.

FOUNDATION, O. 2013. *About OPC* [Online]. Available: http://www.opcfoundation.org/Default.aspx/01_about/01_whatis.asp?MID=AboutOPC [Accessed 10th August 2013].

GANGULY, J. & VOGEL, G. 2006. Process Analytical Technology and Scalable Automation for Bioprocess Control and Monitoring - A Case Study. *ISPE,* 26.

GELADI, P. 1989. Analysis of multi-way (multi-mode) data. *Chemometrics and Intelligent Laboratory Systems,* 7**,** 11-30.

GELADI, P. & KOWALSKI, B. R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta,* 185**,** 1-17.

GUPTA, M. & KOHLI, A. 2006. Enterprise resource planning systems and its implications for operations function. *Technovation,* 26**,** 687-696.

HARSHMAN, R. A. & LUNDY, M. E. 1994. PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis,* 18**,** 39-72.

ICH-Q8 2009. Guidance for Industry: Q8 (R2) Pharmaceutical Development

ISA. 2010. *Technology ISA - 95* [Online]. Available: http://www.isa-95.com/subpages/technology/isa-95.php?PHPSESSID=46047dff8aebb0f3d31fc7812ddd5356 [Accessed 28th July 2013].

JACKSON, J. E. & MUDHOLKAR, G. S. 1979. Control Procedures for Residuals Associated With Principal Component Analysis. *Technometrics,* 21**,** 341-349.

KEITHLEY, R. B., MARK WIGHTMAN, R. & HEIEN, M. L. 2009. Multivariate concentration determination using principal component regression with residual analysis. *TrAC Trends in Analytical Chemistry,* 28**,** 1127-1136.

KHATIBISEPEHR, S., HUANG, B. & KHARE, S. 2013. Design of inferential sensors in the process industry: A review of Bayesian methods. *Journal of Process Control,* 23**,** 1575-1596.

KHERKHOF, P. V. D., VANLAER, J., GINS, G. & IMPE, J. F. M. V. 2013. Contribution plots for Statistical Process Control:analysis of the smearing-out effect. Available: http://www.nt.ntnu.no/users/skoge/prost/proceedings/ecc-2013/data/papers/0912.pdf [Accessed 6th August 2014].

KOURTI, T. & MACGREGOR, J. F. 1995. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems,* 28**,** 3-21.

LEE, J.-M., YOO, C. & LEE, I.-B. 2003. On-line batch process monitoring using a consecutively updated multiway principal component analysis model. *Computers & Chemical Engineering,* 27**,** 1903-1912.

LEE, J.-M., YOO, C. K. & LEE, I.-B. 2004. Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis. *Journal of Biotechnology,* 110**,** 119-136.

MACGREGOR, J. F. 1988. CONTROL OF POLYMERIZATION REACTORS. *In:* MCGREAVY, C. (ed.) *Dynamics and Control of Chemical Reactors and Distillation Columns.* Oxford: Pergamon.

MACGREGOR, J. F. & KOURTI, T. 1995. Statistical process control of multivariate processes. *Control Engineering Practice,* 3**,** 403-414.

MARJANOVIC, O., LENNOX, B., SANDOZ, D., SMITH, K. & CROFTS, M. 2006. Real-time monitoring of an industrial batch process. *Computers & Chemical Engineering,* 30**,** 1476-1481.

MARTENS, H., JENSEN, S. A. & GELADI, P. 1983. Multivariate linearity transformation for near infrared reflectance spectra of meat. Application Spectroscopy. *Applied Statistics***,** 235-267.

MARTENS, H. & STARK, E. 1991. Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis,* 9**,** 625-635.

MARTIN, E. 2014. Multivariate Statistical Process Control and Process Performance Monitoring.

MARTIN, E. B., MORRIS, A. J., PAPAZOGLOU, M. C. & KIPARISSIDES, C. 1996. Batch process monitoring for consistent production. *Computers & Chemical Engineering,* 20, Supplement 1**,** S599-S604.

MCARTHUR, H., & SPALDING, D. 2004. Engineering Materials Science: Properties,Uses, Degradation,Remediation. Woodhead Publishing.

MILLER, C. E. 2005. Chemometrics in Process Analytical Chemistry. *In:* BAKEEV, K. A. (ed.) *Process Analytical Technology.* India: Blackwell Publishing.

MOREL, G., PANETTO, H., ZAREMBA, M. & MAYER, F. 2003. Manufacturing Enterprise Control and Management System Engineering: paradigms and open issues. *Annual Reviews in Control,* 27**,** 199-209.

NOMIKOS, P. 1996. Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis World Batch Forum, Toronto, May 1996. *ISA Transactions,* 35**,** 259-266.

NOMIKOS, P. & F.MACGREGOR, J. 1995. Mutlivariate SPC Charts for Monitoring Batch Processes. *Technometrics,* 37.

NOMIKOS, P. & MACGREGOR, J. F. 1994. Monitoring batch processes using multiway principal component analysis. *AIChE Journal,* 40**,** 1361-1375.

NOMIKOS, P. & MACGREGOR, J. F. 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics,* 37**,** 41-59.

OHSHIMA, M. & TANIGAKI, M. 2000. Quality control of polymer production processes. *Journal of Process Control,* 10**,** 135-148.

PEARSON, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6,* 2**,** 559-572.

PERKINELMER. 2005. FT-IR Spectroscopy Attenuated Total Reflectance (ATR). Available: www.perkinelmer.com [Accessed 15th August 2013].

RATHORE, A. S. 2014. QbD/PAT for bioprocessing: moving from theory to implementation. *Current Opinion in Chemical Engineering,* 6**,** 1-8.

RICHARDS, J. R. & CONGALIDIS, J. P. 2006. Measurement and control of polymerization reactors. *Computers & Chemical Engineering,* 30**,** 1447-1463.

RINNAN, Å., BERG, F. V. D. & ENGELSEN, S. B. 2009a. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry,* 28**,** 1201-1222.

RINNAN, Å., NØRGAARD, L., BERG, F. V. D., THYGESEN, J., BRO, R. & ENGELSEN, S. B. 2009b. Chapter 2 - Data Pre-processing. *In:* SUN, D.-W. (ed.) *Infrared Spectroscopy for Food Quality Analysis and Control.* San Diego: Academic Press.

SANCHEZ, E. & KOWALSKI, B. R. 1990. Tensorial resolution: A direct trilinear decomposition. *Journal of Chemometrics,* 4**,** 29-45.

SAVITZKY, A. & GOLAY, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry,* 36**,** 1627-1639.

SCHAEFER, C., LECOMTE, C., CLICQ, D., MERSCHAERT, A., NORRANT, E. & FOTIADU, F. 2013. On-line near infrared spectroscopy as a Process Analytical Technology (PAT) tool to control an industrial seeded API crystallization. *Journal of Pharmaceutical and Biomedical Analysis,* 83**,** 194-201.

SHARMIN, R., SUNDARARAJ, U., SHAH, S., VANDE GRIEND, L. & SUN, Y.-J. 2006. Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant. *Chemical Engineering Science,* 61**,** 6372-6384.

SUBRAMANIAN, A. & RODRIGUEZ-SAONA, L. 2009. Chapter 7 - Fourier Transform Infrared (FTIR) Spectroscopy. *Infrared Spectroscopy for Food Quality Analysis and Control.* San Diego: Academic Press.

THERMO FISHER SCIENTIFIC, I. 2015. *Fourier Transform Infrared Spectroscopy (FTIR)* [Online]. Available: http://www.thermoscientific.com/en/products/fourier-transform-infrared-spectroscopy-ftir.html [Accessed 3rd August 2015].

UMBLE, E. J., HAFT, R. R. & UMBLE, M. M. 2003. Enterprise resource planning: Implementation procedures and critical success factors. *European Journal of Operational Research,* 146**,** 241-257.

UNSCRAMBLER, T. 2014. The Unscrabler - Help.

VAN DEN BERG, F., LYNDGAARD, C. B., SØRENSEN, K. M. & ENGELSEN, S. B. 2013. Process Analytical Technology in the food industry. *Trends in Food Science & Technology,* 31**,** 27-35.

WESTERHUIS, J. A., GURDEN, S. P. & SMILDE, A. K. 2000. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems,* 51**,** 95-114.

WOLD, S. 1978. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics,* 20**,** 397-405.

WOLD, S., KETTANEH-WOLD, N., MACGREGOR, J. F. & DUNN, K. G. 2009. 2.10 - Batch Process Modeling and MSPC. *In:* WALCZAK, S. D. B. T. (ed.) *Comprehensive Chemometrics.* Oxford: Elsevier.

WOLD, S., KETTANEH, N., FRIDÉN, H. & HOLMBERG, A. 1998. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and Intelligent Laboratory Systems,* 44**,** 331-340.

WOLD, S., RUHE, A., WOLD, H. & DUNN, I. W. 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM Journal on Scientific and Statistical Computing,* 5**,** 735-743.

WOLD, S., SJÖSTRÖM, M. & ERIKSSON, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems,* 58**,** 109-130.

ZALAZNIK, M., KALIN, M. & NOVAK, S. 2015. Influence of the processing temperature on the tribological and mechanical properties of poly-ether-ether-ketone (PEEK) polymer. *Tribology International,* 94**,** 92-97.

8 Appendices

Appendix A: Chapter 4

The following screenshots are the models that were developed within Unscrambler V10.3 and further imported within the ADI system for the FAT's

Results		Level-1		
		Y Predicted	Sample Leverage	X Sample Validation Residuals
		Property 6		
		1	2	3
sample-1	1	9.9680	0.0811	0.6764
sample-2	2	9.5969	0.0848	0.6651
sample-3	3	26.9974	0.0229	0.0633
sample-4	4	27.6463	0.0250	0.0647
sample-5	5	9.9909	0.0809	0.6417
sample-6	6	8.8490	0.0926	0.6437
sample-7	7	26.5662	0.0216	0.0517
sample-8	8	27.3326	0.0239	0.0570
sample-9	9	5.7184	0.1297	0.4666
sample-10	10	6.3215	0.1220	0.4792
sample-11	11	10.7836	0.0734	0.6997
sample-12	12	9.5404	0.0854	0.6696
sample-13	13	9.2385	0.0885	0.6530
sample-14	14	9.8911	0.0819	0.6564
sample-15	15	10.2092	0.0788	0.6558
sample-16	16	10.8771	0.0725	0.6648
sample-17	17	10.9304	0.0720	0.6659
sample-18	18	20.3296	0.0197	0.6914
sample-19	19	10.0275	0.0806	0.6672
sample-20	20	10.0318	0.0805	0.6531

Figure 8-1: Property 6

Results		Level-1		
		Y Predicted	Sample Leverage	X Sample Validation Residuals
		Property 1		
		1	2	3
sample-1	1	73.4595	12.2349	1.2370
sample-2	2	73.2645	11.6902	1.0485
sample-3	3	81.6157	0.3802	0.9523
sample-4	4	89.1909	0.4081	0.8457
sample-5	5	54.6295	5.9851	1.0055
sample-6	6	55.1664	6.4210	0.9971
sample-7	7	89.2289	0.4567	1.0131
sample-8	8	84.6941	0.4419	0.8641
sample-9	9	47.3179	5.5082	1.1655
sample-10	10	59.9787	7.7553	0.6571
sample-11	11	69.3084	9.9696	1.0789
sample-12	12	62.2497	9.1694	1.1706
sample-13	13	65.7522	7.8887	0.6071
sample-14	14	65.4634	7.3226	0.5832
sample-15	15	65.8358	8.7723	0.9590
sample-16	16	58.2320	7.2935	0.7537
sample-17	17	64.0885	9.0473	0.7356
sample-18	18	-21.7514	12.5532	4.7861
sample-19	19	60.6484	8.1512	1.4056
sample-20	20	60.2752	7.5625	1.0863

Figure 8-2: Property1 model

Results		Level-1		
		Y Predicted	Sample Leverage	X Sample Validation Residuals
		Property 2		
		1	2	3
sample-1	1	62.5454	0.0060	0.0211
sample-2	2	62.6061	0.0060	0.0412
sample-3	3	25.8856	0.0032	0.0484
sample-4	4	25.5663	0.0032	0.0487
sample-5	5	60.7616	0.0054	0.0107
sample-6	6	60.7438	0.0054	0.0127
sample-7	7	25.2156	0.0033	0.0529
sample-8	8	26.3635	0.0031	0.0488
sample-9	9	58.5601	0.0047	0.0122
sample-10	10	57.8178	0.0045	0.0118
sample-11	11	63.7574	0.0064	0.0255
sample-12	12	63.6755	0.0064	0.0212
sample-13	13	63.9880	0.0065	0.0173
sample-14	14	62.8736	0.0061	0.0184
sample-15	15	62.3929	0.0059	0.0203
sample-16	16	63.4647	0.0063	0.0210
sample-17	17	62.3667	0.0059	0.0221
sample-18	18	28.1501	0.0028	0.2105
sample-19	19	60.9346	0.0054	0.0290
sample-20	20	61.3864	0.0056	0.0160

Figure 8-3: Property 2 model

Spectral Diagnostic model

Results		RMS 1		Peak Model 1		PCA 1			
		RMS	State	Value	State	Projected Ho	State	Projected Sa	State
		1	2	3	4	5	6	7	8
2	1	0.0033	Warning	285.9850	AlarmHigh	39.6643	Alarm	0.1900	Alarm
17	2	0.0057	Alarm	279.2423	WarningHigh	36.2053	Alarm	0.1738	Alarm
1	3	0.0014	Normal	241.9158	Normal	76.6830	Alarm	0.3630	Alarm
22	4	0.0032	Warning	207.6467	Normal	101.9460	Alarm	0.4810	Alarm
9	5	0.0012	Normal	187.0411	Normal	190.7850	Alarm	0.8962	Alarm

Figure 8-4: Spectral Diagnostic Model

Appendix B: Chapter 6

Appendix B1: Matlab Code for Unfolding

The following Matlab code was used for unfolding the data using the N&M approach and ariable-Wise unfolding Approach

Batch wise unfolded Matlab code

```
x=B1:

transpose_x=x':

reshape_x=reshape(transpose_x,1,6000):

% 6000- this is because the matrix 1500x4 is being reshaped to 1 row to

% unfold the matrix.

B_1=reshape_x:

%this is the reshaped matrix to have

%b1[t1(v1,v2...vn),t2(v1,v2...vn),....tn(v1 to vn)]

UNFOLD_P6_1500=vertcat(B_1,B_2,B_3,B_4,B_5,B_6,B_7,B_8,B_9,B_10,B_11,B_12,B_13,B_14,B_15,B_16,B_17,B_18,B_19,B_20,B_21,B_22,B_23,B_24,B_25,B_26,B_27,B_28,B_29,B_30,B_31,B_32,B_33,B_34,B_35,B_36,B_37,B_38,B_39,B_40,B_41,B_42,B_43,B_44,B_45,B_46,B_47,B_48,B_49,B_50,B_51,B_52,B_53,B_54,B_55,B_56,B_57,B_58,B_59,B_60,B_61,B_62,B_63,B_64,B_65,B_66,B_67,B_68,B_69):

%Put all the unfolded batches one after the other
```

Variable-Wise Unfolded Matlab Code

The following code unfolds the good batches according to Variable-Wise unfolding approach

```
X =
horzcat(B22,B23,B24,B25,B26,B27,B28,B29,B30,B31,B32,B33,B34,B35,B36,B37,B38,B39,B40,B41,B42,B43,B44,B45,B46,B47,B48,B49,B50,B51,B52,B53,B54,B55,B56,B57,B58,B59,B60):

X1 = X':
```

X2=reshape (X1,4,39000):

GoodW=X2':

The following code unfolds the bad batches according to Wolds Approach

BadX = horzcat(B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14,B15,B16,B17,B18,B19,B20,B21):

BadX1=BadX':

BadX2=reshape(BadX1,4,21000):

BadW=BadX2':

The following code was used in Matlab to reshape the scores matrix to monitor the batch as it progressed with time. The calculation of upper control limit and lower control limit for the refolded scores matrix have also been included in this code.

x1 = t1(1:46,:):

y1=t1(47:69,:):

t1=reshape(scores1,69,1000):

t1_1=reshape(scores1,1000,69):

aa=t1_1':

plot(aa(47,:),'DisplayName','aa(47,:)','YDataSource','aa(47,:)'):figure(gcf)

plot(aa(44,:),'DisplayName','aa(44,:)','YDataSource','aa(44,:)'):figure(gcf)

goodt1 = scores1(1:46000,:):

goodt2=scores2(1:46000,:):

badt1=scores1(46001:69000,:):

badt2=scores2(46001:69000,:):

t1g=reshape(goodt1,46,1000):

t2g=reshape(goodt2,46,1000):

```
t1b=reshape(badt1,23,1000):

t2b=reshape(badt2,23,1000):

plot(t1b(1,:),'DisplayName','t1b(1,:)','YDataSource','t1b(1,:)'):figure(gcf)

plot(t1g(4,:),'DisplayName','t1g(4,:)','YDataSource','t1g(4,:)'):figure(gcf)

Meant1g=mean(goodt1):

stdvt1=std(goodt1):

Meant1g=mean(goodt1,1):

Meant1g=mean(goodt1):

Meant1g=mean(t1g):

stdvt1=std(t1g):

UCL=(Meant1g+2*stdvt1):

UCL=(Meant1g+3*stdvt1):

UCL3=(Meant1g+3*stdvt1):

UCL2=(Meant1g+2*stdvt1):

LCL2=(Meant1g-2*stdvt1):

LCL3=(Meant1g-3*stdvt1):

Meant2g=mean(t2g):

stdvt2=std(t2g):

UCL2_2=Meant2g+2*stdvt2):

UCL2_2=(Meant2g+2*stdvt2):

LCL2_2=(Meant2g-2*stdvt2):

LCL3_2=(Meant2g-3*stdvt2):

UCL3_2=(Meant2g+3*stdvt2):
```

Appendix B2: Crystallised batch plots used for investigating rogue batches

The following plots for the crystallised batches were used to investigate additional process data analysis on Polymer 5

