University of Rhode Island

DigitalCommons@URI

2013

Applying QbD and Pat in Biological Manufacturing for "Continued Process Verification"

Prakash Bedre
University of Rhode Island, pbedre@amgen.com

APPLYING QbD AND PAT IN

BIOLOGICAL MANUFACTURING

FOR

"CONTINUED PROCESS VERIFICATION"

BY

PRAKASH BEDRE

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOMEDICAL AND PHARMACEUTICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2013

DOCTOR OF PHILOSOPHY DISSERTATION

OF

PRAKASH BEDRE

APPROVED:

Dissertation Committee:

Major Professor   Clinton Chichester

                       Cenk Undey

                       Manbir Sodhi

                       Navindra Seeram

                    Nasser H. Zawia
            DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2013

ABSTRACT

The objective of this research topic is to show QbD and PAT tools such as multivariate analysis can perform "Continued Process Verification by using a Real-Time Multivariate Process Monitoring (RT-MSPM) system. There are not one but many challenges The pharmaceutical and bio-pharmaceutical manufacturers are facing multiple challenges such as changing regulatory requirements, healthcare reforms, economic pressure and availability of advance manufacturing technology to make better quality products at reduced costs.

Due to the recent technological developments, significant opportunities exist for improving pharmaceutical development, manufacturing and quality assurance through innovation in product and process development, process analysis, and process controls. The latest FDA guidelines such as QbD, PAT and the 2011 process validation have opened the doors for "Real-Time Process Monitoring" concepts for "Continued Process Verification".

The regulatory agencies have taken the initiative by providing guidelines in last ten years such as, Pharmaceutical cGMPs for the 21st Century - A Risk Based Approach, Final Report in September 2004 [1], Guidance for Industry: PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance [2], Guidance for Industry Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations [3], Internal Commerce for Harmonization (ICH) - guidelines [4, 5, 6, and 7], QbD, a perspective from the "Office of Biotechnology Products" (OBP) [8] and lastly, Guidance for Industry Process Validation General Principles and practices utilizing three stages during Process Validation [9].

The objective of agency is to ensure that the most up-to-date concepts of risk management and quality systems approaches are incorporated into the manufacturing.

The application of multivariate statistical models for process monitoring can provide information on the challenges that are routinely encountered by drug manufacturers and process can be monitored in real-time to achieve continued process verification (CPV). The outcome of the study is intended to become a benchmark for biological manufacturers who are interested in applying the "PAT tools" for existing legacy products or any new manufacturing process to address challenges [10, 11] such as raw material variation and control of process variability, identifying and monitoring of relevant process parameters in the operating space, RT-MSPM with early fault detection and diagnosis of process upsets and trends.

PCA (Principal Component Analysis (PCA) and PLS (Projection to Latent Structure (PLS) are the two popular techniques are used to create the multivariate (MV) models. MV statistical models for process monitoring are used in this study to address the challenges in biologics manufacturing process such as <u>raw material variation and control of process variability, identification and monitoring of relevant process parameters in the operating space and RT-MSPM for Early Detection and Diagnosis of Process Upsets and Trends.</u>

The implementation of RT-MSPM assists in meeting the latest process validation guidance requirement to achieve continued process verification (CPV) by monitoring each and every batch in real time. With the use of RT-MSPM tool, every run can be considered as a process validation run. If the process is monitored in real time then the sampling frequency can be reduced significantly, which can result in tremendous cost saving.

ACKNOWLEDGEMENTS

I am grateful to Kathleen Hayes, Principal Clerk, College of Pharmacy, for her continuous help for over seven years. Kathleen was very prompt in responding to calls for assistance. She kept me on track with all of my academic requirements. Kathleen guided me to ensure that I fulfill all university graduate requirements in time.

I acknowledge the help, support and comradeship of my Amgen colleagues David Binkley, Pranav Patel and Kelly Quesada.

I am grateful to my Parents, who always wanted me to pursue my Ph. D., because they have tirelessly followed my progress and taught me to complete every task I took upon.

Last but not the least; I am grateful to my wife, Mrunalini, and my daughters, Ishani and Shivani, for their sacrifice and patience during some of the toughest times and encouragement for completing this work. My daughters were always amazed to see me studying since they think grown-ups don't need to study. I am grateful to all of my family members and friends for their encouragement and support during this time.

PREFACE

This thesis is prepared according to the manuscript format. The manuscripts are included.

MANUSCRIPT 1: APPLYING QBD AND PAT IN COMMERCIAL BIOLOGICS MANUFACTURING, VALIDATION OF PROCESS MONITORING TOOL AND BENEFIT EVALUATION

MANUSCRIPT 2: USE OF MULTIVARIATE STATISTICAL PROCESS MONITORING TO ACHIEVE CONTINUED PROCESS VERIFICATION

The first manuscript will be submitted for publication in Journal of Pharmaceutical Innovation. The second manuscript will be submitted for publication in PDA's Journal of Pharmaceutical Sciences and Technology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

Regulatory Guidance Related to PAT (Process Analytical Technology) and QbD (Quality by Design):

PAT is a system for designing, analyzing, and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality.

The term *analytical* in PAT is viewed broadly and includes chemical, physical, microbial, mathematical, and risk analysis conducted in an integrated manner. The goal of PAT is to enhance understanding and control the manufacturing process. Quality cannot be tested into products; it should be built-in or should be by design. Consequently, the tools and principles described in this guidance should be used for gaining process understanding and can also be used to meet the regulatory requirements for validating and controlling the manufacturing process [2].

Using the approach of building quality into products, PAT guidance highlights the necessity for process understanding and opportunities for improving manufacturing efficiencies through innovation and enhanced scientific communication between manufacturers and the agency. Increased emphasis on building quality into products allows more focus to be placed on relevant multi-factorial relationships amongst material, manufacturing process, environmental variables, and their effects on quality. This enhanced focus provides a basis for identifying and understanding relationships among various critical formulation and process factors and for developing effective risk mitigation strategies (e.g., product specifications, process controls, training, etc.). The data and information to help understand these relationships can be leveraged through pre-formulation programs, development and scale-up studies, as well as from improved analysis of manufacturing data collected over the life of a product.

A desired goal of the PAT framework is to design and develop well understood processes that will consistently ensure a predefined quality at the end of the manufacturing process. Such procedures would be consistent with the basic tenet of QbD and could reduce risks to quality and regulatory concerns while improving efficiency. Gains in quality, safety,

and/or efficiency will vary depending on the process and the product, and are likely to come from reducing production cycle times by using on-line, in-line, and/or at-line measurements and controls ; preventing rejects, scrap, and re-processing ; real time release; increasing automation to improve operator safety and reduce human errors ; improving energy and material use and increasing capacity; facilitating continuous processing to improve efficiency and manage variability.

This guidance facilitates innovation in development, manufacturing and quality assurance by focusing on process understanding. These concepts are applicable to all manufacturing situations.

Process Understanding [2, 8]

A process is generally considered well understood, when variability from batch to batch is explained, a good run from bad run is predicted, and all the factors that can alter quality, are accounted for and are understood.

A focus on process understanding can reduce the burden for validating systems by providing more options for justifying and qualifying systems intended to monitor and control biological, physical, and/or chemical attributes of materials and processes. Structured product and process development on a small scale, using experimental design and on-line or in-line process analyzers to collect data in real time, can provide increased insight and understanding for process development, optimization, scale-up, technology transfer, and control. Process understanding then continues in the production phase when other variables (e.g., environmental and supplier changes) may possibly be encountered. Therefore, continuous learning over the life cycle of a product is important.

Real-time multivariate statistical process monitoring provides a means to proactively monitor this overall process variability. It build the necessary foundation towards predictive monitoring which is aligned with the regulatory agency expectation on risk management and continual process improvement post-commercialization [11,12].

Principles of PAT [12]

Pharmaceutical manufacturing processes often consist of a series of unit operations, each intended to modulate certain properties of the materials being processed. To ensure acceptable and reproducible modulation, consideration should be given to the quality attributes of incoming materials and their process-ability for each unit operation.