Batch 1

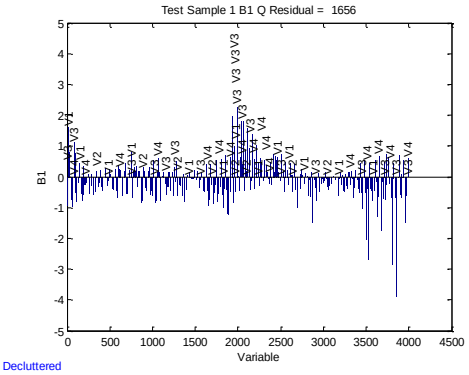

Figure 8-5: Batch 1 Q residual contribution

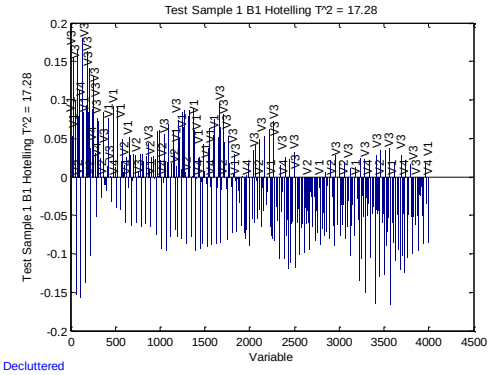

Figure 8-6: Batch 1 Hotelling's T2 contribution

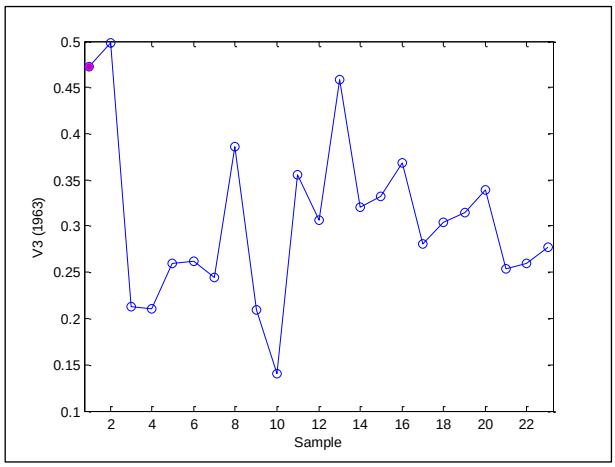

Figure 8-7: Batch 1 KwRise value at time point 1963

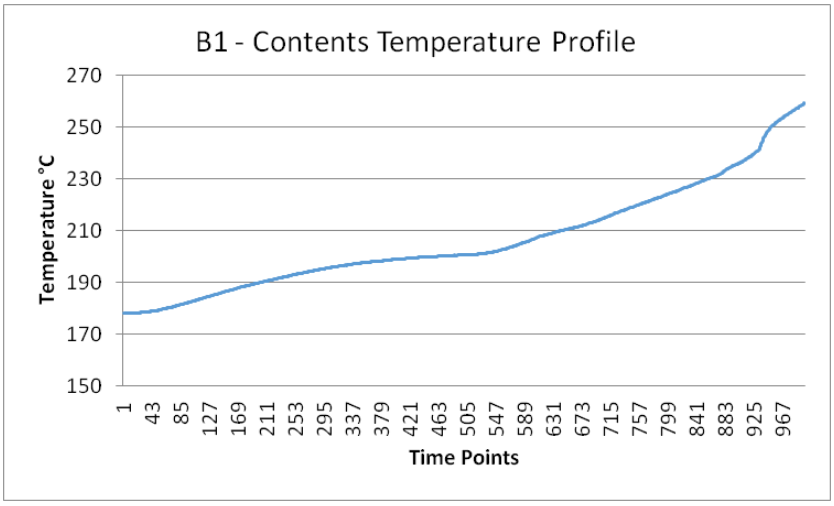

Figure 8-8: B1 Contents Temperature Profile

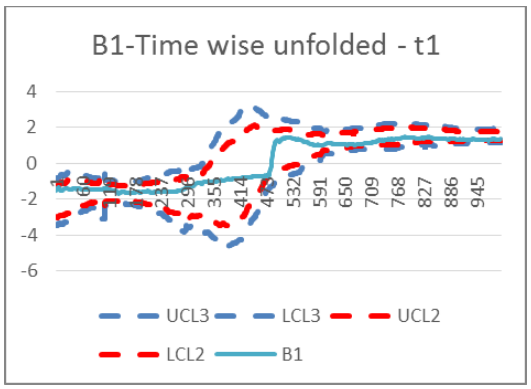

Figure 8-9: B1 Variable-Wise unfolding approach unfolded Scores 1
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

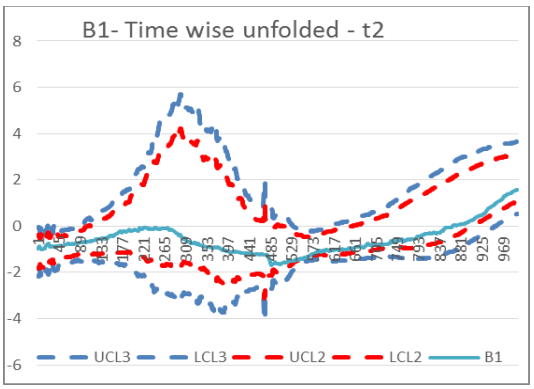

Figure 8-10: B1 Variable-Wise unfolding approach unfolded Scores 2
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

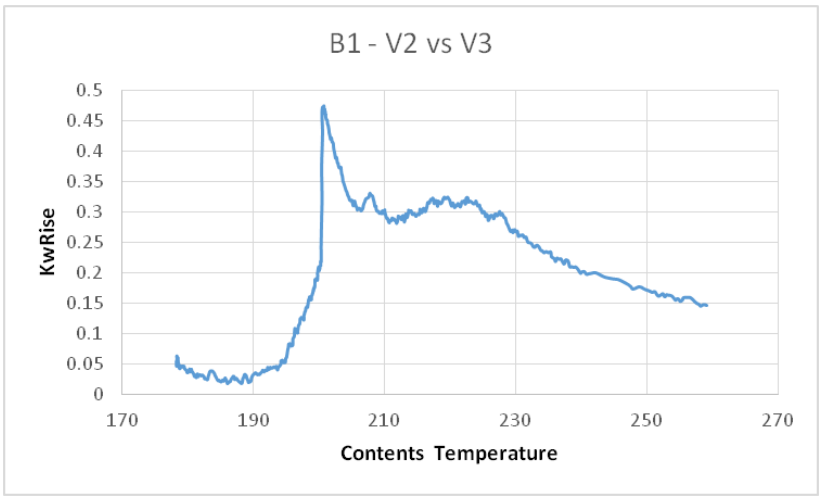

Figure 8-11: B1 Contents temperature vs. KwRise

Batch 2

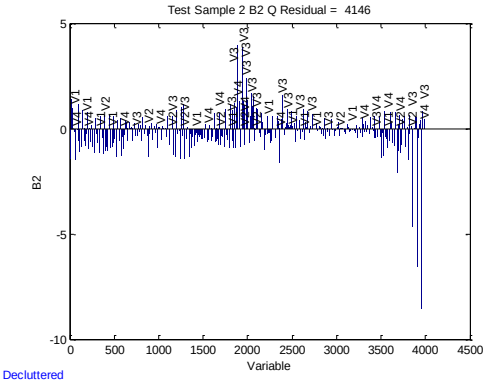

Figure 8-12: P5 B2 Q-residual contribution

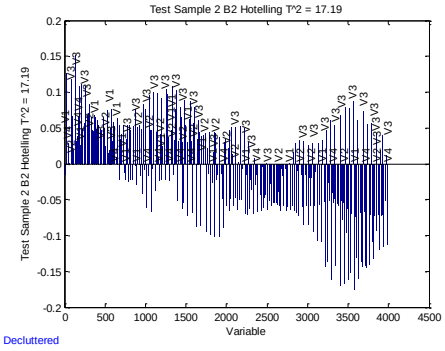

Figure 8-13: P5 B2 Hotelling's T2 contribution

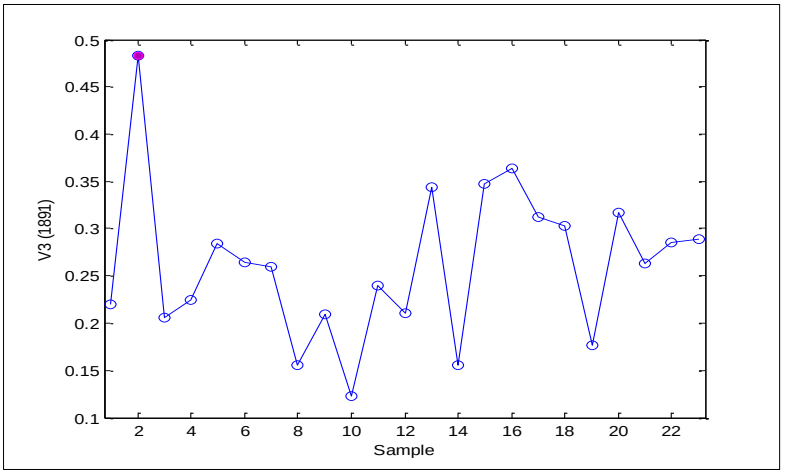

Figure 8-14: B2 KwRise contribution at time point 1891

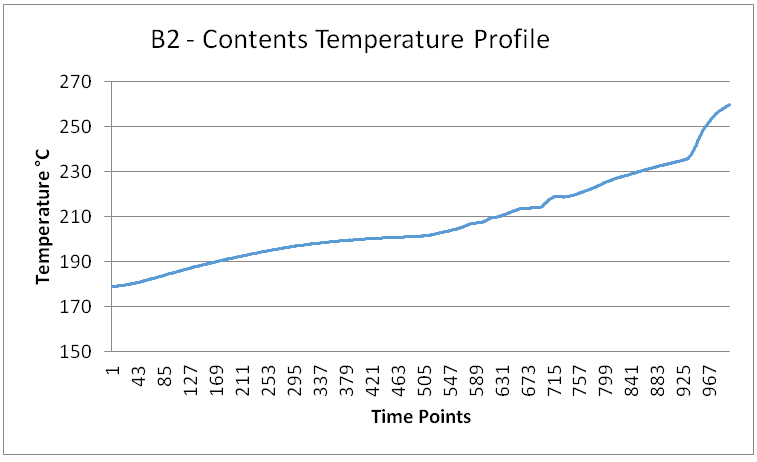

Figure 8-15: B2 Contents Temperature Profile

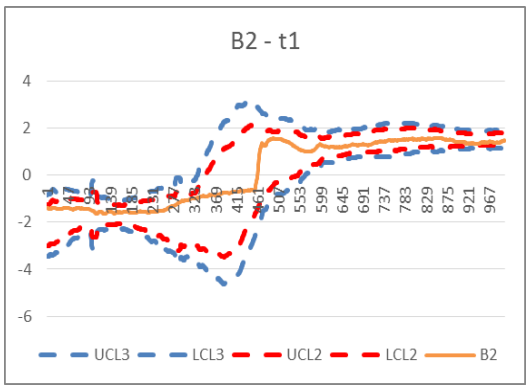

Figure 8-16: B2 Variable-Wise unfolding approach unfolded Scores 1
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

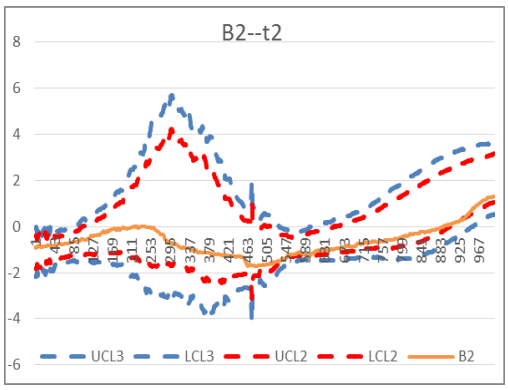

Figure 8-17: B2 Variable-Wise unfolding approach unfolded Scores 2
Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

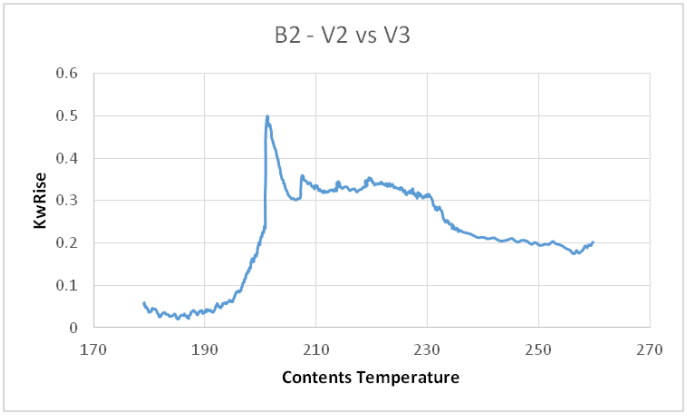

Figure 8-18: B2 Contents Temperature vs. KwRise

Batch 3

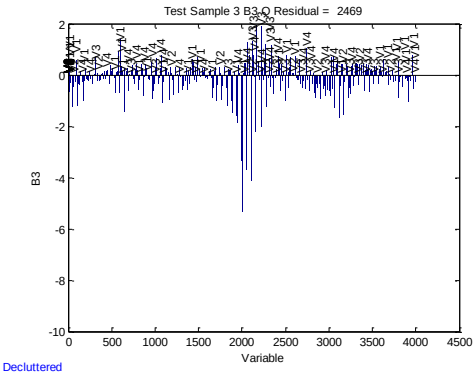

Figure 8-19: P5 B3 Q-residual contribution

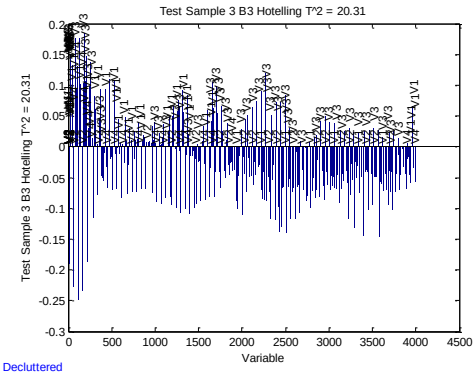

Figure 8-20: P5 B3 Hotelling's T2 contribution

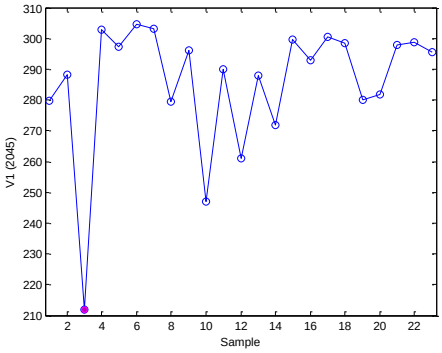

Figure 8-21: P5 B3 V1 contribution at time point 2045

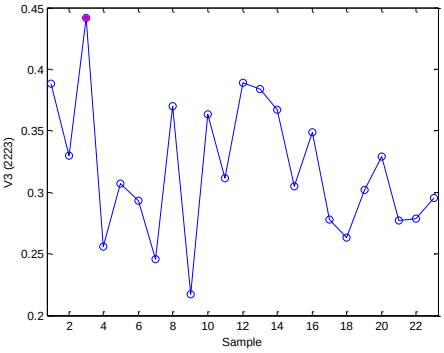

Figure 8-22: P5 B3 V3 contribution at time point 2223

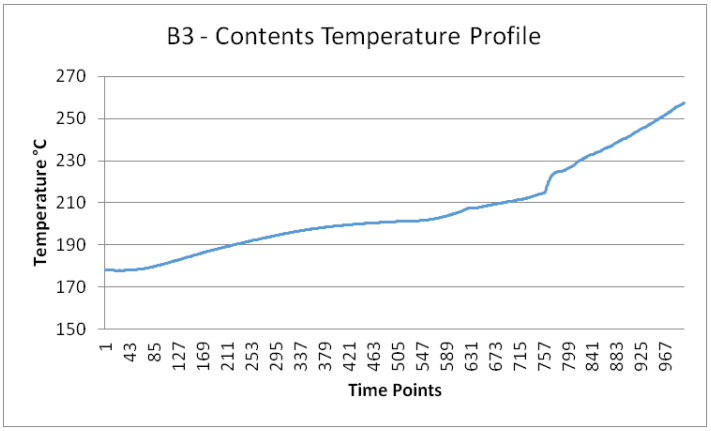

Figure 8-23: B3 Contents Temperature Profile

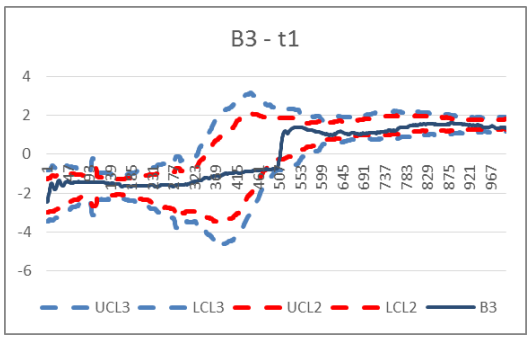

Figure 8-24: B3 Variable-Wise unfolding approach unfolded Scores 1

Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

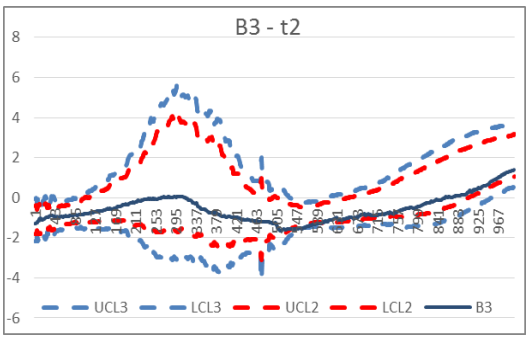

Figure 8-25: Variable-Wise unfolding approach unfolded Scores 2

Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

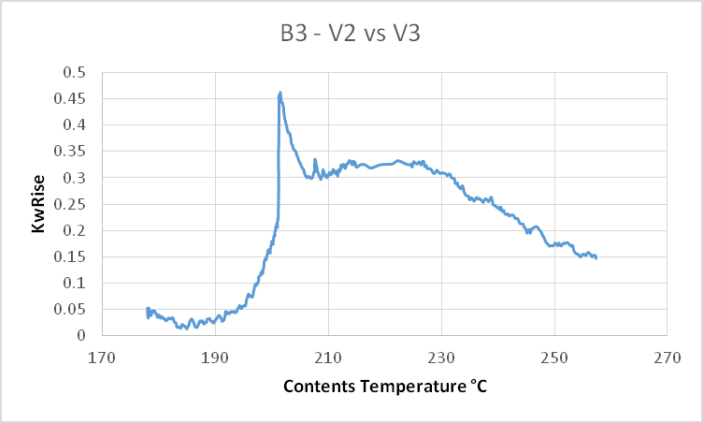

Figure 8-26: B3 Contents temperature vs. KwRise

Batch 4

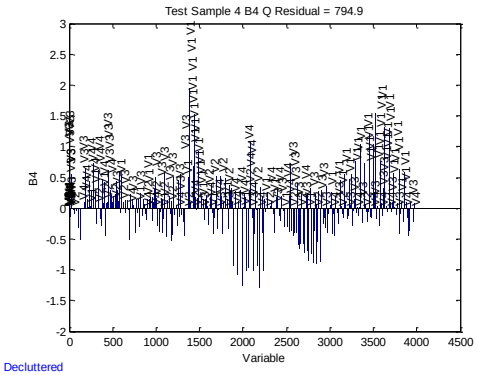

Figure 8-27: P5 B4 Q-residual contribution

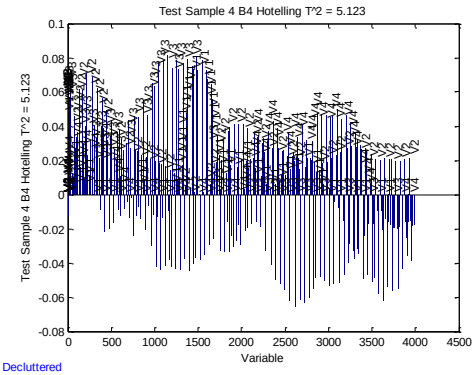

Figure 8-28: P5 B4 Hotelling's T2 contribution

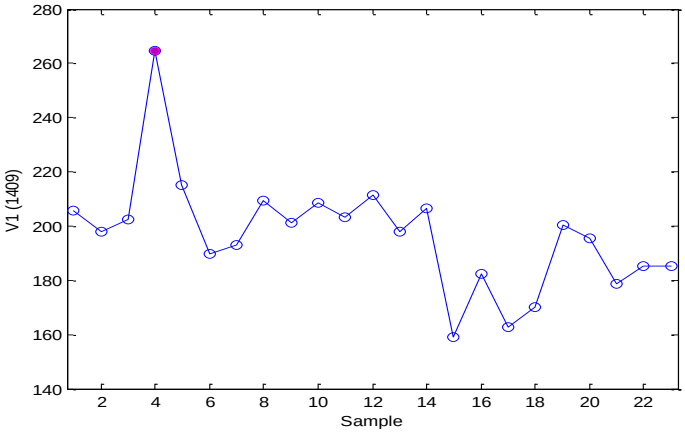

Figure 8-29: P5 B4 Q residual contribution V1 at time 1409

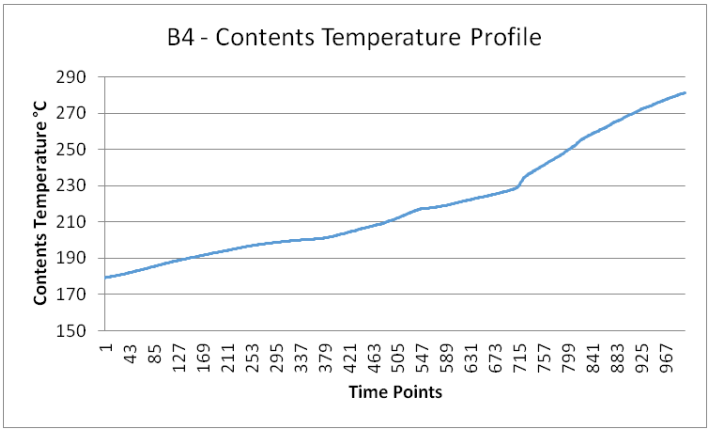

Figure 8-30: B4 Contents Temperature Profile

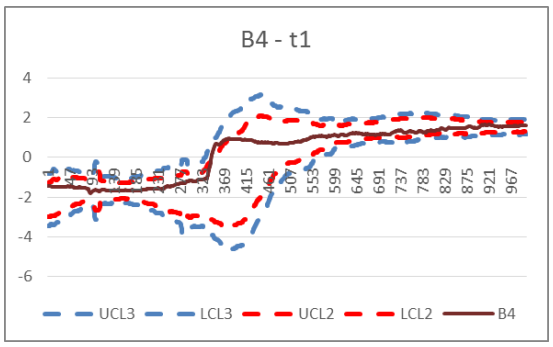

Figure 8-31: Batch 4 Variable-Wise unfolding approach unfolded Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

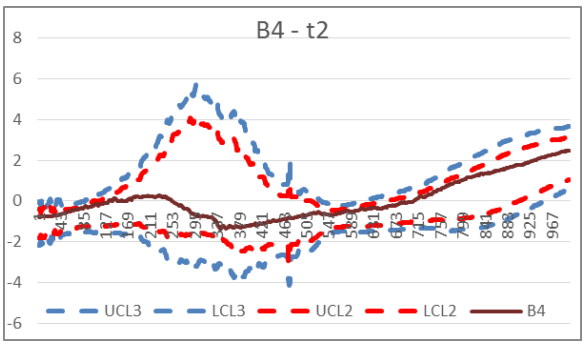

Figure 8-32: Batch 4 Variable-Wise unfolding approach unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

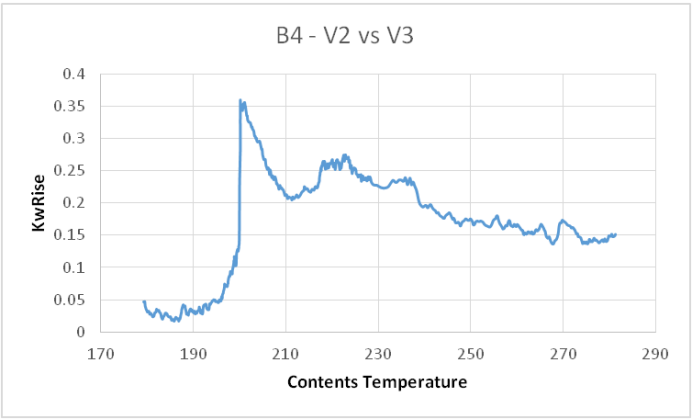

Figure 8-33: B4 Contents Temperature Vs KwRise

Batch 5

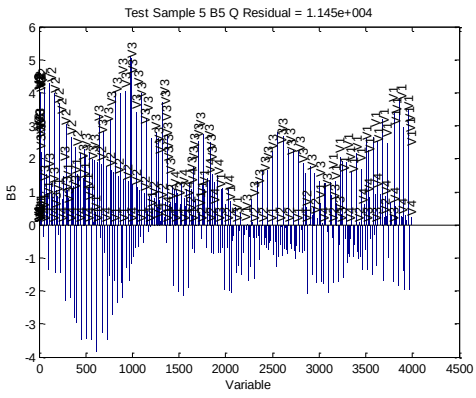

Figure 8-34: B5 Hotelling's T^2 contribution plot

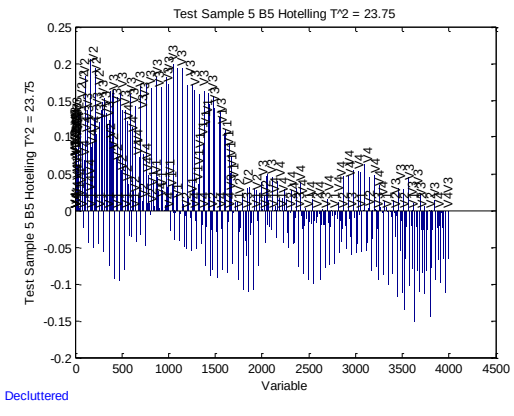

Figure 8-35: B5 Q-residuals contribution plot

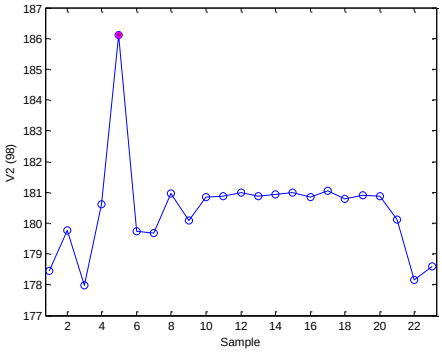

Figure 8-36: P5 B5 Q residual contribution of V2 at time point 98

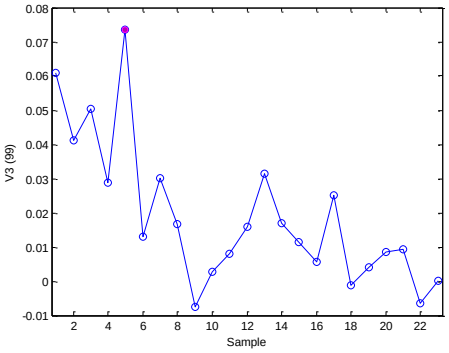

Figure 8-37: P5 B5 Q residual contribution of V3 at time point 98

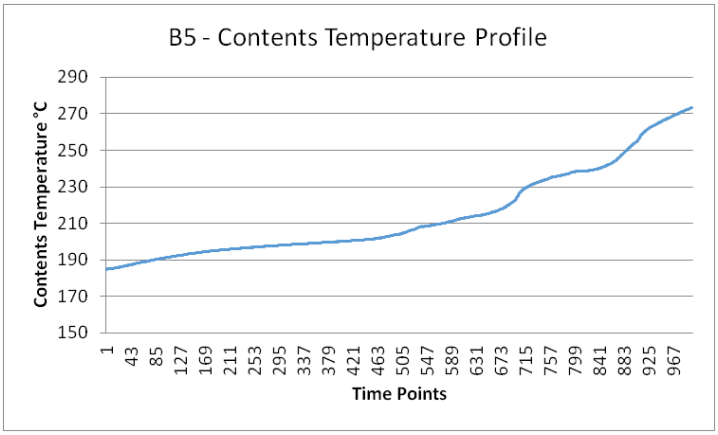

Figure 8-38: B5 Contents Temperature Profile

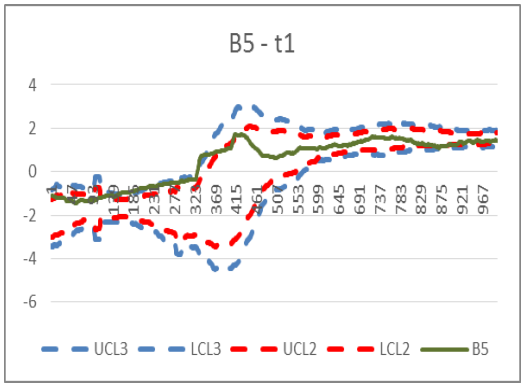

Figure 8-39: B5 Variable-Wise unfolding approach unfolded Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

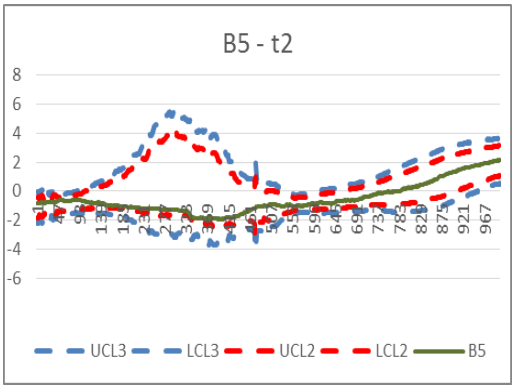

Figure 8-40: B5 Variable-Wise unfolding approach unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

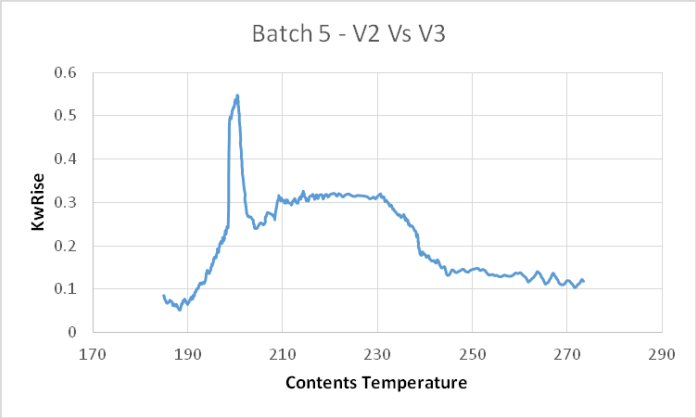

Figure 8-41: B5 Contents temperature profile vs KwRise

Batch 6

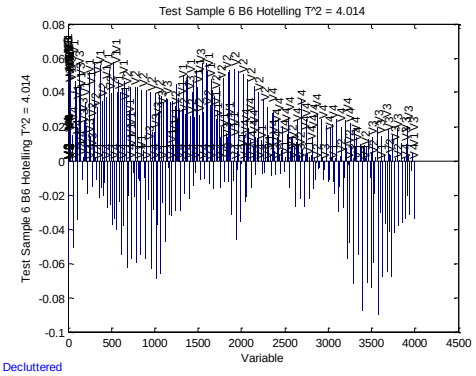

Figure 8-42: P5 B6 Hotelling's T2 contribution

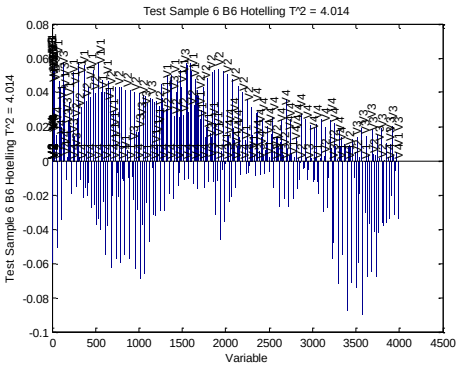

Figure 8-43: P5 B6 Q-residuals contribution

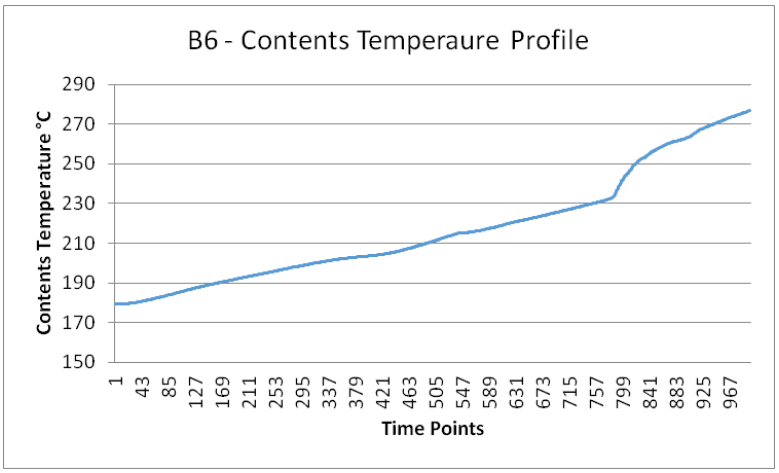

Figure 8-44: B6 Contents Temperature Profile

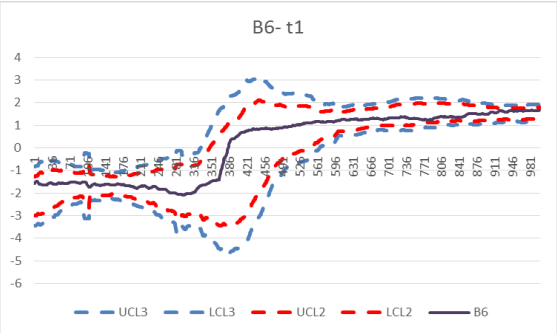

Figure 8-45: P5 B6 Variable-Wise unfolding approach unfolded Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

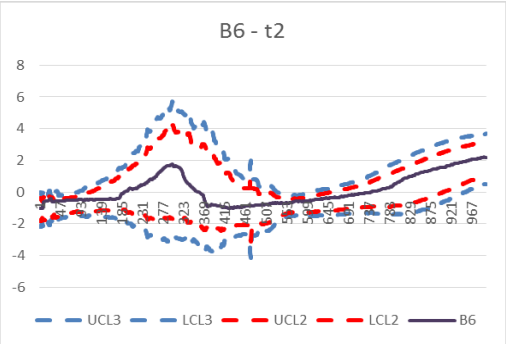

Figure 8-46: P5 B6 Variable-Wise unfolding approach unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

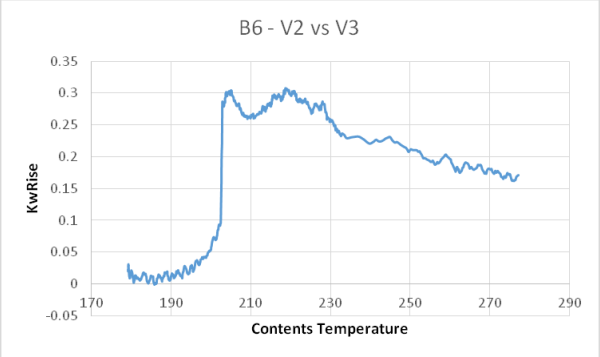

Figure 8-47: B6 Contents Temperature vs KwRise

Batch 7

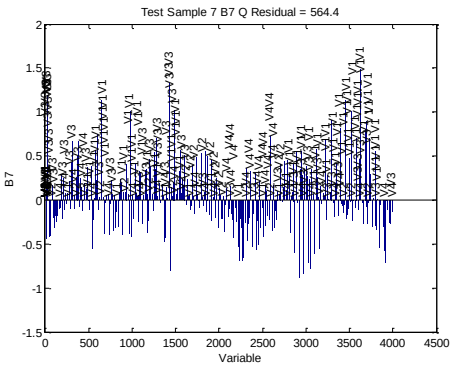

Figure 8-48: P5 B7 Q-residual contribution plot

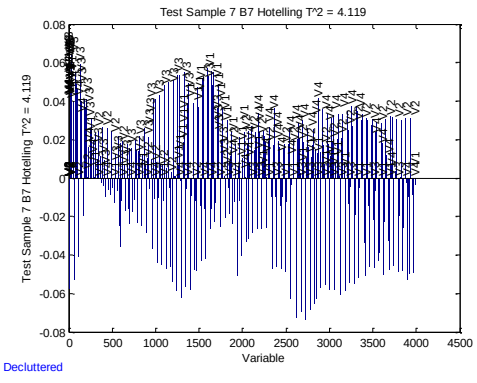

Figure 8-49: P5 B7 Hotelling's T2 contribution plot

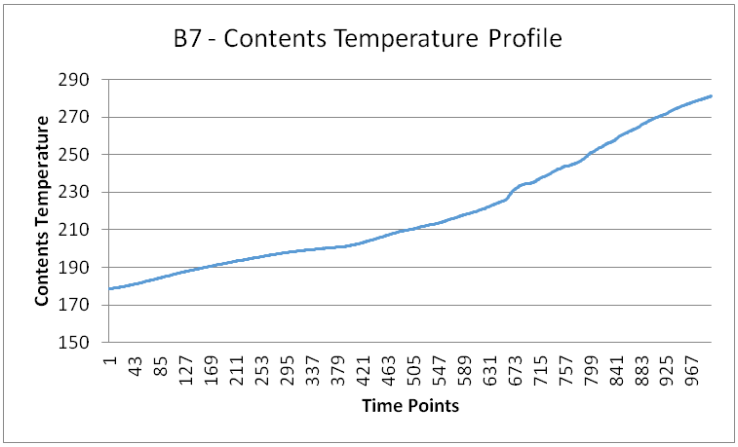

Figure 8-50: B7 Contents Temperature Profile

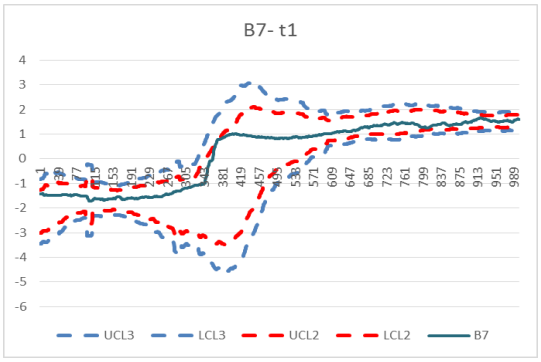

Figure 8-51: B7 Variable-Wise unfolding approach unfolded Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