During the last three decades, significant progress has been made in developing analytical methods for chemical attributes (e.g., identity and purity). However, certain physical and mechanical attributes of pharmaceutical ingredients are not necessarily well understood. Consequently, the inherent, undetected variability of raw materials may be manifested in the final product.

Establishing effective processes for managing physical attributes of raw and in-process materials requires a fundamental understanding of attributes that are critical to product quality. Such attributes may pose a significant challenge because of their complexities and difficulties related to collecting representative samples. Since the Formulation design strategies are not generalized, the quality of these formulations can be evaluated only by testing samples of in-process materials and end products.

Currently, these tests are performed off line after preparing collected samples for the analysis. Different tests are needed because they only address one attribute of the active ingredient following sample preparation (e.g., chemical separation to isolate it from other components). During sample preparation, other valuable information pertaining to the formulation matrix is often lost.

Several new technologies are now available that can acquire information on multiple attributes with minimal or no sample preparation. These technologies provide an opportunity to assess multiple attributes, often nondestructively.

Appropriate use of PAT tools and principles (described below) can provide relevant information relating to physical, chemical, and biological attributes. The process understanding gained from this information will enable process control and optimization, address the limitation of the time-defined end points discussed above, and improve efficiency.

Process Analytical Technology Tools [2]

There are many new tools available that enable scientific, risk-managed pharmaceutical development, manufacture, and quality assurance. These tools, when used within a system can provide effective and efficient means for acquiring information to facilitate process understanding, develop risk-mitigation strategies, achieve continuous improvement, and share information and knowledge.

Producing a product consistently rests on four key areas of technology: multivariate data analysis, process analyzers, process automation/control and knowledge management.

When all of these ingredients are added to the mix, powerful solutions can be realized. Typically, collecting information from sensor and instruments is not complicated. Servers are bursting with data about processes. However, getting the process engineer the information he or she needs requires intensive IT involvement.

Even more importantly is getting access to this data in real time to make decisions about quality.

In the PAT framework, these tools can be categorized as:

I. Multivariate (more than one variable) data acquisition and analysis
II. Modern process analyzers or process analytical chemistry tools
III. Process and endpoint monitoring and control tools
IV. Continuous improvement and knowledge management tools

An appropriate combination of some, or all, of these tools may be applicable to a single-unit operation, or to an entire manufacturing process and its quality assurance.

Multivariate (more than one variable) data acquisition and analysis:

From a physical, chemical, or biological perspective, pharmaceutical products and processes are complex, multi-factorial systems. There are many development strategies that can be used to identify optimal formulations and processes. The knowledge acquired in these development programs is the foundation for product and process design.

This knowledge base can help to support and justify flexible regulatory paths for innovation in manufacturing and post approval changes. A knowledge base can be a benefit when it consists of scientific understanding of the relevant multi-factorial relationships (e.g., between formulation, process, and quality attributes), as well as a means to evaluate the applicability of this knowledge in different scenarios (i.e., generalization). This benefit can be achieved through the use of multivariate mathematical approaches, such as statistical design of experiments, response surface methodologies, process simulation, and pattern recognition tools, in conjunction with knowledge management systems. The applicability and reliability of knowledge in the form of mathematical relationships and models can be assessed by statistical evaluation of model predictions.

Methodological experiments based on statistical principles of orthogonality, reference distribution, and randomization; provide effective means for identifying and studying the effect and interaction of product and process variables. Traditional one-factor-at-a-time experiments do not address interactions among product and process variables.

When used appropriately, these tools enable the identification and evaluation of product and process variables that may be critical to product quality and performance. The tools may also identify potential failure modes and mechanisms and quantify their effects on product quality.

Modern process analyzers or process analytical chemistry tools:

Process analysis has advanced significantly during the past several decades, due to an increasing appreciation for the value of collecting process data. Industrial drivers of productivity, quality, and environmental impact have supported major advancements in this area. Available tools have evolved from those that predominantly take univariate process measurements, such as pH, temperature, and pressure, to those that measure biological, chemical, and physical attributes. Indeed some process analyzers provide nondestructive measurements that contain information related to biological, physical, and chemical attributes of the materials being processed. These measurements can be <u>At-line: Measurement, On-line Measurement and In-line Measurement.</u>

Process analyzers typically generate large volumes of data. Certain data is likely to be relevant for routine quality assurance and regulatory decisions. In a PAT environment, batch records should include scientific and procedural information indicative of high process quality and product conformance. For example, batch records could include a series of charts depicting acceptance ranges, confidence intervals, and distribution plots (inter- and intra-batch) showing measurement results. Ease of secure access to these data is important for real time manufacturing control and quality assurance. Installed information technology systems should accommodate such functions.

Measurements collected from these process analyzers need not be absolute values of the attribute of interest. The ability to measure relative differences in materials before (e.g., within a lot, lot-to-lot, different suppliers) and during processing will provide useful information for process control. A flexible process may be designed to manage variability of the materials being processed. Such an approach can be established and justified when differences in quality attributes and other process information are used to control (e.g., feed-forward and/or feed-back) the process.

The advances in process analyzers made the real time control and quality assurance during manufacturing feasible. However, multivariate methodologies are often necessary to extract critical process knowledge for real time control and quality assurance.

Comprehensive statistical and risk analyses of the process are generally necessary to assess the reliability of predictive mathematical relationships. Based on the estimated risk, a simple correlation function may need further support or justification, such as a mechanistic explanation of causal links among the process, material measurements, and target quality specifications. For certain applications, sensor-based measurements can provide a useful process signature that may be related to the underlying process steps or transformations. Based on the level of process understanding, these signatures may also be useful for process monitoring, control, and end point determination when these patterns or signatures relate to product and process quality.

Design and construction of the process equipment, the analyzer, and their interfaces are critical to ensure that collected data are relevant and representative of process and product attributes. Robust design, reliability, and ease of operation are important considerations.

Installation of process analyzers on existing process equipment in production should be done after risk analysis to ensure this installation does not adversely affect process or product quality.

A review of current standard practices (e.g., ASTM International) for process analyzers can provide useful information and facilitate discussions with the Agency. A few examples of such standards are listed in the bibliography section. Additionally, standards forthcoming from the ASTM Technical Committee E55 may provide complimentary information for implementing the PAT Framework. We recommend that manufacturers developing a PAT process consider a scientific, risk-based approach relevant to the intended use of an analyzer for a specific process and its utility for understanding and controlling the process.

Process and endpoint monitoring and control tools:

It is important to emphasize that a strong link between product design and process development is essential to ensure effective control of all critical quality attributes. Process monitoring and control strategies are intended to monitor the state of a process and actively manipulate it to maintain a desired state. Strategies should accommodate the

attributes of input materials, the ability and reliability of process analyzers to measure critical attributes, and the achievement of process end points to ensure consistent quality materials and the final product.

The design and optimization of drug formulations and manufacturing processes within the PAT framework can include steps such as identifying critical attributes, measurement of the critical attributes, design process control to monitor and maintain these attributes within the operating space.

Within the PAT framework, a process end point is not a fixed time; rather it is the achievement of the desired material attribute. This, however, does not mean that process time is not considered. A range of acceptable process times (process window) is likely to be achieved during the manufacturing phase and should be evaluated, and considerations for addressing significant deviations from acceptable process times should be developed.

Where PAT spans the entire manufacturing process, the fraction of in-process materials and final product evaluated during production could be substantially greater than what is currently achieved using laboratory testing. Thus, an opportunity to use more rigorous statistical principles for a quality decision is provided. Rigorous statistical principles should be used for defining acceptance criteria for end point attributes that consider measurement and sampling strategies. Multivariate Statistical Process Control can be feasible and valuable to realizing the full benefit of real time measurements. Quality decisions should be based on process understanding and the prediction and control of relevant process/product attributes. This is one way to be consistent with relevant CGMP requirements, as such control procedures that validate the performance of the manufacturing process (21 CFR 211.110(a)).

Systems that promote greater product and process understanding can provide a high assurance of quality on every batch and provide alternative, effective mechanisms to demonstrate validation (per 21 CFR 211.100(a), i.e., production and process controls are designed to ensure quality). In a PAT framework, validation can be demonstrated through

continuous quality assurance where a process is continually monitored, evaluated, and adjusted using validated in-process measurements, tests, controls, and process end points.

Risk-based approaches are suggested for validating PAT software systems. The recommendations provided by other FDA guidance, such as General Principles of Software Validation [17] should be considered. Other useful information can be obtained from consensus standards, such as ASTM.

Continuous improvement and knowledge management tools:

Continuous learning through data collection and analysis over the life cycle of a product is important. These data can contribute to justifying proposals for post approval changes. Approaches and information technology systems that support knowledge acquisition from such databases are valuable for the manufacturers and can also facilitate scientific communication with the Agency.

Opportunities need to be identified to improve the usefulness of available relevant product and process knowledge during regulatory decision making. A knowledge base can be of most benefit when it consists of scientific understanding of the relevant multi-factorial relationships (e.g., between formulation, process, and quality attributes) as well as a means to evaluate the applicability of this knowledge in different scenarios (i.e., generalization). Today's information technology infrastructure makes the development and maintenance of this knowledge base practical.

Process Validation Guidance

A typical biologics manufacturing process starts with inoculation phase and end up into final product which is distributed to patients as shown in Figure 1 [25]. This process involves several upstream unit operations such as series of cell culture bioreactors, centrifuges, filtration steps and downstream unit operations such as chromatography, ultra-filtration and diafiltration (UF/DF), viral inactivation, etc.

Traditional Process Validation Approach: Per 21 CFR Parts 210 and 211, and of the Good Manufacturing Practice Regulations for Medical Devices, 21 CFR Part 820[15],

every pharmaceutical or biologics manufacturing organization has to go through rigorous testing and qualification phase before they seek approval for large scale manufacturing of drug substance or drug product. The term "qualification" and "validation" are separate terms but are interchangeably used in the industry [16]. The FDA's definition of validation is "Establishing documented evidence that a process or system, when operated within established parameters can perform effectively and reproducibly to produce a medicinal product, that meets its pre-determined specifications and quality attributes [17]".

In other words, each and every piece of equipment used in the manufacturing facility needs to undergo Installation, Operational and Performance Qualification processes to meet the guidance. The automated and computerized systems needs to go through "Software Validation [16] and Part 11 Compliance for electronics records and electronic signature validation [18] to ensure that the data inputs and outputs of these systems are secured and trust worthy just like paper records. It ensures that all the critical equipment is installed correctly; operate with the operating ranges and performs within the acceptable criteria.

Upon completion of above mentioned validation process, the process validation is performed. Process validation is a federal requirement therefore; it is applicable to all the manufacture of pharmaceuticals and medical devices. Per "Guideline on general principles of process validation, May 1987", manufacturing processes needed to be validated. Assurance of product quality was derived from careful attention to a number of factors including selection of quality parts and materials, adequate product and process design, control of the process, and in-process and end-product testing [17].

As stated in old process validation guidance, the manufacturers needed to perform confirmation runs a.k.a process validation runs to prove that the process is capable of effectively meeting the key and critical process parameter acceptance criteria. The key and critical operating parameters were within operating ranges and the process was able to generate the product in a controlled manner. The analytical assays tested the incoming

raw materials; in process material and finished product material to ensure that that they meet the specifications. The guidelines suggested establishing robust test protocols to specify "A sufficient number of replicate process runs to demonstrate reproducibility and provide an accurate measure of variability among successive runs [17]". It did not exactly specify how many runs. The manufacturing industry started performing three process validation runs and soon it became a norm. The standard industry practice became three consecutive process validation runs. Upon completion of this torturous, expensive and time consuming process FDA inspected the facilities gave approvals to manufacture the drug substances or drug products. The process was never validated after that throughout its lifecycle.

FDA's lifecycle Approach for Process validation: In 2011 FDA came up with new process validation guidance. This guidance is aligned with existing FDA guidance, including the FDA/International Conference on Harmonization (ICH) guidance for industry, Q8 - Pharmaceutical Development [4], Q9- Quality Risk Management [5], Q10 - Pharmaceutical Quality System [6] and Q11- Development and manufacture of drug substances [7]. Although; this guidance does not repeat the concepts and principles explained in other guidance's, FDA encourages the use of modern pharmaceutical development concepts, quality risk management, and quality systems at all stages of the manufacturing process lifecycle [9].

Per this new guidance, manufacturers are required to adopt the lifecycle approach by performing the process validation activities in three stages [9] after completing the equipment and facility qualification.

The three stages in the lifecycle approach outlines the development phase where the product knowledge and process understanding is gained to establish the operating space. The stage 1 is linked to process qualification stage for process validation. Utilizing the stage 1 and 2, the new expectation is to perform the continued process verification (CPV) to ensure that process remains in the control and consistently make the quality product.

Three stages of process validation are outlined below:

I. In Stage 1, process design, the commercial process is defined based on knowledge gained through development and scale-up activities.

II. In Stage 2, process qualification, the process design is evaluated and assessed to determine if the process is capable of reproducible commercial manufacturing.

III. In Stage 3, continued process verification, ongoing assurance is gained during routine production that the process remains in a state of control.

Pre-Requisites of PAT Implementation [2]:

In order to get PAT on a more practical and operational level, we can list a number of prerequisites:

Infrastructure: Automated data acquisition systems, databases, networks, and synchronization procedures must be in place. The greatest hurdle involved in almost any analysis is generation, integration and organization of data. This is particularly true for the pharmaceutical industry where data are often stored in vast warehouses but rarely, if ever, retrieved and used. Past regulatory environments did not provide incentives for analysis of manufacturing processes because implementing improvements required re-validation and the current condition of pharmaceutical data infrastructures reflects this. As a result large efforts are required to assemble meaningful datasets. This challenge is further complicated given that laboratory and production data are scattered in various disconnected databases. Examples of these databases include Laboratory Information Management Systems (LIMS), Manufacturing Execution Systems (MES), Enterprise Resource Planning systems (ERP) as well as Supervisory Control and Data Acquisition systems (SCADA) and process historian databases. Product quality is influenced by all stages of production including variability of the raw materials. Developing process understanding of a finished product can only be realized through uncovering the cumulative influence of all processing steps and their interactions. Integrating, synchronizing and aligning data from all relevant sources is therefore a pre-requisite before analysis can begin.

Multivariate characterization: Adequate and informative data must be measured on all steps and ingredients of the process.

Multivariate evaluation of all data: All data should be analyzed together. The data analysis should not focus on variable selection, should not be univariate in nature, and should not involve methods with many adjustable parameters which are prone to over fit. The data analysis phase should entail simple, transparent, informative, and reversible projection models.

Data and information integration and communication: All data flows and data bases should be integrated onto one common platform. This facilitates use of data, visualization of data, and communication of results.

Design of Experiment (DOE): A suitable use of DOE combined with some of the steps above can augment the analysis and help to ensure that critical system parameters are varied together in a simultaneous to get the optimum information from the experiments.

Strategy for Implementation of PAT [2]

The Agency understands that to enable successful implementation of PAT, flexibility, coordination, and communication with manufacturers is critical. The Agency believes that current regulations are sufficiently broad to accommodate these strategies. Regulations can effectively support innovation when clear, effective, and meaningful communication exists between the Agency and industry, for example, in the form of meetings or informal communications.

The first component of the PAT framework described above addresses many of the uncertainties with respect to innovation and outlines broad principles for addressing anticipated scientific and technical issues. This framework should assist a manufacturer in proposing and adopting innovative manufacturing and quality assurance. The Agency encourages such proposals and has developed a regulatory strategy to consider such proposals.

Ideally, PAT principles and tools should be introduced during the development phase. The advantage of using these principles and tools during development is to create opportunities to improve the mechanistic basis for establishing regulatory specifications. Manufacturers are encouraged to use the PAT framework to develop and discuss approaches for establishing mechanistic-based regulatory specifications for their products. The recommendations provided in this guidance are intended to alleviate concerns with approval or inspection when adopting the PAT framework.