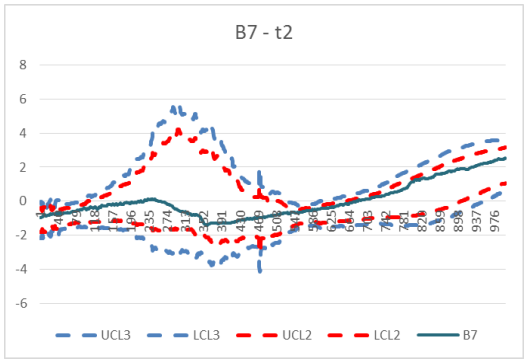

Figure 8-52: B7 Variable-Wise unfolding approach unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

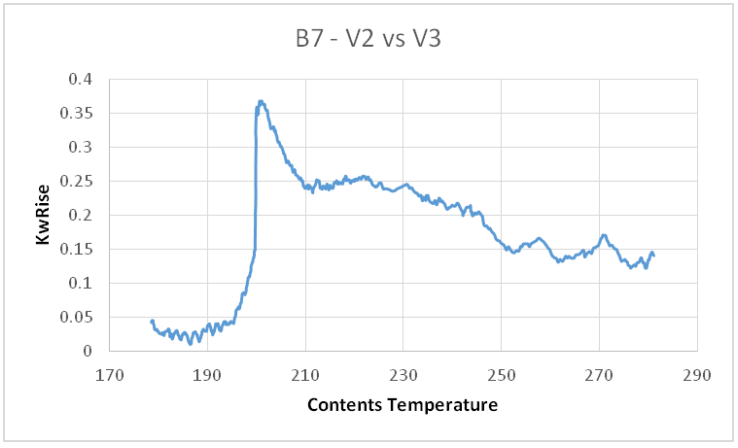

Figure 8-53: B7 Contents Temperature profile vs KwRise

Batch 8

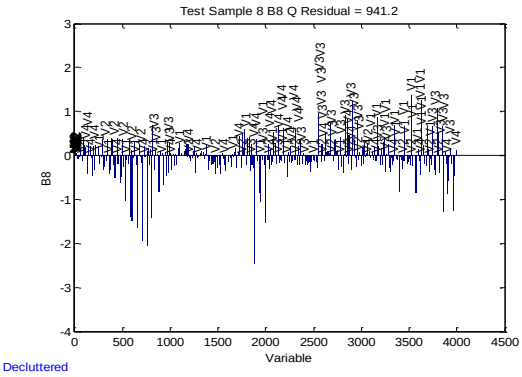

Figure 8-54: B8 Q-residual contribution plot

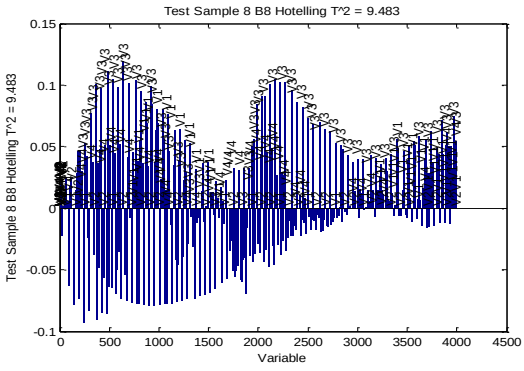

Figure 8-55: B8 Hotelling's T2 contribution plot

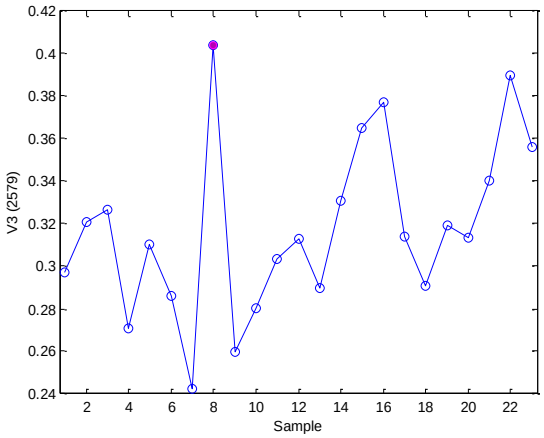

Figure 8-56: B8 KwRise contribution at time point 2579

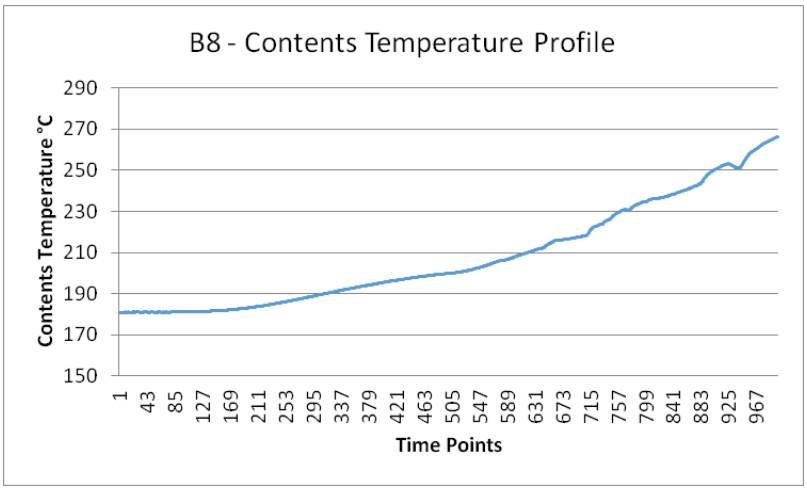

Figure 8-57: B8 Contents Temperature Profile

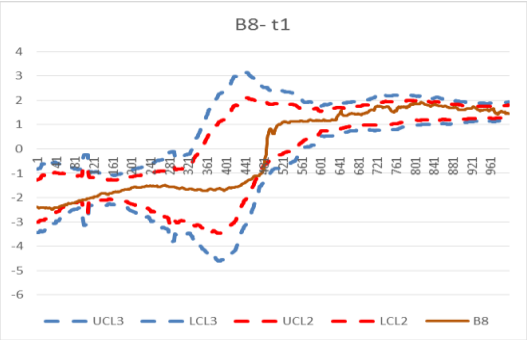

Figure 8-58: B8 Variable-Wise unfolding approach unfolded Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

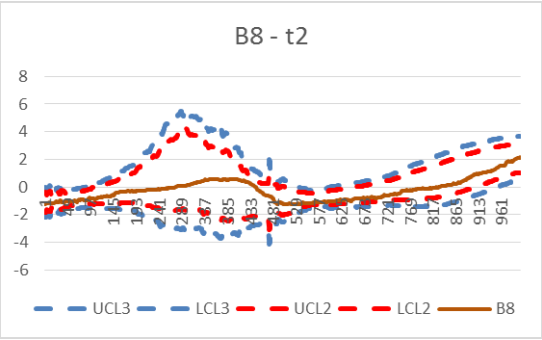

Figure 8-59: B8 Variable-Wise unfolding approach unfolded Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

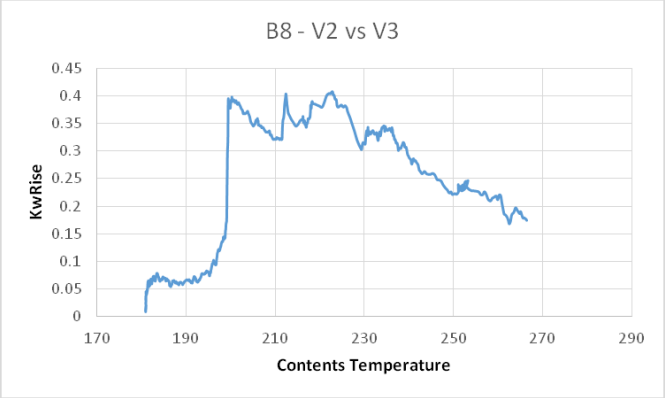

Figure 8-60: B8 Contents Temperature profile vs KwRise

Batch 9

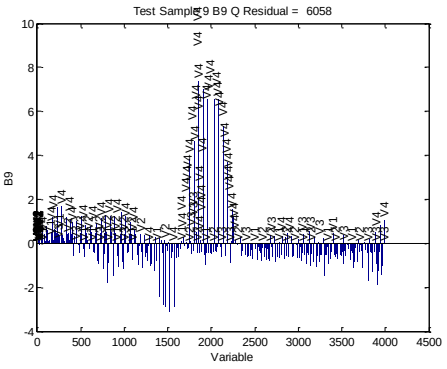

Figure 8-61: B9 Q-residual contribution

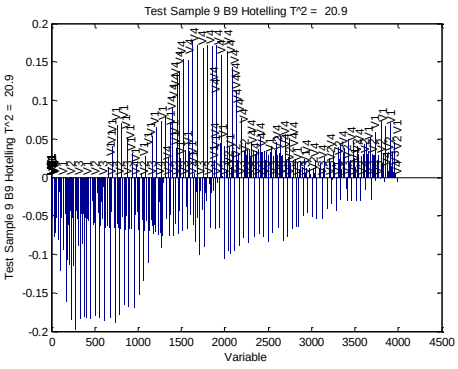

Figure 8-62: B9 Hotelling's T2 contribution plot

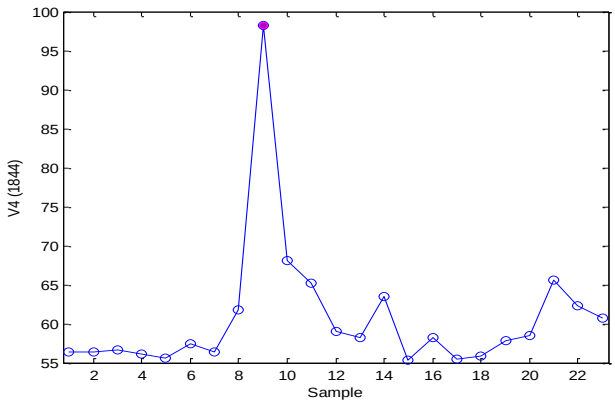

Figure 8-63: B9 Level at time point 1844

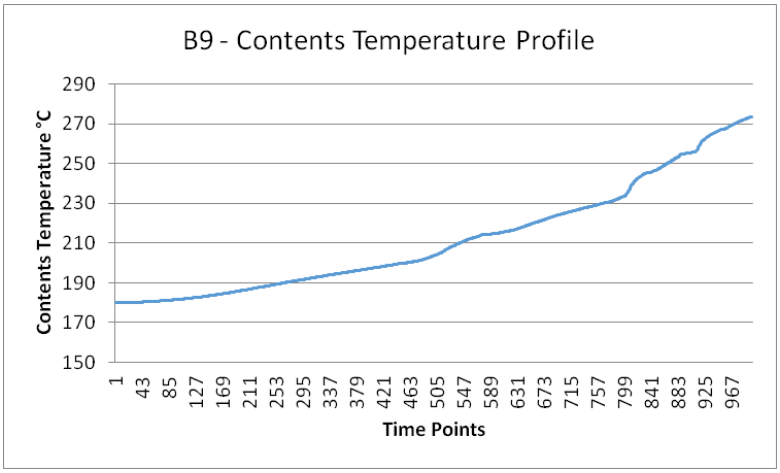

Figure 8-64: B9 Contents Temperature profile

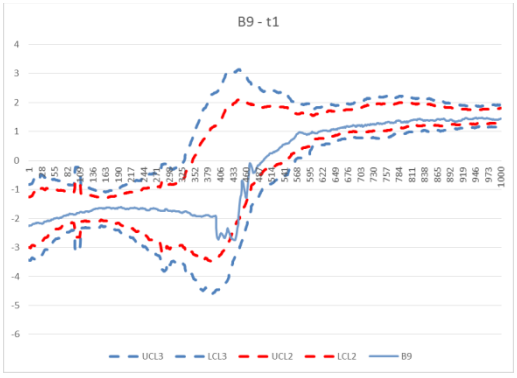

Figure 8-65: B9 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

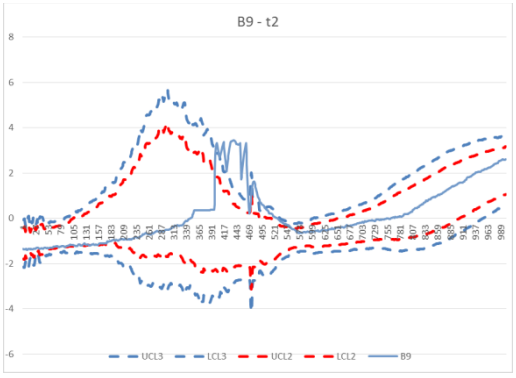

Figure 8-66: B9 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

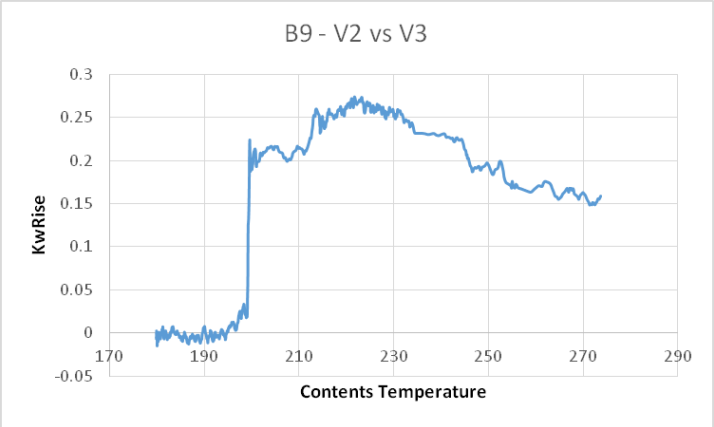

Figure 8-67: B9 Contents temperature vs KwRise

Batch 10

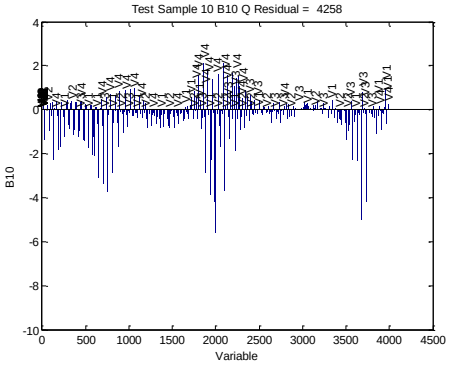

Figure 8-68: B10 Q-residual contribution plot

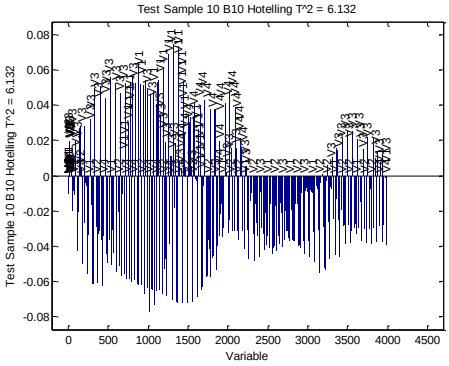

Figure 8-69: B10 Hotelling's T2 contribution plot

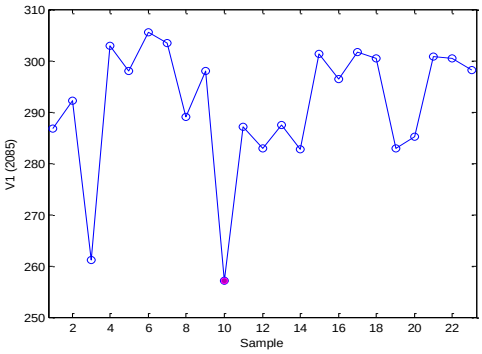

Figure 8-70: B10 V1 value at time point 2085

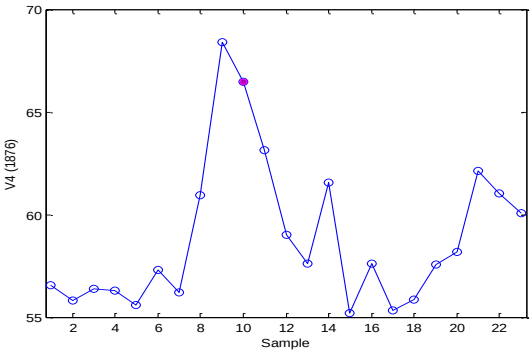

Figure 8-71: B10 V4 value at time point1876

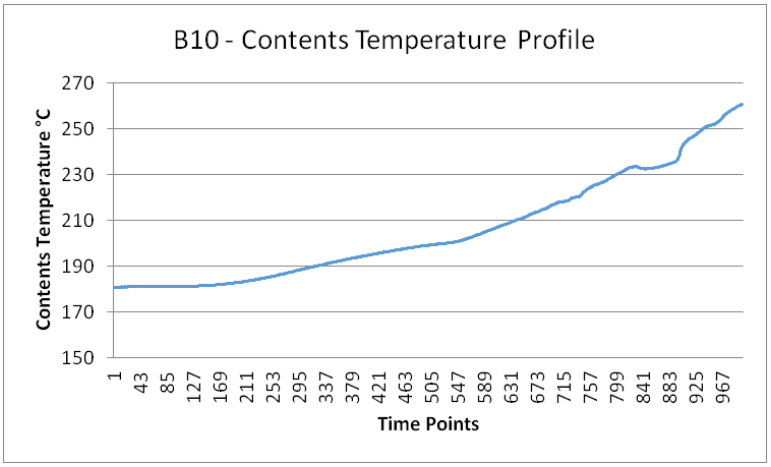

Figure 8-72: B10 Contents Temperature Profile

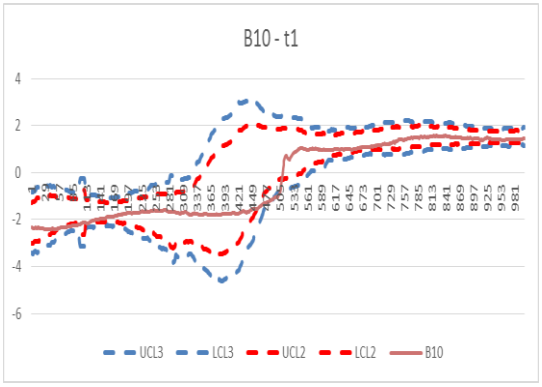

Figure 8-73: B10 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

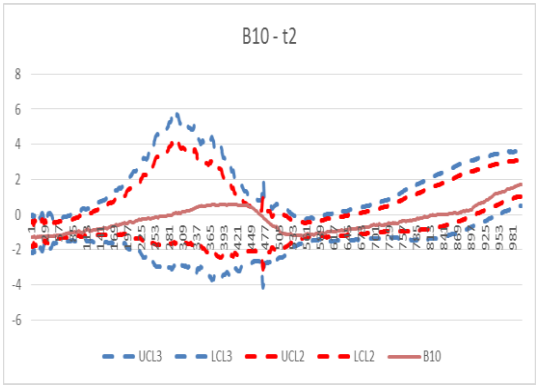

Figure 8-74: B10 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

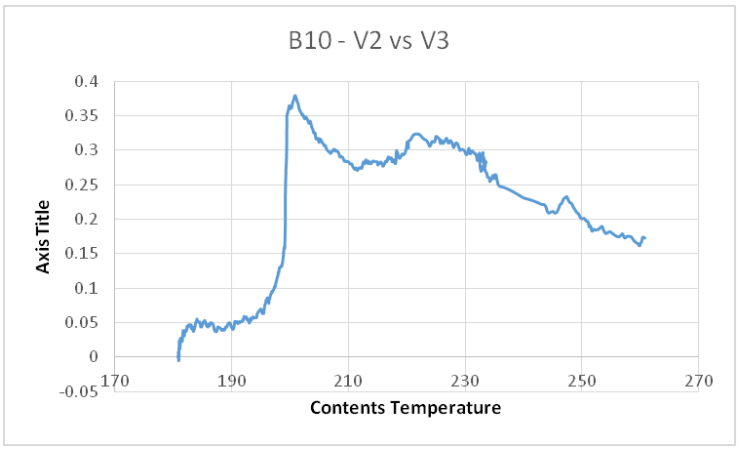

Figure 8-75: B10 Contents Temperature profile vs KwRise

Batch 11

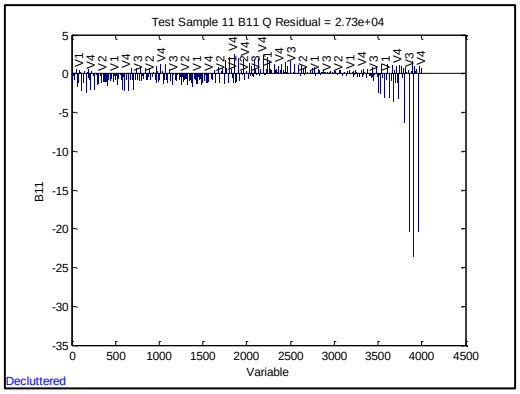

Figure 8-76: B11 Q-residual contribution plot

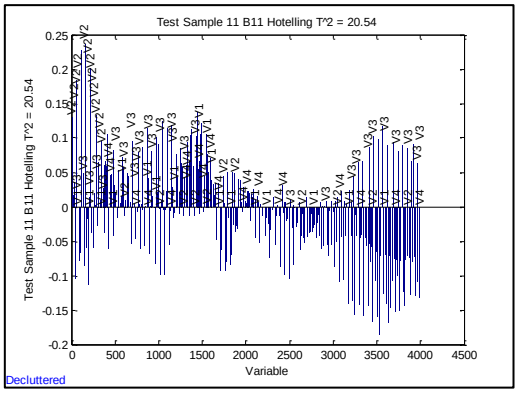

Figure 8-77: B11 Hotelling's T2 Contribution plot

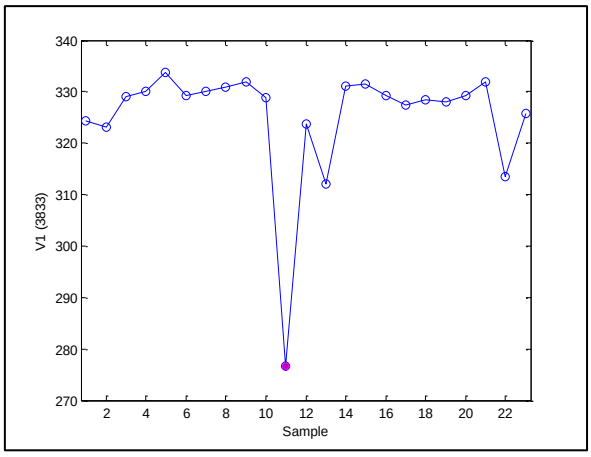

Figure 8-78: B11 V1 value at time point 3833

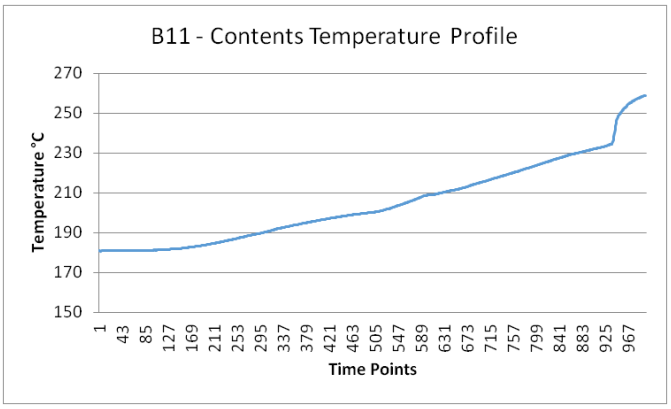

Figure 8-79: B11 Contents Temperature profile

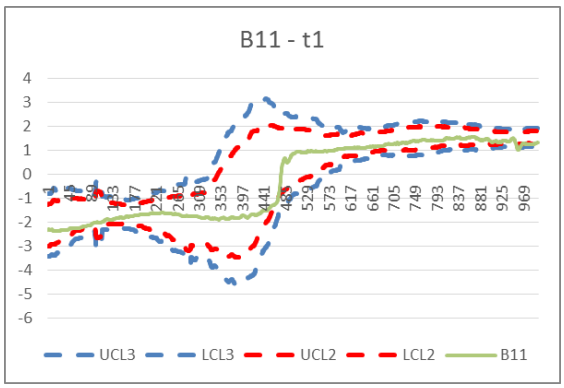

Figure 8-80: B11 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

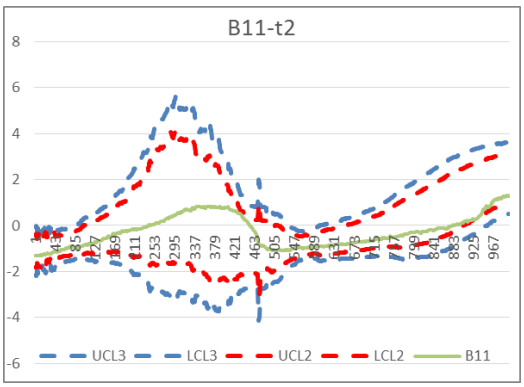

Figure 8-81: B11 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

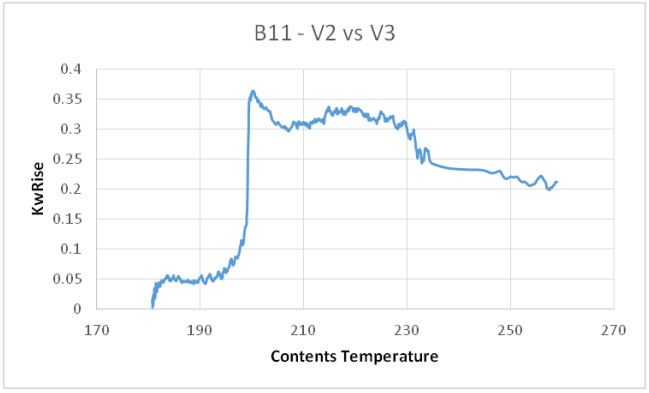

Figure 8-82: B11 Contents Temperature profile vs KwRise

Batch 12

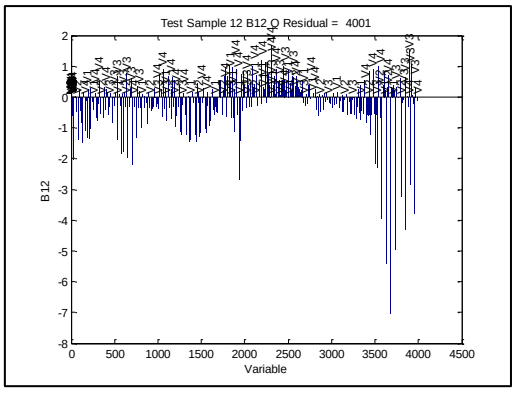

Figure 8-83: B12 Q-residual contribution plot

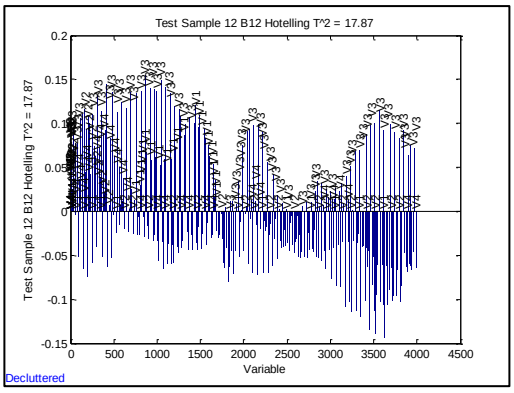

Figure 8-84: B12 Hotelling's T2 contribution plot

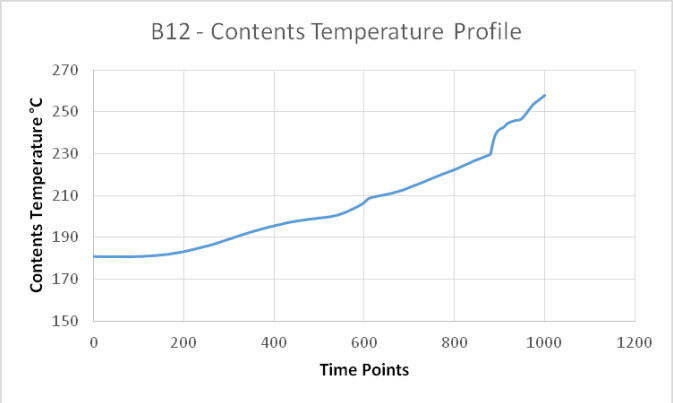

Figure 8-85: B12 Contents Temperature Profile

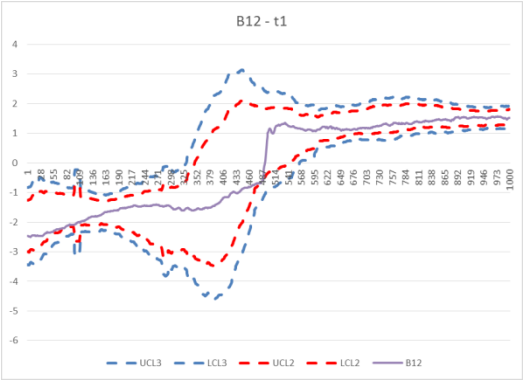

Figure 8-86: B12 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

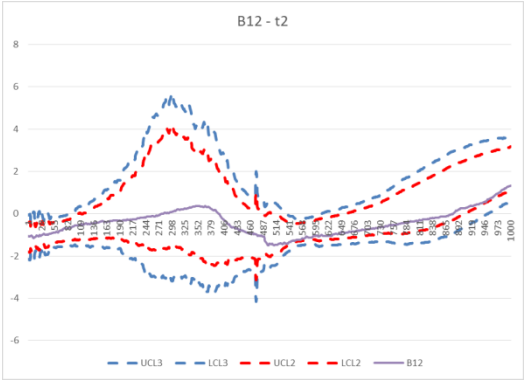

Figure 8-87: Variable-Wise unfolding Approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