In the course of implementing the PAT framework, manufacturers may want to evaluate the suitability of a PAT tool on experimental and/or production equipment and processes. For example, when evaluating experimental on- or in-line process analyzers during production, it is recommended that risk analysis of the impact on product quality be conducted before installation. This can be accomplished within the facility's quality system without prior notification to the Agency. Data collected using an experimental tool should be considered research data. If research is conducted in a production facility, it should be under the facility's own quality system.

When using new measurement tools, such as on- or in-line process analyzers, certain data trends, intrinsic to a currently acceptable process, may be observed. Manufacturers should scientifically evaluate these data to determine how or if such trends affect quality and implementation of PAT tools. FDA does not intend to inspect research data collected on an existing product for the purpose of evaluating the suitability of an experimental process analyzer or other PAT tool. FDA's routine inspection of a firm's manufacturing process that incorporates a PAT tool for research purposes will be based on current regulatory standards (e.g., test results from currently approved or acceptable regulatory methods). Any FDA decision to inspect research data would be based on exceptional situations similar to those outlined in Compliance Policy Guide Sec. 130.300.4 Those data used to support validation or regulatory submissions will be subject to inspection in the usual manner.

Challenges of PAT Implementation [2, 8]

For successful implementation of PAT combination of following areas must be considered:

Organization support:

One of the most important factors in ensuring the success of process analytical methods is strategic organizational support, which can afford to design, implement and maintain the PAT systems. The most PAT systems have significant upfront costs and efforts, they require management support.

Implementation of process analytical methods requires the highest level of interaction with plant personnel and management. The aspect of ownership and roles and responsibilities are handled effectively as long there is continuity in management support.

The development of PAT applications by central groups and subsequent transfer of ownership to the plant require the highest level of interaction with plant personnel and technology transfer teams. It is critical to success that the interests of central organizations and sites are aligned to effectively develop and support the process analytical systems, particularly those with relatively high complexity.

Necessary roles in PAT implementation can be filled by personnel from different departments:

Management: Provides overall project management, business support and expertise as well as strategic oversight

Procurement: Works with vendors to facilitate and coordinate purchase of equipment

Site/Plant Operations: Contributes technical know-how as well as process/manufacturing logistics

Process Analytics: Contributes analytical knowledge, analyzer technology development, analyzer engineering, implementation support and training

Automation and control: Handles efforts in the area of process control and automation of process equipment

Maintenance: Responsible for care and feeding of process analytical equipment

Vendor: Provide products and services for PA implementation

Quality: Potential oversight of data and product quality

Regulatory expertise: Necessary for regulatory compliance and strategic partnerships with regulatory agencies.

Profile of a process analytical scientist/engineer: For successful implementation of PAT project, process analytical personnel should have combination of competencies such as technical, interpersonal effectiveness, initiative, business focus, innovative, learning, and leadership skills.

Industrial Application of PAT tools: Multivariate Data Analysis (MVDA):

Due to the recent advancement in the computer technology, we have the capability to collect larger amounts of data. This trend will continue to accelerate in the next decades as the technological developments continues. Multivariate data analysis (MVDA) is becoming increasingly popular because the on-going data collection tends to overload our computers and data-bases with tons of data. It is therefore necessary to work on bigger samples if full advantage is to be taken of all accessible information. It is also necessary to derive as much information as possible from the diversity of the data, rather than restricting attention to subsets of it. The use of multivariate data analysis techniques provides this opportunity and it can be used to reveal the information otherwise impossible to know. By using MVDA we can extract much more information than univariate data analysis techniques for the selected variables and observations. This data then needs to be analyzed so that meaningful information can be extracted from it. MVDA methods are being used in pharmaceutical and biotechnology industries in

several areas [12] such as 1) Process monitoring, and early fault detection and classification; 2) Process Analytical technology tool; 3) Quality control analysis; 4) Data mining and integration; 5) Structure activity relationship; 6) Multivariate characterization; 7) Multivariate calibration; and 8) Multivariate characterization and discrimination analysis.

In the biological manufacturing process, tremendous amount of data is generated from various sensors during each phase for the respective unit operation. If the manufacturing operation is comprised of the high tech data collection sensors, then the amount of data can be generated for each phase from few seconds interval to several days interval. The in- process test results helps to measure and ensure that the process is running under control. This data is stored on the computer and server databases.

The use of various multivariate analysis techniques such as Principal Component analysis (PCA) and Projection to Latent Structures (PLS) modeling can give a meaning to this data. 'The masses of process data can provide easy to grasp graphical information about the state of the process, and relations between important sets of process variables. These multivariate methods make efficient use of all pertinent data, with little loss of information'. [19]

In order to perform MVDA, it is important to understand the variability, complexity of the data and the type of data being analyzed. The MVDA is well suited to deal with variability in the complex data, thereby reducing the risk of incorrect inferences. However; all the data points are needed. One should not disregard the data because the variables are often collinear, either partially or completely. If part of the data is ignored then there is a substantial risk of overlooking the important information.

In MVDA, most common and widely-used methods are PCA and PLS. These methods present the modeling results graphically and the observations and variables are easily available for diagnostics and interpretation. PCA and PLS methods are mainly popular because they can deal with the problems related to dimensionality, co-linearity, noise and missing data. These methods offer a number of diagnostic tools, which facilitate the identification of assignable causes.

PCA and PLS can be used to address three main types of data issues such as overview of data, classification & discrimination and regression modeling.

Principal Components Analysis (PCA) [14, 20, 39]:

The PCA method was first introduced by Pearson in 1901. In 1933, Harold Hotelling formulated the PCA theory. But, 1980 onwards, due to development of personal computers, application of PCA exploded [20].

PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. The main advantage of PCA is that once you have found these patterns in the data, and you compress the data, i.e. by reducing the number of dimensions, without much loss of information [21]. Statistically, PCA finds a new lines or planes in the multi-dimensional space that approximate the data as well as possible in the least square sense. The goal of PCA is to reduce the number of variables of interest into a smaller set of components. First principal component (1st PC) explains maximum variation and subsequent PC's explain remaining variation in a descending order [12].

PCA method is useful for over viewing a data matrix X, because it is able to explore relationships both among variables and observations. PCA provides the understanding of the relationships between variables which contribute similar information to the PCA model and which variables provide unique information about the observations.

Principal Components Analysis have several objectives [22] such as dimensionality reduction; determining of linear combinations of variables; choosing of the most useful variables; visualization of multidimensional data; identification of underlying variables; identification of groups of objects or outliers.

The tasks required of the analyst to carry these out are as follows:

Dimensionality Reduction:

In case of a table of dimensions N x M, each of the *N* rows or observations can be regarded as an *M*-dimensional vector or variables. Finding a set of M' < M principal axes allows the objects to be adequately characterized on a smaller number of (artificial) variables. This is advantageous as a prelude to further analysis as the M – M' dimensions may often be ignored as constituting noise; and, secondly, for storage economy (sufficient information from the initial table is now represented in a table with M' < M columns). Reduction of dimensionality is practicable if the first M' new axes account for approximately 75 % or more of the variance. There is no set threshold, the analyst must judge. The cumulative percentage of variance explained by the principal axes is consulted in order to make this choice.

The determining of linear combinations of variables:

The data matrix X is projected into multidimensional space. PCA method provides the understanding of the relationships between variables. This relationship is transformed into a covariance matrix. The eigenvalues and eigenvectors are the properties of matrix. The relationship between the variables, their length and the direction of PC vectors is explained by eigenvalues and eigenvectors. The eigenvector are found in square matrix, its direction is not affected by scaling and they are orthogonal to each other. The eigenvalues are closely related to eigenvectors because they always come in pairs. It is important in PCA that each eigenvector to be of unit length [21] that means the variance of the eigenvector is one. If the eigenvalue is zero, the variance of projections on the associated eigenvector is zero. Hence the eigenvector is reduced to a point. If this point is additionally the origin (i.e. the data is centered), then this allows linear combinations between the variables to be found. In fact, we can go a good deal further: by analyzing second-order variables, defined from the given variables, quadratic dependencies can be straightforwardly sought. This means, for example, that in analyzing three variables, y_1, y_2, and y_3, we would also input the variables y_1^2, y_2^2, y_3^2, y_1y_2, y_1y_3, and y_2y_3. If the linear combination $y_1 = c_1 y_2^2 + c_2 y_1y_2$ exists, then we would find it. Similarly we could feed in the logarithms or other functions of variables.

Feature selection: the choosing of the most useful variables:

In feature selection we want to simplify the task of characterizing each object by a set of attributes. Linear combinations among attributes must be found; highly correlated attributes (i.e., closely located attributes in the new space) allow some attributes to be removed from consideration; and the proximity of attributes to the new axes indicate the more relevant and important attributes. As stated earlier, PCA method calculates the eigenvectors and eigenvalues from the relationship matrix. The eigenvector with the *highest* eigenvalue is the *principal component* of the data set. In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. This gives you the components in order of significance [21]. This step assists in in choosing the most useful variables.

Visualization of multidimensional data:

In order to provide a convenient representation of multidimensional data, planar plots are necessary. An important consideration is the adequacy of the planar representation: the percentage variance explained by the pair of axes defining the plane must be looked at here.

Identification of underlying variables:

PCA is often motivated by the search for latent variables. Often it is relatively easy to label the highest or second highest components, but it becomes increasingly difficult as less relevant axes are examined. The objects with the highest loadings or projections on the axes (i.e. those which are placed towards the extremities of the axes) are usually worth examining: the axis may be characterisable as a spectrum running from a small number of objects with high positive loadings to those with high negative loadings.

Identification of groups of objects or of outliers:

A visual inspection of a planar plot indicates which objects are grouped together, thus indicating that they belong to the same family or result from the same process. Anomalous objects can also be detected, and in some cases it might be of interest to redo the analysis with these excluded because of the perturbation they introduce.

In this process, the principal components are derived as [23]:

$$PC1 = b11X1 + b21X2 + \dots + bk1Xk \qquad \text{(Eqn. 1)}$$

$$PC2 = b12X1 + b22X2 + \dots + bk2Xk \qquad \text{(Eqn. 2)}$$

$$PCf = b1fX1 + b2fX2 + \dots + bkfXk \qquad \text{(Eqn. 3)}$$

PCA modeling shows the correlation structure of data matrix X, approximating it by a matrix product of lower dimension (TP'), called principal components plus a matrix of residuals (E). The PCA model is shown by following equation [12]:

$$X = 1 * \bar{x} + TP' + E \qquad \text{(Eqn. 4)}$$

where,

T = Matrix of scores that summarizes the X-variables

P' = Matrix of loadings showing the influence of the variables

E = Matrix of residuals showing the variation in the data which is left out of modeling

The PCA model as shown in equation (4) can monitor new batch in real time without having to estimate the future portion of the new data point. The score for the new batch variable is calculated using equation (5).

$$T_{pred, k} = X_{new, k} * W (P'W)^{-1} \qquad \text{(Eqn. 5)}$$

where,

W = Weights of the Matrix of scores that summarizes the X-variables

Projections to Latent Structures (PLS) [12, 14, 24, 35, 39]:

The PLS approach was originated in 1975 by Herman Wold. He developed a simple way to estimate parameters in the model called NIPALS (Nonlinear Iterative partial least squares). These are later called PLS models. In PLS model, P indicates 'partial' because it is a partial regression since parameter vector (X variable) is considered fixed in the estimation. In 1980, the PLS started to interpret as "Projection to Latent Structures".

PLS is a similar technique which also reduces the dimensionality of the input space X, however, it does this while finding the best regression fit against a response variable Y. PLS method utilizes regression modelling between two data metrices, usually denoted by X and Y, with the aim of predicting Y from X for new observations. This is achieved by "Linear Multivariate" modelling. In PLS modelling, the aim is to predict complex response or output variables (Y) based on the input variables (X). The precision of PLS model increases with increasing number of X variables.

In process modelling the PLS method finds the relationship between input (X variables) measured on the process at N time points and corresponding values of Y (output variables). The PLS model can project and data table as long as there is a similarity between observations.

PLS model consists of a structural part, which reflects the relationships between the latent variables, and a measurement component, which shows how the latent variables and their indicators are related.

PLS starts by calculating case values. For this purpose, the "unobservable variables are estimated as exact linear combinations of their empirical indicators", and PLS treats these estimated proxies as perfect substitutes for the latent variables. The weights used to determine these case values are estimated so that the resulting case values capture most of the variance of the X variables that is useful for predicting the Y variable. This is based on the implicit assumption that all measured variance of the variables in the model is useful variance that should be explained. Using these weights, it is then possible to determine a value for each unobservable variable, simply by calculating a weighted average of its indicators. This results in a model in which all unobservable variables are approximated by a set of case values and that can, therefore, be estimated by a set of simple, first-generation, ordinary least squares regressions.
The basic idea of PLS is quite straightforward:

First, the weight relations, to their respective unobservable variables, are estimated.

Second, case values for each unobservable variable are calculated, based on a weighted average of its indicators, using the weight relations as an input.

Finally, these case values are used in a set of regression equations to determine the parameters for the structural relations.

This explanation makes it obvious that the most crucial part of a PLS analysis is the estimation of the weight relations. Of course, it would be easier simply to <u>assume equal weights for all indicators</u>, but this approach has two disadvantages:

First, there is no theoretical rationale for all indicators to have the same weighting. Because it can be assumed that the resulting parameter estimates of the structural model depend on the type of weighting used, at least as long as the number of indicators is not excessively large, the (exogenous) assumption of equal weights makes the results highly arbitrary. Second, such a procedure does not take into account the fact that some indicators may be more reliable than others and should, therefore, receive higher weights.

Consequently, PLS uses a more complex, two-step estimation process to determine the weights (w): First, it starts with an outside approximation, in which case values for each latent variable are estimated, based on a weighted average of their respective indicators. The weights used to calculate this aggregation is determined in a manner similar to a principal-components analysis for reflective or regression analysis for formative indicators. In the next step, the inside approximation, improved case values are determined as a weighted average of neighboring latent variables. For this process, there are three different weighting schemes available, but one can demonstrate that the choice between them has only a minor impact on the final results. Using this second estimate of the case values, the weight relations are modified and the process of inside and outside approximation starts from the beginning again and is repeated until convergence of the case values is achieved.

Hence, being a limited information approach PLS has the advantage that it "involves no assumptions about the population or scale of measurement" and consequently works without distributional assumptions and with nominal, ordinal, and interval scaled variables. However, one has to bear in mind that PLS, like any statistical technique, also requires certain assumptions to be fulfilled. Beyond those known from the standard regression model, the most important assumption is predictor specification. This requirement states that the systematic part of the linear regression must be equal to the conditional expectation of the dependent variable and can be considered as fulfilled in most cases. PLS is quite robust with regard to several inadequacies (e.g., skewness or multicollinearity of the indicators, misspecification of the structural model) and that the latent variable scores always conform to the true values.

However, there is also another side of the coin, namely, the problem of consistency at large. In general, a consistent estimator can be described as "one that converges in probability to the value of the parameter being estimated as the sample size increases". However, because the case values for the latent variables in PLS are aggregates of manifest variables that involve measurement error, they must be considered as inconsistent.

Therefore, "the path coefficients estimated through PLS converge on the parameters of the latent-variable model [only] as both the sample size and the number of indicators of each latent variable becomes infinite" a problem known under the term consistency at large. Hence in all real-life situations, in which both the number of cases in the sample and the number of indicators per latent variable will be finite, PLS tends to underestimate the correlations between the latent variables and overestimate the loadings. Only when the number of cases in the sample and the number of indicators per latent variable increase to infinity do the latent variable case values approach the true values and this problem disappears.