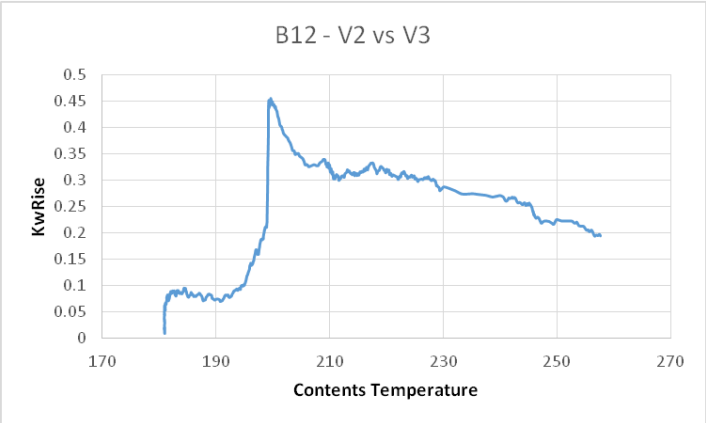

Figure 8-88: Contents temperature profile vs KwRise

Batch 13

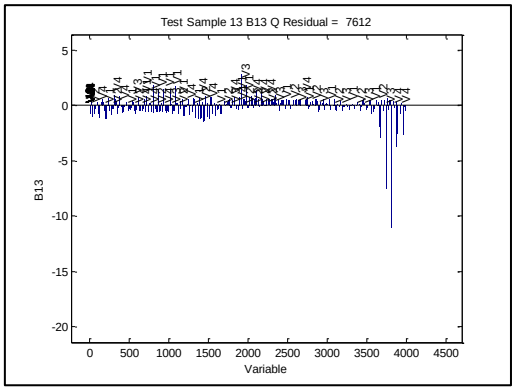

Figure 8-89: B13 Q-residual contribution plot

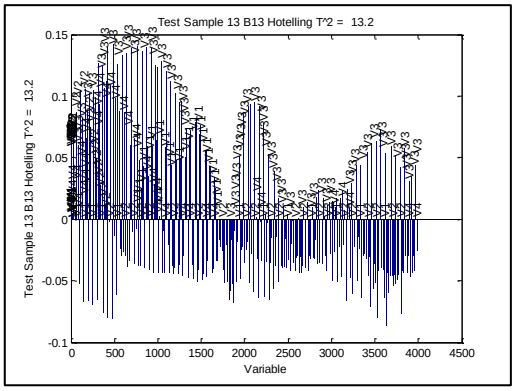

Figure 8-90: B13 Hotelling's T2 contribution plot

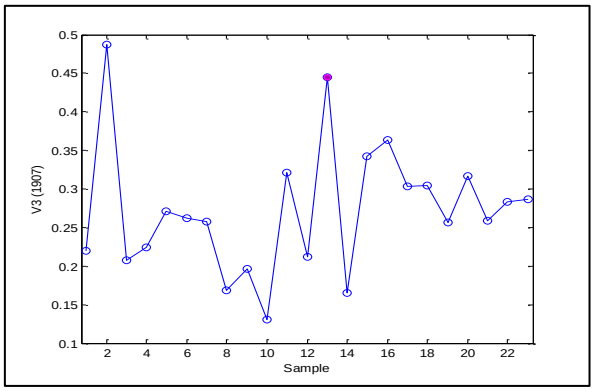

Figure 8-91: B13 KwRise value at time point 1907

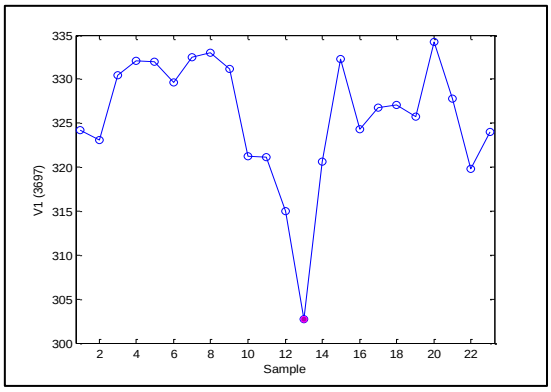

Figure 8-92: B13 Jacket temperature value at time point 3997

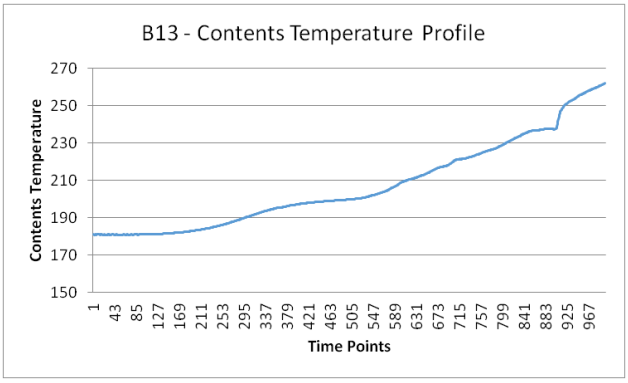

Figure 8-93: B13 Contents Temperature Profile

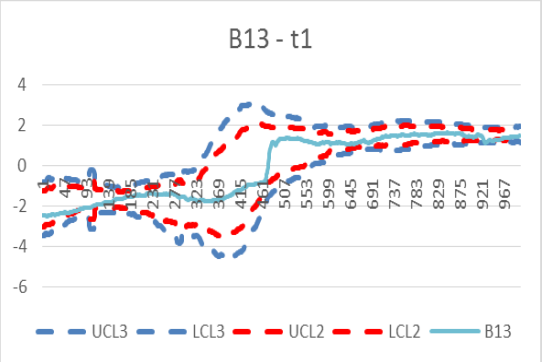

Figure 8-94: B13 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

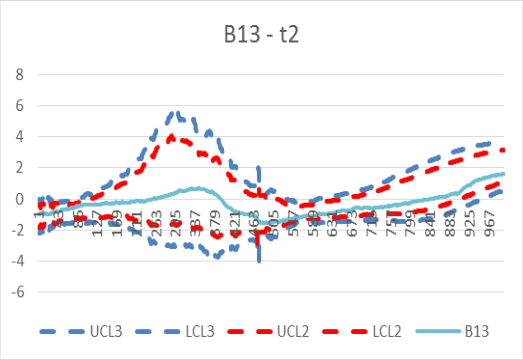

Figure 8-95: B13 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

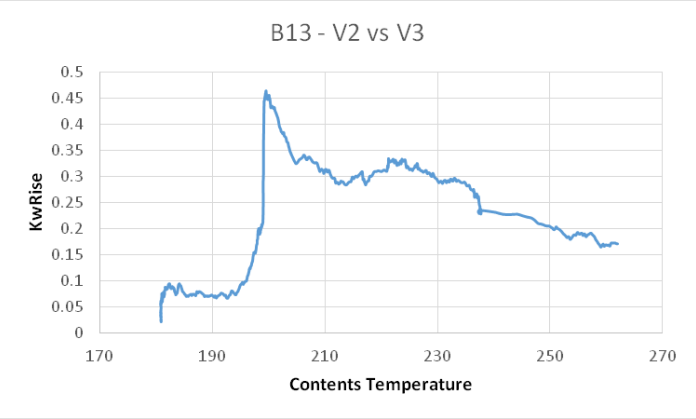

Figure 8-96: B13 Contents Temperature profile vs KwRise

Batch 14

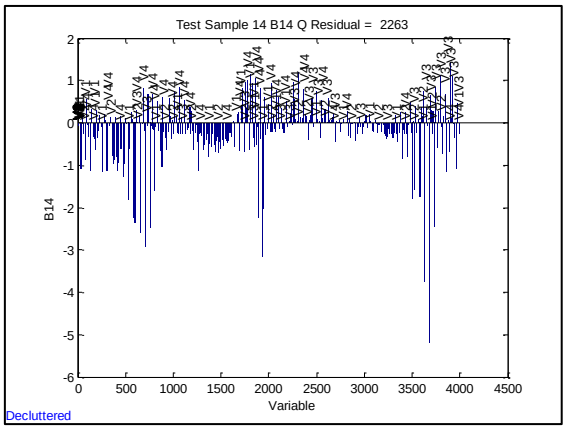

Figure 8-97: B14 Q-residual contribution plot

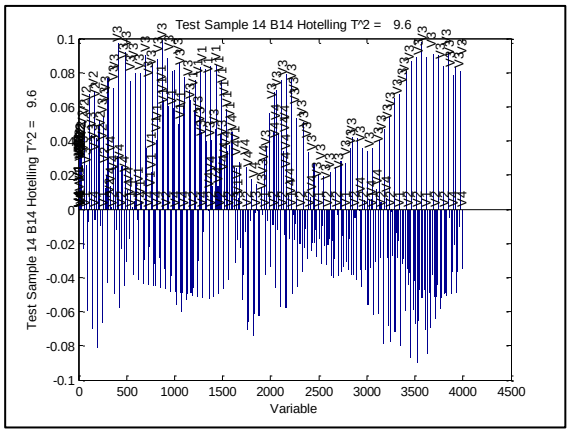

Figure 8-98: B14 Hotelling's T2 contribution plot

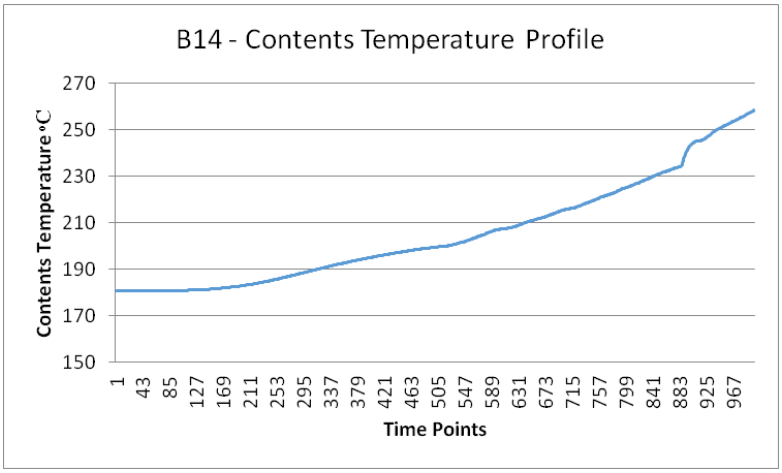

Figure 8-99: B14 Contents Temperature Profile

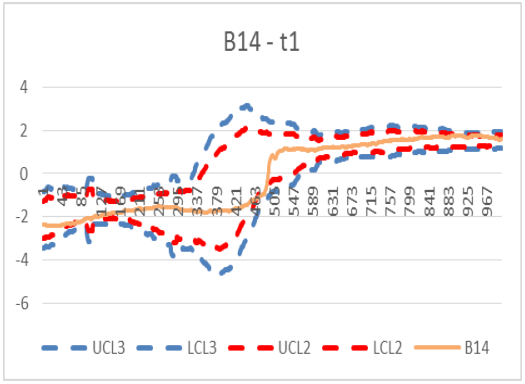

Figure 8-100: B14 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

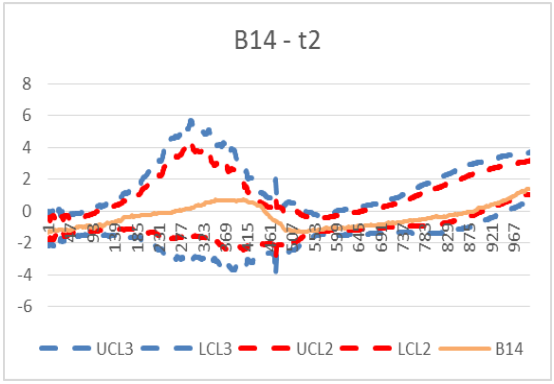

Figure 8-101: B14 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

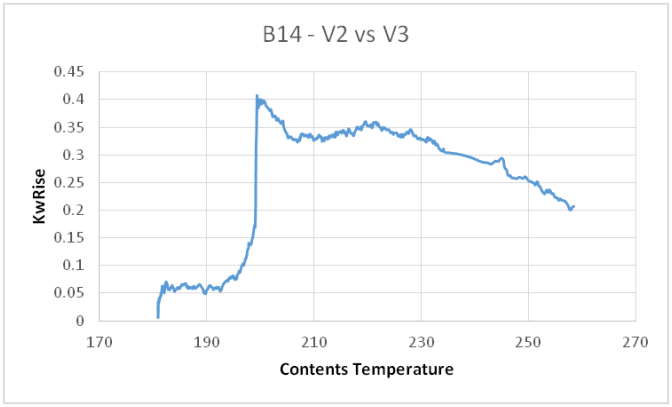

Figure 8-102: B14 Contents temperature profile vs KwRise

Batch 15

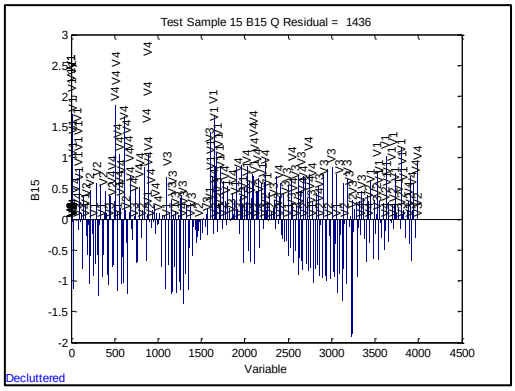

Figure 8-103: B15 Q-residual contribution plot

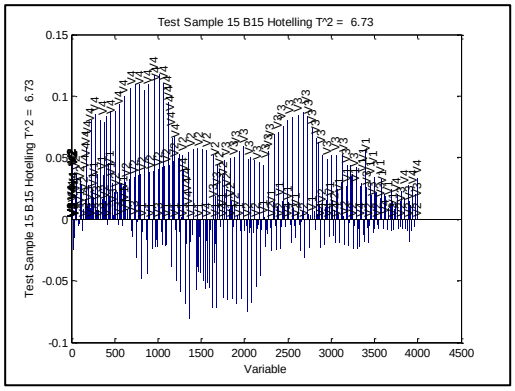

Figure 8-104: B15 Hotelling's T2 contribution plot

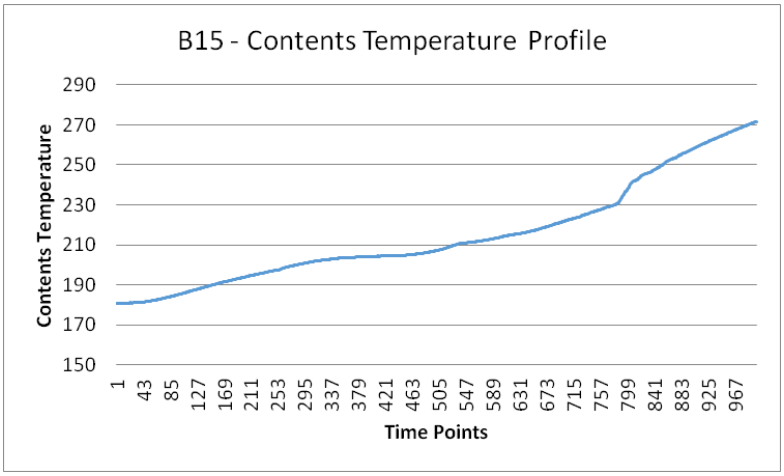

Figure 8-105: B15 Contents Temperature Profile

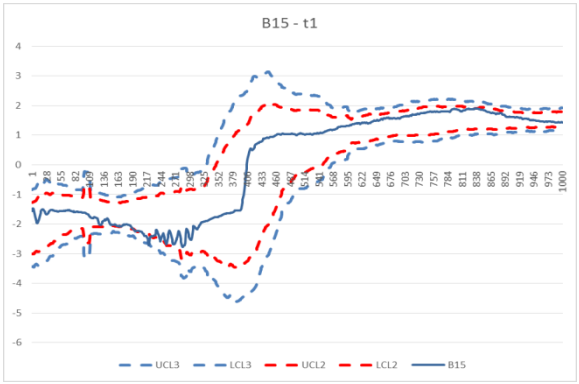

Figure 8-106: B15 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

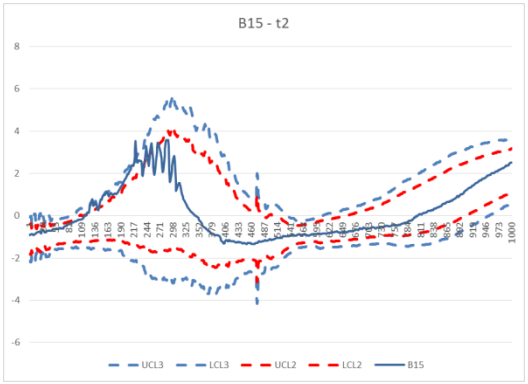

Figure 8-107: Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

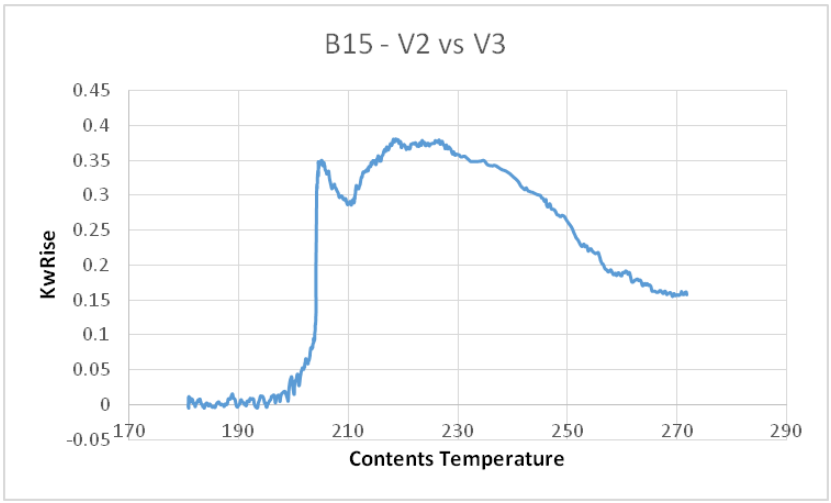

Figure 8-108: B15 Contents Temperature vs KwRise

Batch 16

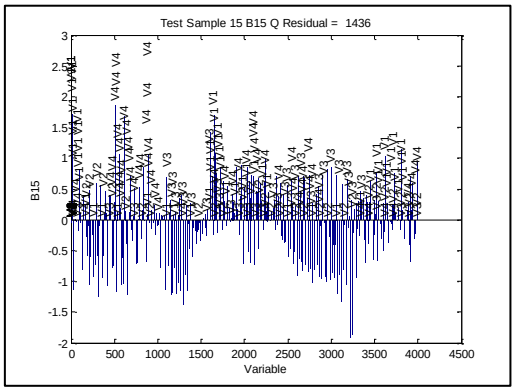

Figure 8-109: B16 Q-residuals contribution plot

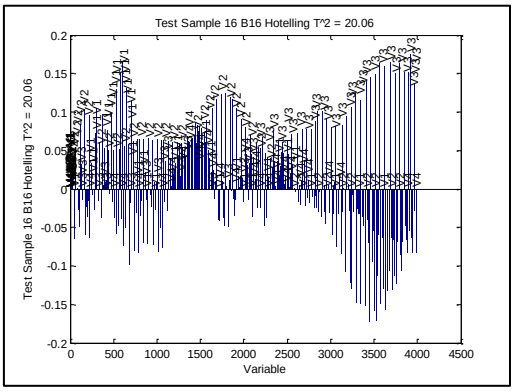

Figure 8-110: B16 Hotelling's T2 contribution plot

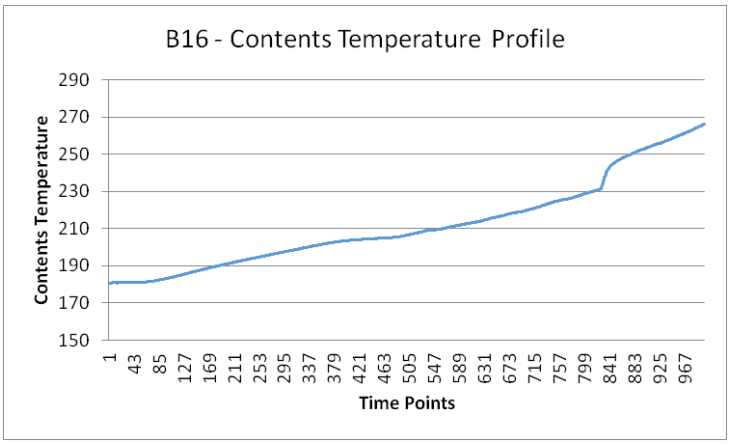

Figure 8-111: Contents Temperature profile

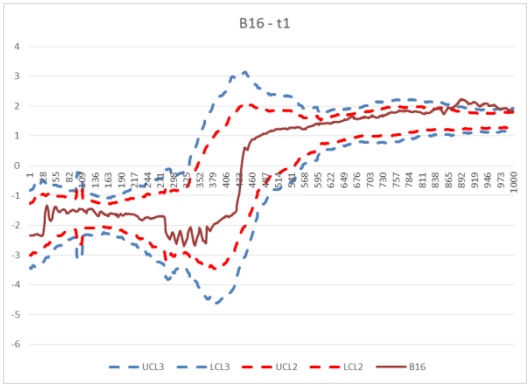