PLS modeling consists of simultaneous projections of both the X and Y spaces. The coordinates of the points on the X and Y dimensions constitutes the elements of the T and

U score matrices, P' and C' loading matrices and E and F residual matrices as shown in equation-4 and 5. The objective here is [12] to well approximate the X and Y spaces and to maximize the correlation between X and Y.

$$X = 1 * \bar{x} + TP' + E \qquad \text{(Eqn. 4)}$$

$$Y = 1 * \bar{Y} + UC' + F \qquad \text{(Eqn. 6)}$$

where,

T and U= Matrix of scores that summarizes the X-variables and Y variables

P' and C' = Matrix of loadings showing the influence of the variables in X and Y matrix

E & F= Matrix of residuals showing the variation in the data which is left out of modeling for X matrix and Y matrix

The batch level model is used to predict the final performance variable using the X matrix T scores as shown in equation-5 and equation-6 [12]. $T_{pred, k}$ and $Y_{pred, k}$ in the following equation calculate the estimated scores and quality / performance attributes at time slice 'k' in a given batch.

$$T_{pred, k} = X_{new, k} * W \, (P'W)^{-1} \qquad \text{(Eqn. 5)}$$

$$Y_{pred, k} = T_{pred, k} * C' \qquad \text{(Eqn. 7)}$$

Pre-Treatment of Data and Scaling Techniques

Like any other statistical application, PCA require the data to be pre-processed prior to using. The variables often have different numerical ranges. A variable with large range has a large variance and a variable with small range has small variance. Since PCA is a maximum variance projection method, it follows that a variable with large variance is more likely to be expressed in the modeling than a low-variance variable. In order to give equal weight to all the variables, the data from the variables needs to standardized. This

process is called scaling. A combination of scaling techniques can also be used as shown in Figure 2 [26]. There are following ways of scaling the data:

Mean Centering: For PCA to work properly, you have to subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. This produces a data set whose mean is zero.

Scaling to Unit Variance: The long variables are shrunk and short variables are stretched, so that, all variables are set to equal length. The UV eliminates the differences due to unit of measures so that the variable with high values does dominate the model.

This is achieved by multiplying each variable by $1/S_k$ where S_k denotes standard deviation of each variable.

Auto scaling: When unit variance scaling is combined with mean centering, it is denoted as 'auto scaling'.

Pre-Requisites and Challenges of Multivariate Model Creation [12, 27]:

Handling of missing data: It is common in process systems that some batches or phases within batches having missing data points or a particular phase is not run for various operational reasons. PCA and PLS techniques inherently capable of predicting missing data for historical batches that are considered as representing normal operation and are highly repeatable therefore using the estimates for the missing data is acceptable [28].

Removal of the outliers: PCA discovers strong outliers and moderate outliers. Outliers are the observations which do not fit the model. Prior to modeling, outliers in the data set should also be detected and removed where applicable. The strong outliers pull the PCA model to themselves. The moderate outliers are captured by residual plots. Detection and elimination of outliers are critical for reliable and robust modeling and monitoring. There also are batch level outliers where a batch level score would be grossly different from the rest of the batches most likely due to a difference in the operational batch characteristics (e.g., different set points) or too many spikes. The outliers can be detected and removed via visual inspection; this can also be done via PCA-based modeling. A PCA model can

be developed and scores space is inspected for outlying data points or batches. PCA can discover strong and moderate outliers. Strong outliers are found in the PC scores plots and Hotelling's T^2 plots and moderate outliers are in the residuals (DModX at batch level).

Alignment of Batch Trajectories: Variability in total duration of the batches (such as operational condition switches like set points as well as biological variability such as cell growth peak maximum) result in unequal and unaligned batch trajectories across batch history. Alignment of the trajectories prior to modeling is important to ensure that (as much as possible) the variables or scores at any point during one batch correspond to those at the same biological or operational state in other batches.

There are many different techniques available for alignment of batch trajectories such as, use of an indicator or maturity variable [27, 29], dynamic time wrapping technique [12] and curve registration technique [27]

The most commonly used technique, is to align the batches based on a "maturity" (also known as an indicator) variable. The maturity variable is selected such a way that it should be monotonically increasing/decreasing, smooth and it indicates the end of a batch or a phase (such as reaction conversion, column volume totalized or simply time elapsed). Measurements on other variables are sampled with respect to the equal intervals of this variable so that batch trajectories are also aligned and set to equal size for modeling purposes. When there is no maturity variable the local batch time is used. SIMCA software by default selects $Time maturity variable for PLS model. The value of local batch time or maturity predicted by PLS model is suitable for a "maturity index" that can be used to indicate how far the batch has evolved [12].

Unfolding of three dimensional data into two dimensional data [12]:

Using SIMCA software, we can do two levels of batch monitoring; the batch evolution level monitoring (BEM) and batch level monitoring (BLM).
BEM Modelling:

The aim of batch evolution level monitoring is to develop a model of the good batches and monitor new batches against this model as they evolve to find out if they are evolving within the confidence limits. A data generated by a batch processes for a biological manufacturing process is arranged in the data blocks as shown in Figure 3a & 3b [30]. The batches are depicted as "I", Variables are depicted as "J" and time points are depicted as "K". In order to do the observation level modeling, the three way batch data table must be unfolded in such a way that the direction of the variable is preserved as shown in Figure 3a [30]. The resulting two-way matrix then has "I*J" rows and "K" columns. Each row contains data points X_{ijk} from a single batch observation (batch I, time K and variable J). If the regression is made against local batch time, the resulting PLS scores reflect linear t_1, t_2, and t_3 relationships (assuming a model with three latent variables) to local time batch time. A PCA on the on the three way data matrix will show how the individual observations relate to each other.

BL Modelling:

In batch level modelling, all the data from input variable matrix (X) and output variable matrix (Y) is available. Therefore, the data from the whole batch is used to create a model. The aim of the whole batch model is to verify the new whole batch is a good batch or bad batch. A data generated by a batch processes for a biological manufacturing process is unfolded in such a way that the direction of the batch is preserved. The resulting two-way matrix then has "I" rows and "J*K" columns as shown in Figure (3b). The resulting PLS model can be used to classify new batches as good or bad. Another important objective of batch level model is to understand how Y (output) variable is influenced by the X (input) variables.

Model Diagnostics - Selection of correct number of PC

Once the model is created, the required number of components is determined by cross validation technique. In order to ensure that the model is effective, it is important to have optimal balance between the goodness of fit and its predictive ability. The good ness of fit is given by the parameter R^2 (explained variation) and the goodness of prediction is given by Q^2 (predicted variation). Usually, R^2 and Q^2 vary differently as the complexity

of the model increases. Therefore, selection of number of parameters is based on the trade-off between goodness of fit and goodness of prediction as shown in Figure 4 [12].

There is another way of selecting number of components. You can plot eigenvalues of each component against the number of components as shown in Figure 5 [31]. The eigenvalue of the components represent the variation for those respective components. Any component whose eigenvalue is less than 1.0 is in most cases are eliminated because it reflects the lowest and negligible variance [23].

Setting the Control Limits [12]:

Once the batch evolution level and the models are created, the SIMCA software performs the rearrangement of reference model scores to create a new matrix. From this new matrix the averages and ± 3 standard deviation confidence limits are calculated. A number of multivariate statistical monitoring observation level and batch level plots and diagnostics are readily available in SIMCA for monitoring new batches in real-time such as; score plot, loading plot, DModX plot, Hotelling's T^2 plot, contribution Plot, Coefficient plot, VIP (variable importance in projection plot), and Observed vs. Predicted relationship plots.

Score Plots and Loading Plots: These plots are typically used at batch level to generate displaying the observations as situated in the multi-dimensional plane. These charts reveal the information about the how the variables are grouped, the covariance trend, if there are any outliers and similarities between them.

DModX (Distance to the model X) Plot: It is an estimate of how far from the model plane, in the X or Y space, the observation is positioned. DModX is used for process deviation detection where events are not necessarily explained by the model. The distance to model plot is displayed in normalized units after the last component with the default significance level of 0.05.

The DModX is calculated by using following formula [32]:

$$DModX = (\Sigma_i \, e_{ik}^2 \, / \, D.F.)^{1/2} \qquad \text{(Eqn. 8)}$$

where,

$\Sigma_i \, e_{ik}^2$ = Residual variable variation

D.F. = Degrees of freedom (K-A). Where, K= number of variables and A = number of components

If the DModX is larger than the critical limit then it indicates that the observation is an outlier in the X space.

Hotelling's T^2 Plot: These charts are also used for process deviation detection. They detect deviations that are explained by the process model (if DModX is in control) and within the overall variability but represent unusually high variation comparing to the average process behavior.