Figure 8-112: B16 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

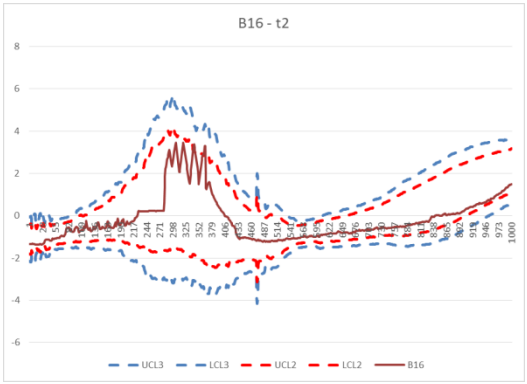

Figure 8-113: B16 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

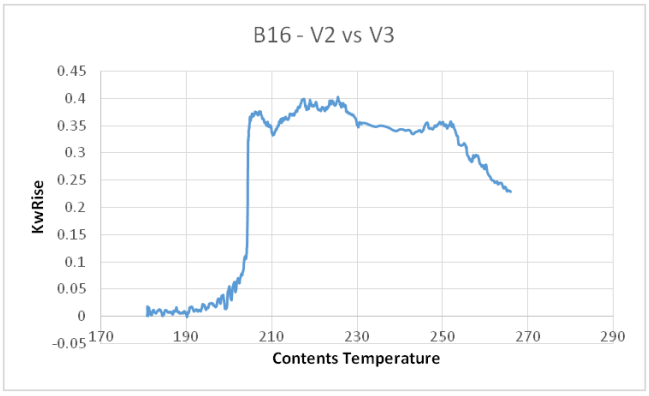

Figure 8-114: B16 Contents temperature vs KwRise

Batch 17

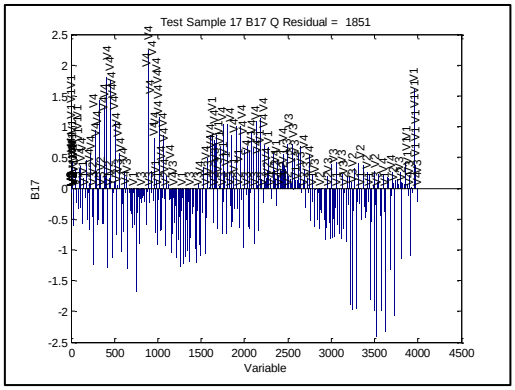

Figure 8-115: B17 Q-residual contribution plot

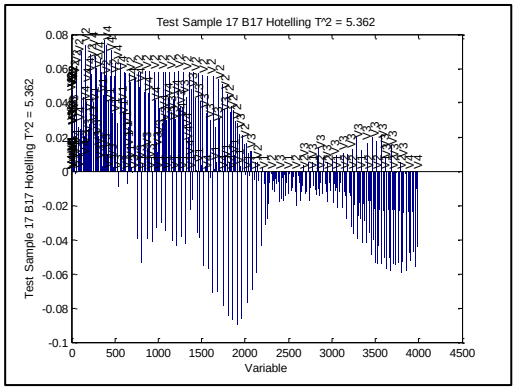

Figure 8-116: Hotelling's T2 contribution plot

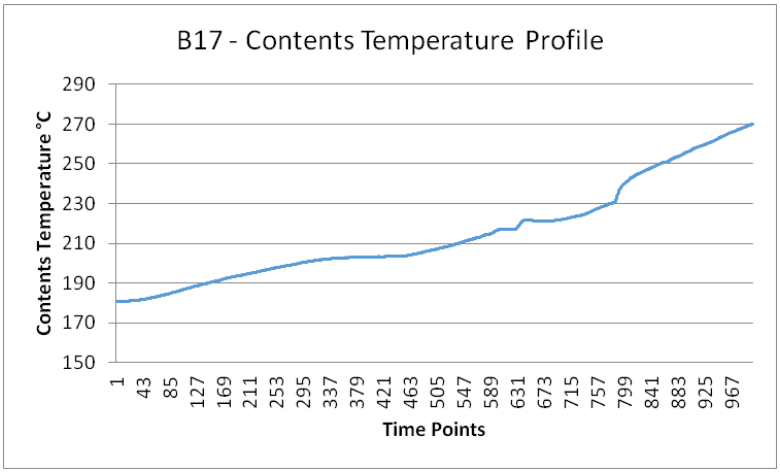

Figure 8-117:B17 Contents Temperature Profile

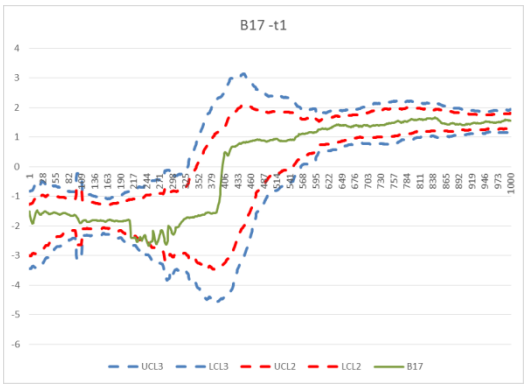

Figure 8-118: B17 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

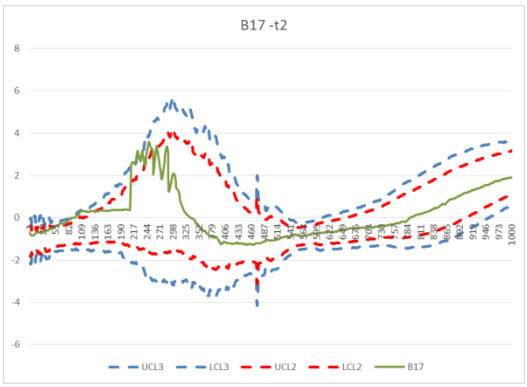

Figure 8-119: B17 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

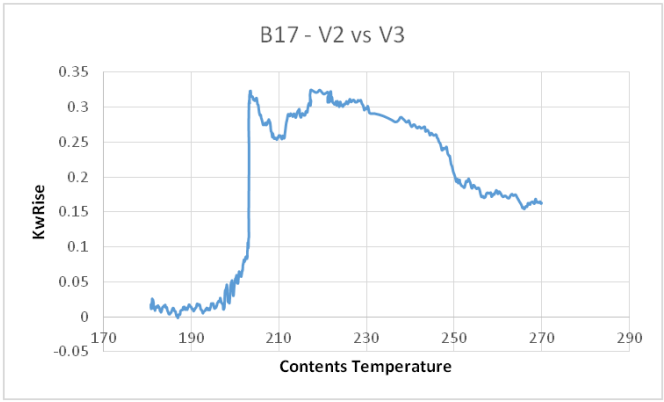

Figure 8-120: B17 Contents temperature vs KwRise

Batch 18

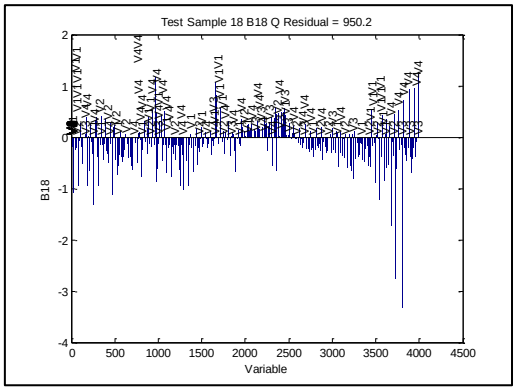

Figure 8-121: B18 Q-residual contribution plot

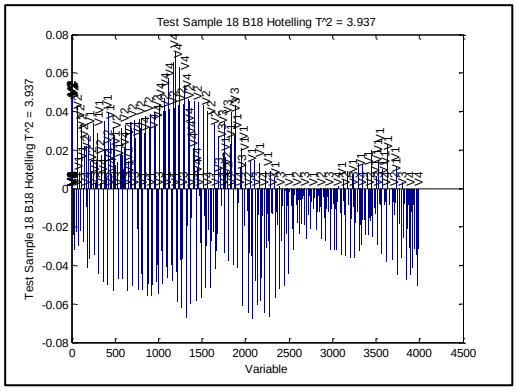

Figure 8-122: B18 Hotelling:s T2 contribution plot

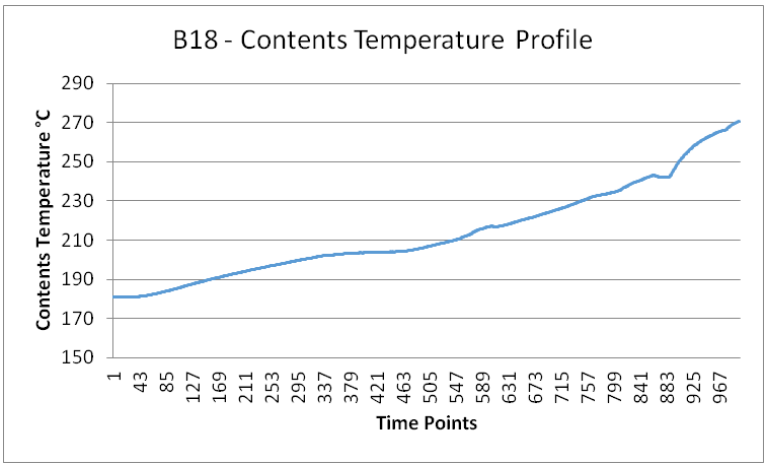

Figure 8-123: B18 Contents Temperature Profile

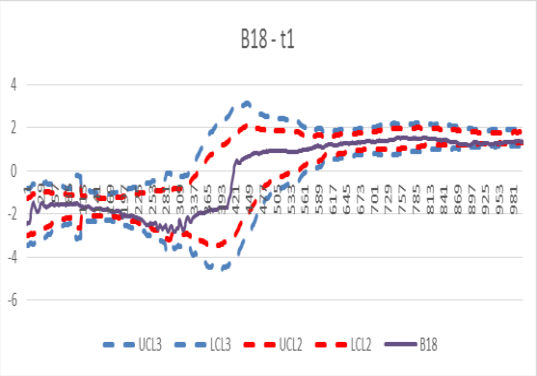

Figure 8-124: B18 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

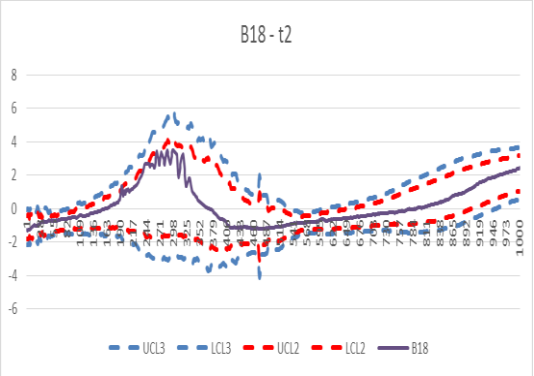

Figure 8-125: B18 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

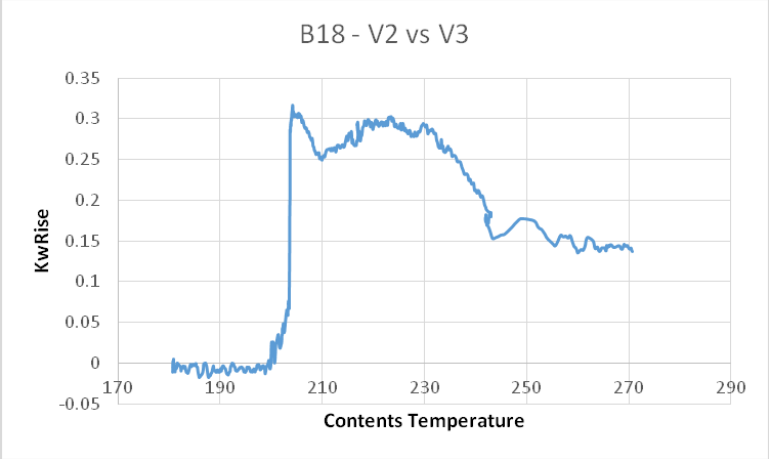

Figure 8-126: B18 Contents Temperature vs KwRise

Batch 19

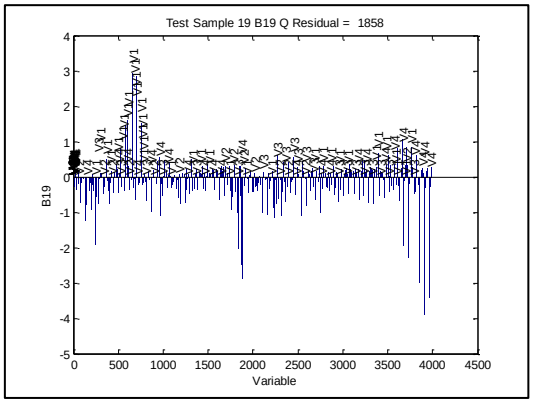

Figure 8-127: B19 Q-residual contribution plot

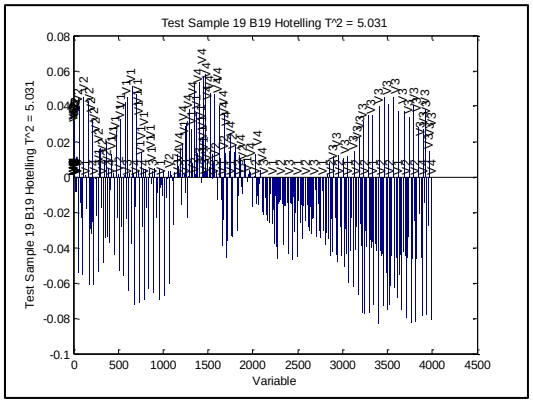

Figure 8-128: B19 Hotelling's T2 contribution plot

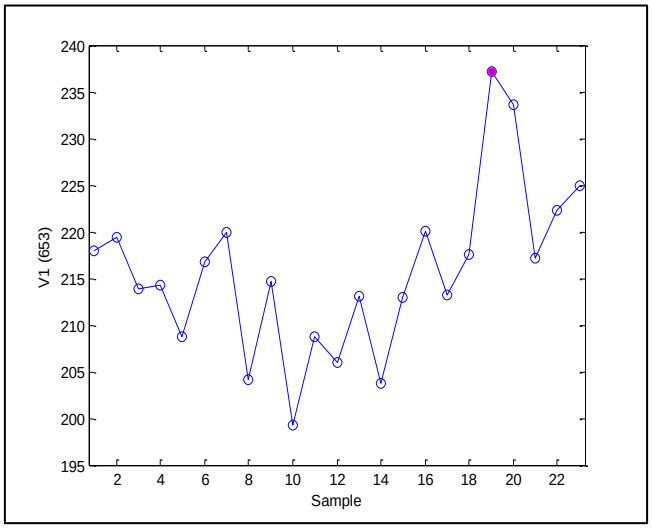

Figure 8-129: B19 Jacket temperature value at time point 853

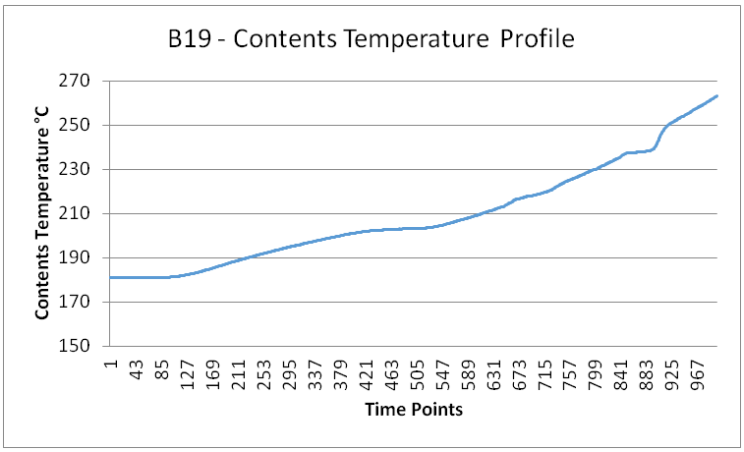

Figure 8-130: B19 Contents temperature profile

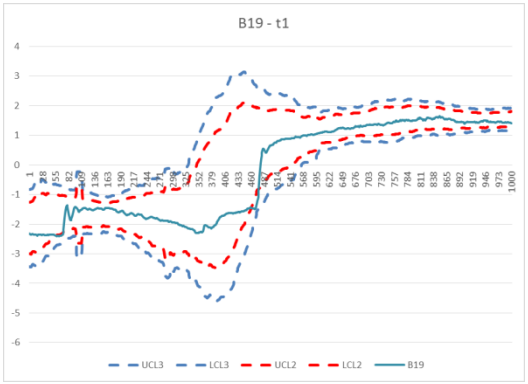

Figure 8-131: B19 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

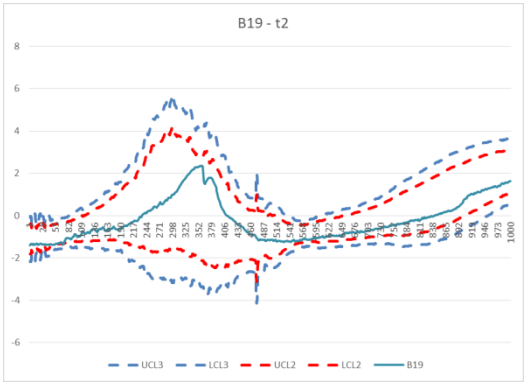

Figure 8-132: B19 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

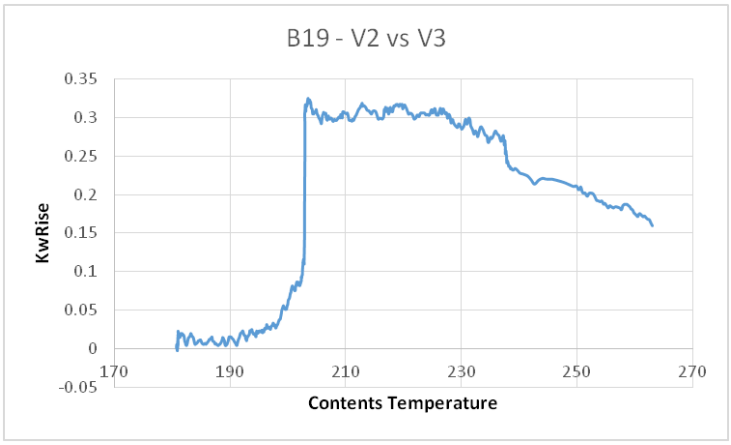

Figure 8-133: B19 Contents Temperature vs KwRise

Batch 20

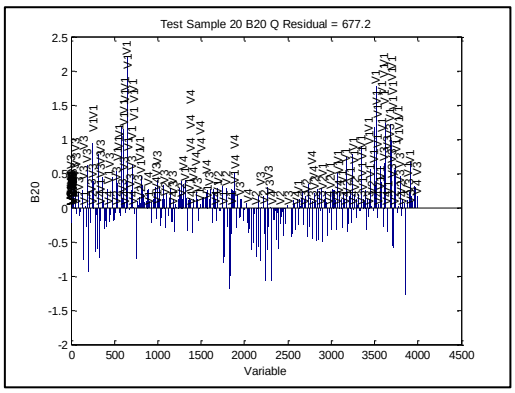

Figure 8-134: B20 Q-residual contribution plot

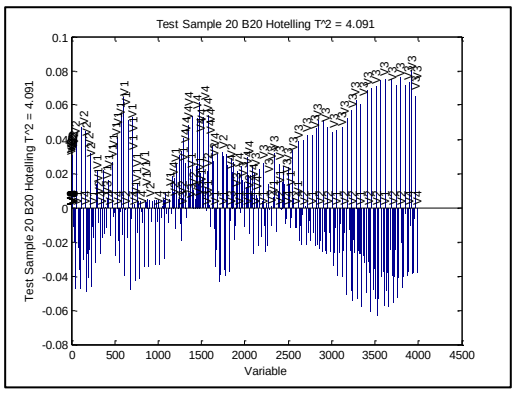

Figure 8-135: B20 Hotelling's T2 contribution plot

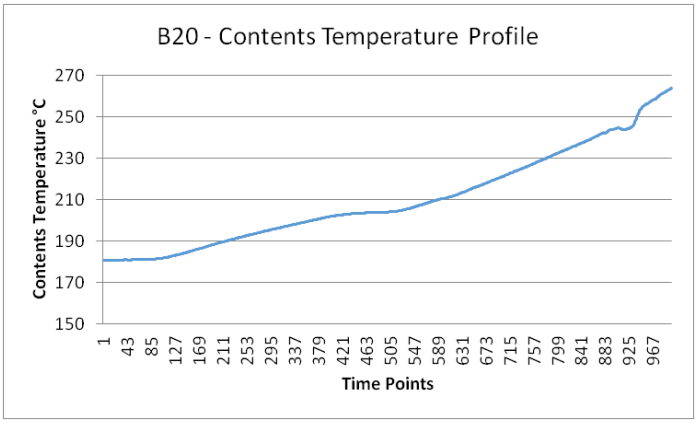