The Hotelling's T^2 for observation i, based on A components is calculated by using following formula [12]:

$$T_i^{\,2} = \Sigma \, ((t_{ia} - t_{avg})^2 \, / \, s_{ta}^2) \qquad \text{(Eqn. 9)}$$

where;

s^2ta = Variance of ta according to the class model.

$Ti^2 * (N - A) / A (N - 1)$

Follows a F distribution with A and N-A degrees of freedom.

N = Number of observations in the workset.

A = Number of components in the model or the selected number of components.

Hence if

$Ti^2 > A (N - 1) / (N - A) * F_{critical}$ (p=0.05)

then observation i is outside the 95% confidence region of the mode.

Contribution plots:

When the Hotelling's T^2 and/or DModX plots show a deviation (outlier from the confidence limit), indicating that some variable(s) are deviating from the historical average behavior without diagnosing which variables are contributing the most. Contribution plots are then used to explore into the original variable level and inspect which variable(s) are contributing to the inflated statistic [19].

VIP (Variable Importance in projection Plot):

The VIP (Variable Importance in Projection) plot summarizes the importance of the variables both to explain X matrix and to correlate to Y matrix. The VIP values are calculated for each X by summing the squares of the PLS loading weights, weighted by the amount of sum of squares explained in each model component. The sum of squares of all VIP's is equal to the number of terms in the model. Hence, the average VIP is equal to 1. VIP-values larger than 1 indicates "important" X-variables, and values lower than 0.5 indicate "unimportant" X-variables. The interval between 1 and 0.5 is a gray area, where the importance level depends on the size of the data set. The VIP plot is sorted from high to low, and shows confidence intervals for the VIP values, normally at the 95% level [19].
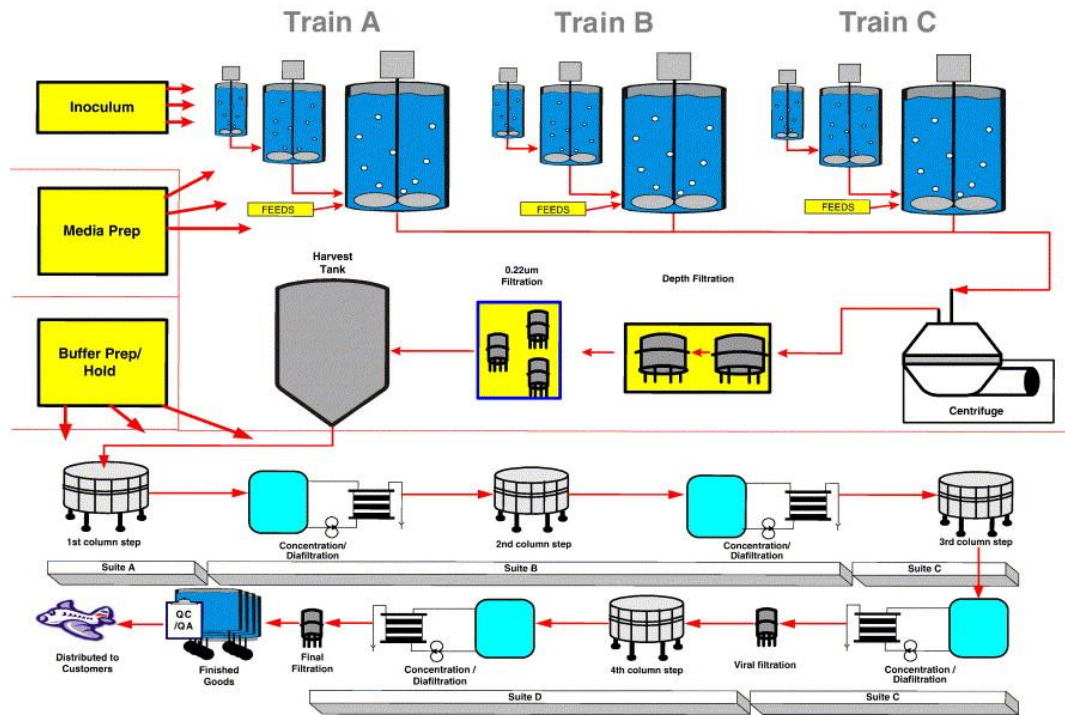
Figure 1: Typical (Biologics) Manufacturing Process [Source: 25]: The process involves several upstream unit operations such as inoculation, series of cell culture bioreactors, harvest tanks, filtration tanks, centrifuges and downstream unit operations such as multiple chromatography operations, multiple ultra-filtration and diafiltration (UF/DF) operations, viral inactivation, final filtration and filling and packaging.
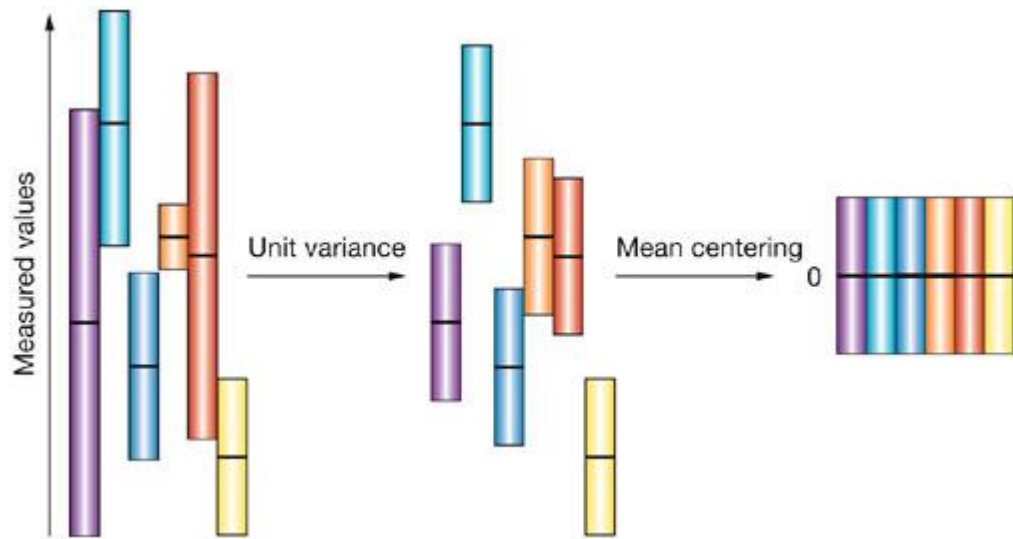
Figure 2: Data Scaling Techniques [Source: 26]: After mean –centering and unit variance scaling all variables will have equal "length" and mean value zero.
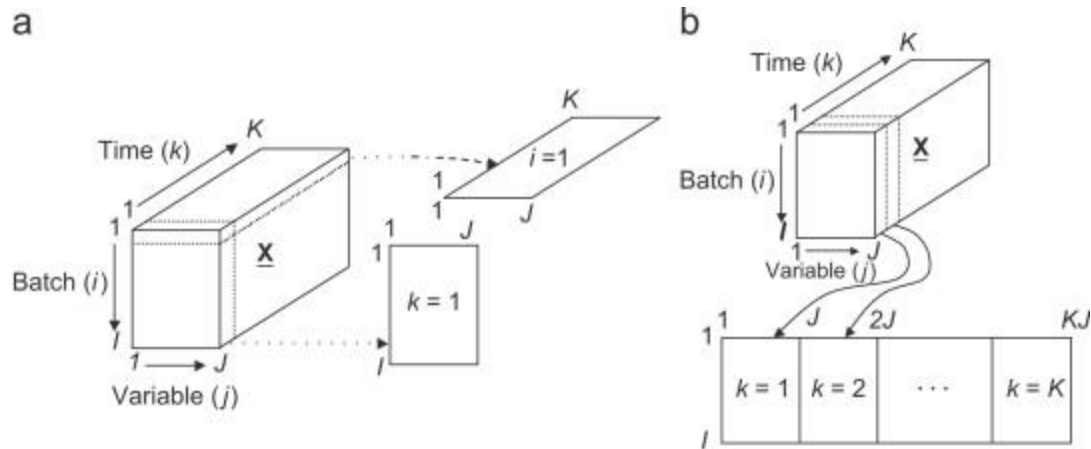
Figure 3a: Unfolding three way data matrix [Source: 30] the three way table of historical batch process data comprises "I", J variables and K time points. In batch evolution model (BEM), this three-way data table is unfolded by preserving the variable direction. This gives a two-way matrix with I*K rows and J columns. Each row contains data points X_{ijk} from a single batch evolution.