Figure 8-136: B20 Contents Temperature Profile

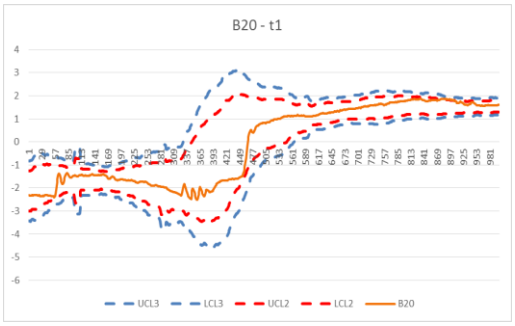

Figure 8-137: B20 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

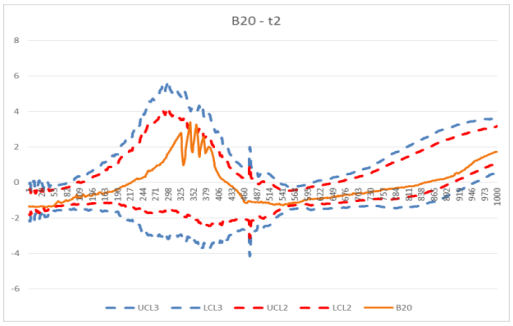

Figure 8-138: Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

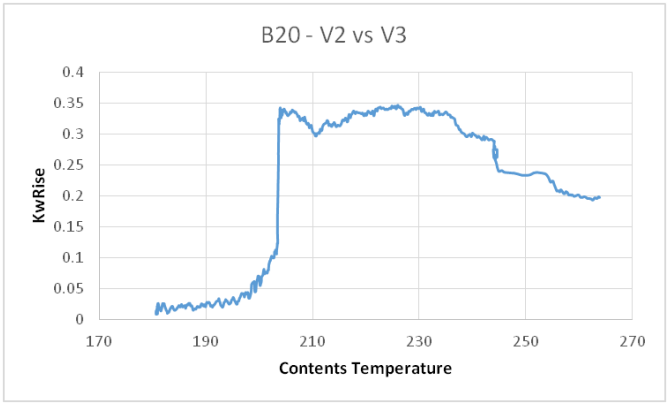

Figure 8-139: B20 Contents Temperature vs KwRise

Batch 21

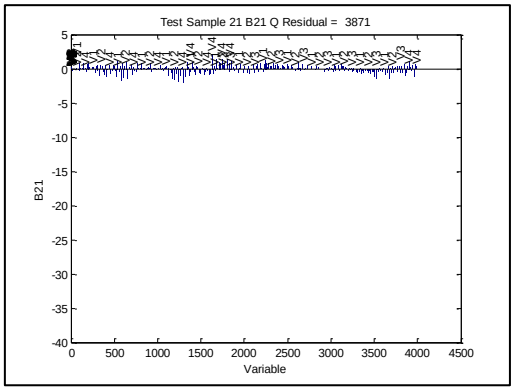

Figure 8-140: B21 Q-residual contribution plot

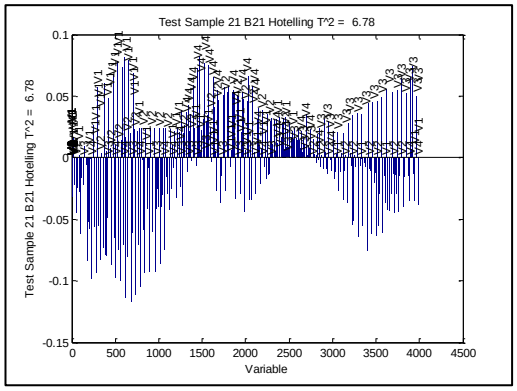

Figure 8-141: B21 Hotelling's T2 contribution plot

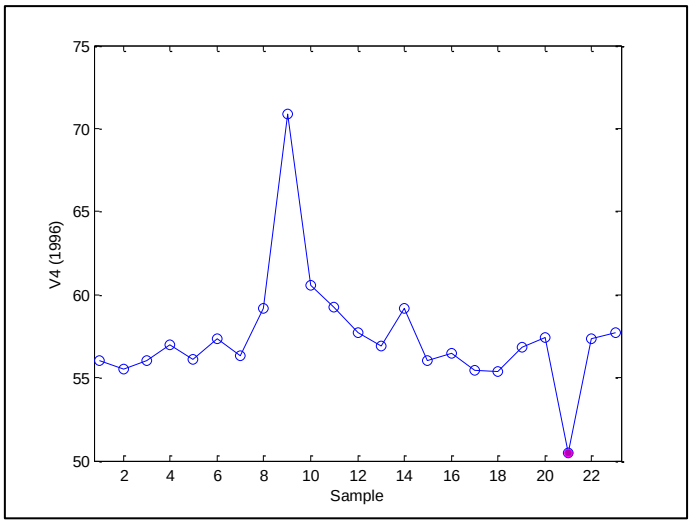

Figure 8-142: B21 Level value at time point

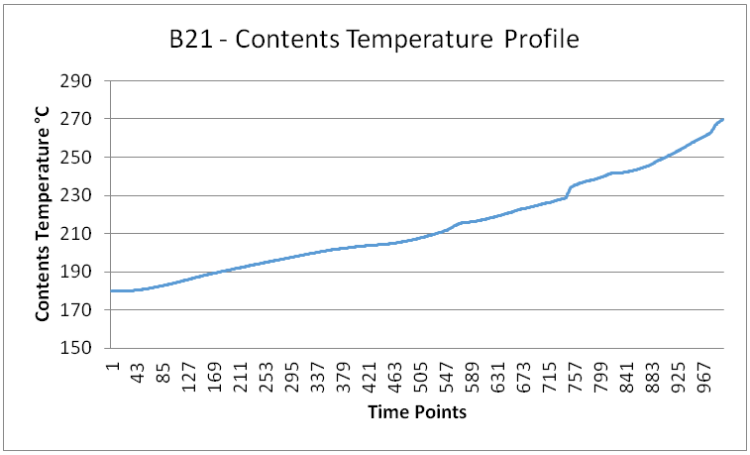

Figure 8-143: B21 Contents Temperature Profile

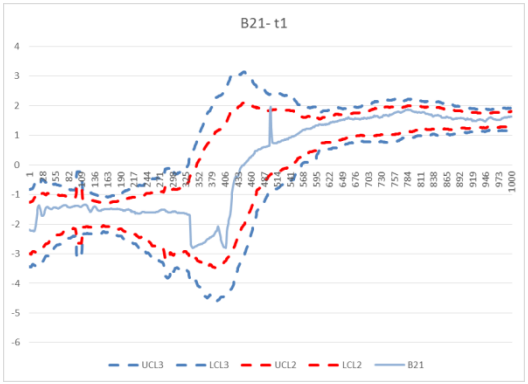

Figure 8-144: B21 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

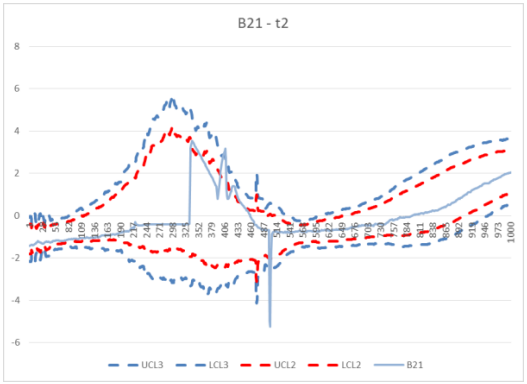

Figure 8-145: B21 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

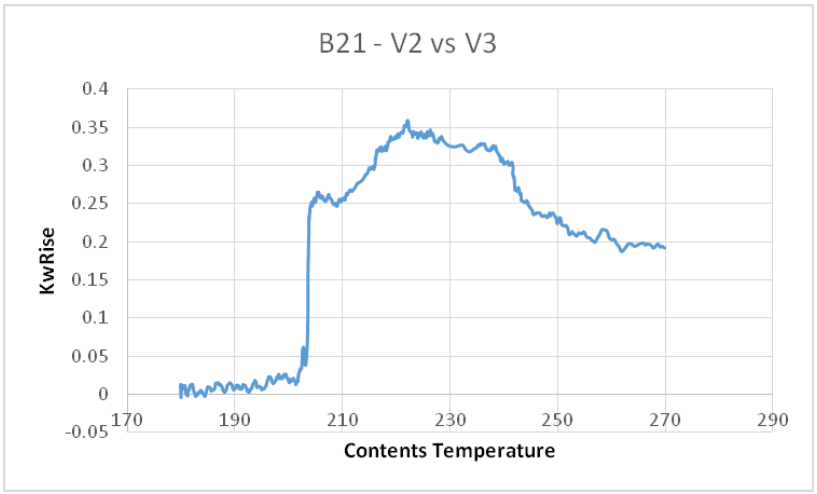

Figure 8-146: B21 Contents Temperature vs KwRise

Batch 22

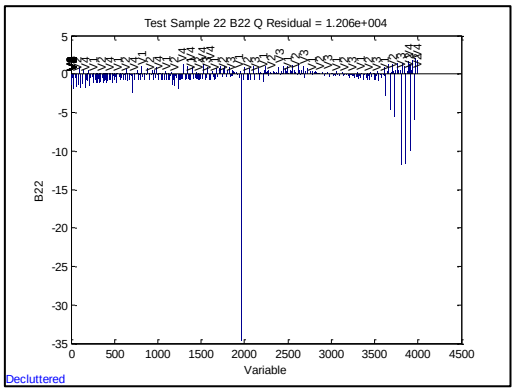

Figure 8-147: B22 Q-residual contribution plot

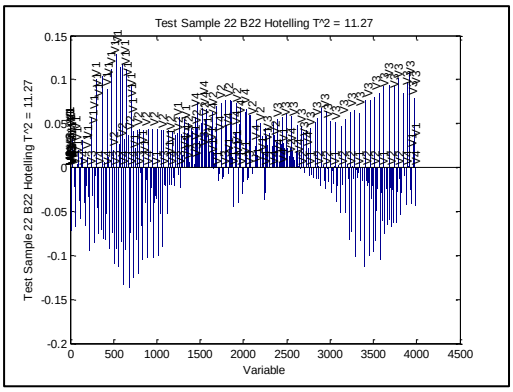

Figure 8-148: B22 Hotelling's T2 contribution plot

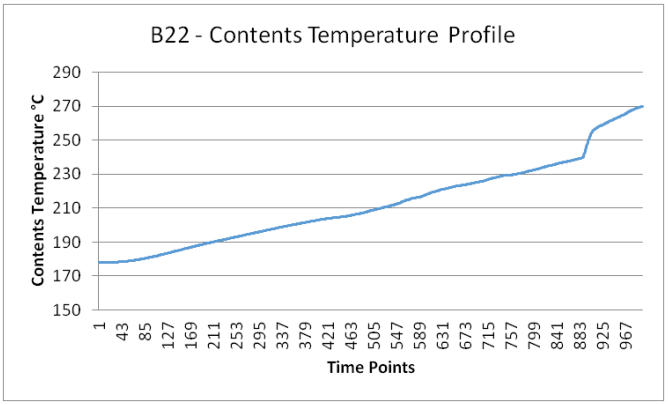

Figure 8-149: B22 Contents Temperature Profile

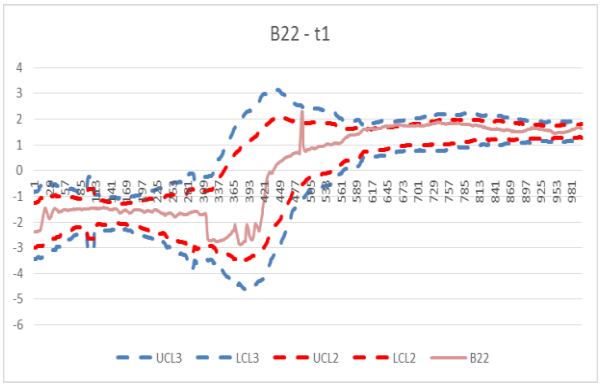

Figure 8-150: B22 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

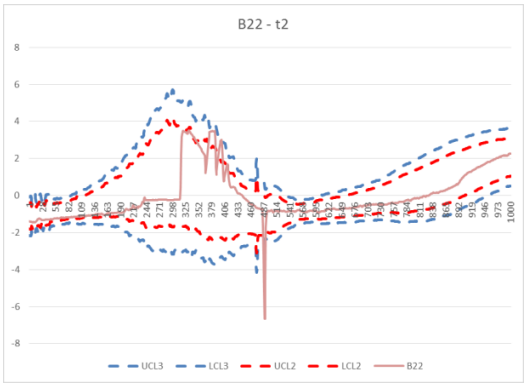

Figure 8-151: B22 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

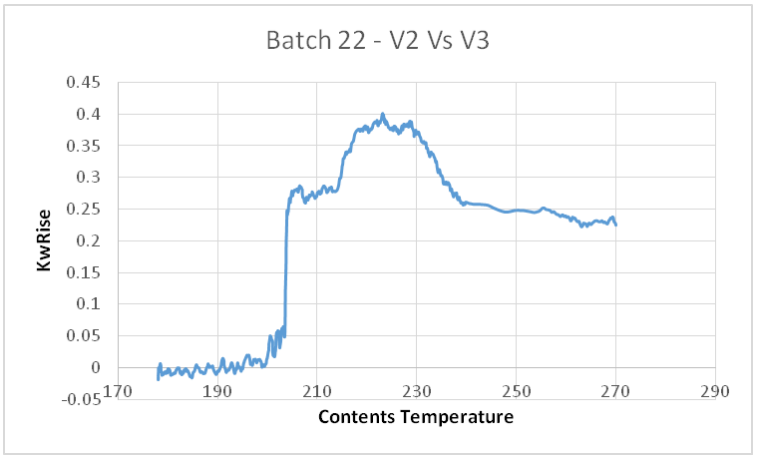

Figure 8-152: B22 Contents temperature vs KwRise

Batch 23

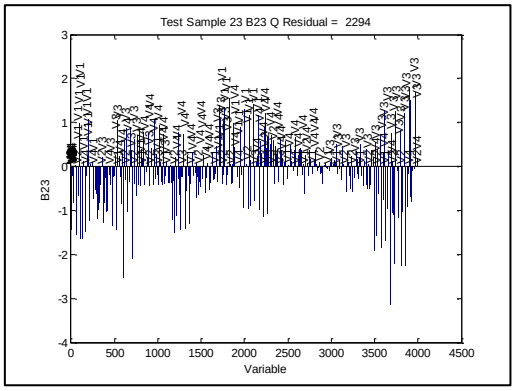

Figure 8-153: B23 Q-residual contribution plot

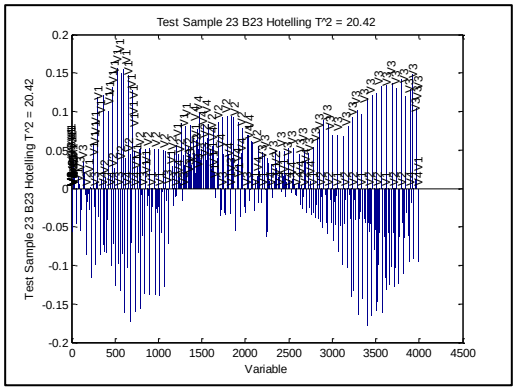

Figure 8-154: B23 Hotelling's T2 contribution plot

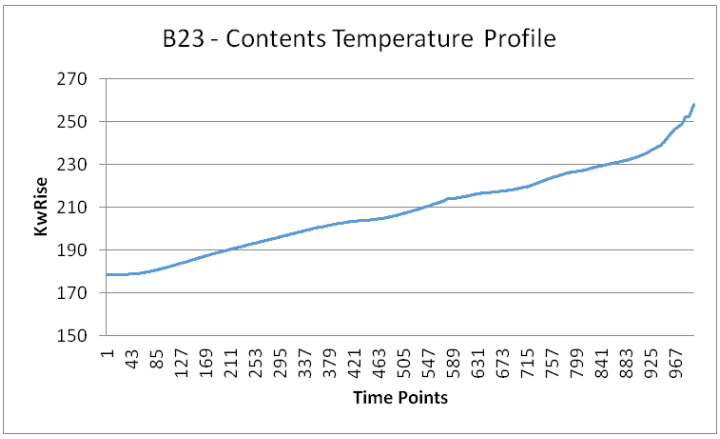

Figure 8-155: B23 Contents Temperature profile

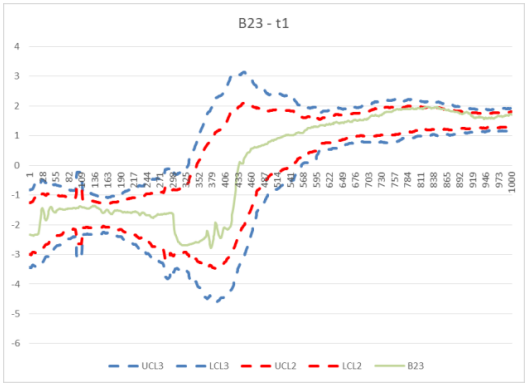

Figure 8-156: B23 Variable-Wise unfolding approach Scores 1 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

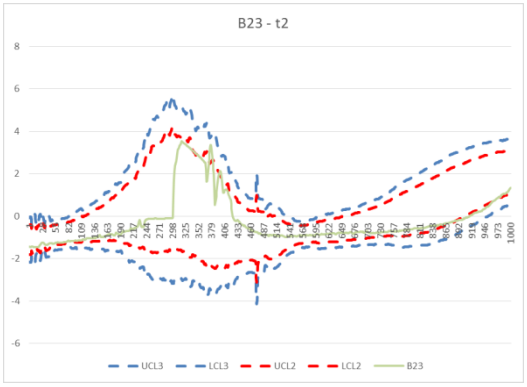

Figure 8-157: B23 Variable-Wise unfolding approach Scores 2 Control limits: 1) UCL3 and LCL3 are upper control limits of + and – 3 standard deviation 2) UCL2 and LCL2 are the lower control limits of + and – 2 standard deviation

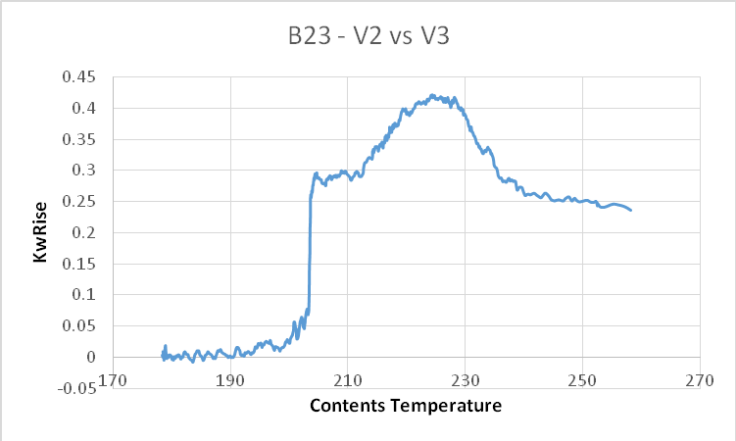

Figure 8-158: B23 Contents Temperature vs KwRise