Figure: 3b, the three way table of historical batch process data comprises "I", J variables and K time points. In batch level model (BLM), this three-way data table is unfolded by preserving the batch direction. This gives a two-way matrix with I rows and J*K columns. Each row contains data points from one single batch.
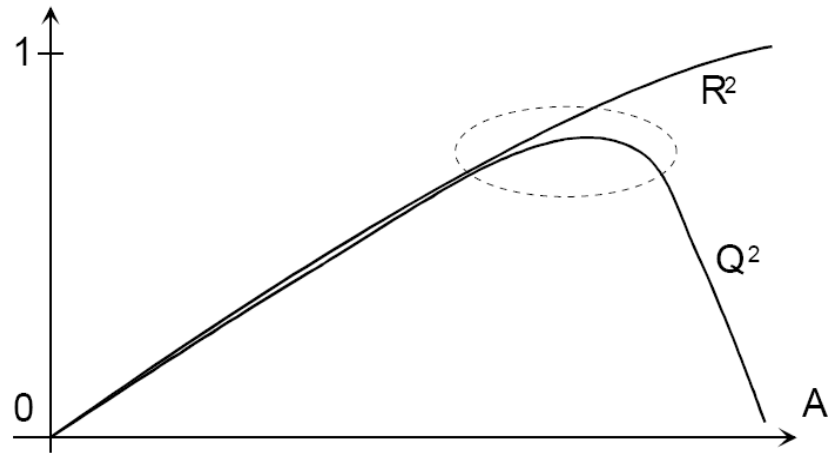
Figure 4: A trade-off between the goodness-of-fit (R^2) and goodness-of-prediction (Q^2). The vertical axis corresponds to the amount of variation and the horizontal axis corresponds to the total number of PCs (A) [Source: 12]. The number of components for the model is selected on the basis of optimal balance between fit and predictability.
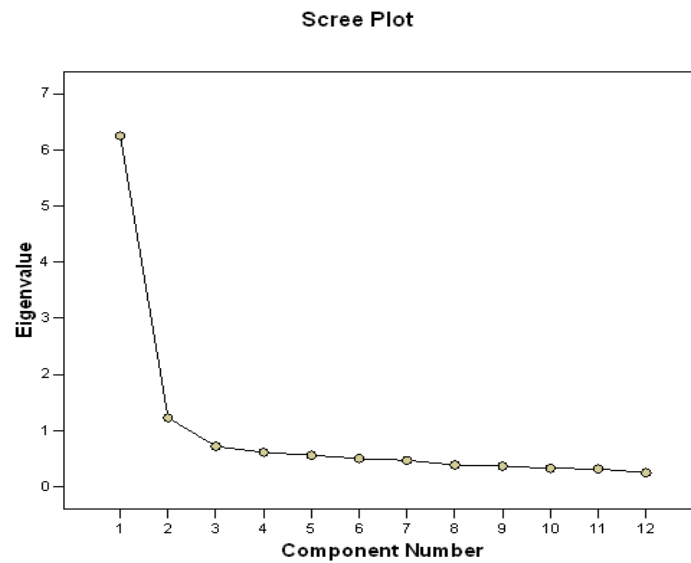
Figure 5: Another way of selecting number of components [Source: 31]. The vertical axis corresponds to the eigenvalues and the horizontal axis corresponds to the total number of PCs. The eigenvalue of the components represent the variation for those respective components. Any component whose eigenvalue is less than 1.0 is in most cases are eliminated because it reflects the lowest and negligible variance.

DEFINITIONS

Term	Definition
Batch level Model [19]	The MV model used for monitoring the batch finger print and predicting the final performance variable
Batch Evolution model [19]	The MV model used for monitoring the batch evolution with respect to a maturity variable in real-time
Chemometrics [24]	A way of analyzing chemical data, in which elements of both statistical and chemical thinking are combined.
Continuous Process Verification [9]	An alternative approach to process validation in which manufacturing process performance is continuously monitored and evaluated.
Design Space [4]	The multidimensional combination and interaction of input variables (e.g. Material attributes) and process parameters that have been demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process. Design space is proposed by the applicant and is subject to regulatory assessment and approval.
DModX Plot [12]	The statistic showing the distance of the observation to the MV model plane.
Hotelling's T^2 Plot [12]	The statistic summarizes the selected scores. It is a measure of how far away an observation is from the center of the MV model
QbD [8, 10]	A strategic approach to drug development, Quality by Design requires getting the product, process, Packaging and manufacturing "right the first time."
Quality [33]	Per ISO: "Degree to which a set of inherent characteristic fulfills requirements"
Loading Plot [12]	It is a summary of variables for observations (batches). It is a means to interpret the patterns in score plot.
Maturity Variable [12]	The variable indicating the evolution of a batch. It is used to understand how far the batch is evolved compared to the historical batches

Term	Definition
PAT [2]	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 the goal of ensuring final product quality
Process Analytics [2]	Chemical or Physical analysis of material in the process through the use of an in-line or on-line analyzer
Process Validation [17]	Establishing by objective evidence that a process consistently produces a result or product meeting its predetermined specifications.
Score Plot [12]	It is a summary of observations (batches)
SIMCA [19]	Software application supporting creation of multivariate models

ACRONYMNS

Abbreviation	Meaning
ACPS	Advisory Committee for Pharmaceutical Science
API	Active Pharmaceutical Ingredient
ASTM	American Society of Testing Materials
BLA	Biological License Application
BSPC	Batch statistical process control
CAPA	Corrective Action and Preventative Action
CBER	Center for Biological Evaluation and Research
CDER	Center for Drug Evaluation and Research
CFR	Code of Federal Regulations
cGMP	Current Good Manufacturing Practices
CMC	Chemistry, Manufacturing and Controls
CQA	Critical to Quality Attributes
CVM	Center for Veterinary Medicine
DOE	Design of experiments
DModX	Distance to the model in the X-data
DS	Drug Substance
ERP	Enterprise Resource Planning systems
FDA	Food and Drug Administration
ICH	International Conference on Harmonization
LIMS	Laboratory Information Management system
MES	Manufacturing Execution system
MSPC	Multivariate statistical process control
MVDA	Multivariate Data Analysis
NDA	New Drug Application
ONDC	Office of New Drug Chemistry
OOC	Out of control
ORA	Office of Regulatory Affairs

Abbreviation	Meaning
QbD	Quality by Design
PAT	Process Analytical Technology
PAI	Pre-Approval Inspection
PCA	Principal Component Analysis
PLS	Partial Least Squares or Projection to Latent Structures

REFERENCES

1. Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach, Final Report, September 2004

2. Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, September 2004

3. Guidance for Industry Quality Systems Approach to Pharmaceutical Current Good Manufacturing Practice Regulations, September 2006

4. Guidance for Industry Q8, Pharmaceutical Development, ICH, May 2006

5. Guidance for Industry Q9, Quality Risk Management, ICH, June 2006

6. Guidance for Industry Q10, Pharmaceutical Quality System, ICH, April 2009

7. Guidance for Industry Q11, Development and Manufacture of Drug Substances, ICH, November 2012

8. Quality-by-Design, A Perspective from the Office of Biotechnology Products (OBP), http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4187B1_01_13-%20OBP-QbD.pdf , accessed on 10/20/13 at 10:41am

9. Guidance for Industry; Process Validation General Principles and Practices, January 2011, Revision 1

10. PharmaQbD.com, http://www.putman.net/brands/qbd/downloads/pharmaqbd_positioning_overview.pdf, accessed on 11/05/13 at 10:41am

11. SIMCA-P and SIMCA-P+ 11, User guide and Tutorial, Umetrics

12. Multivariate and Megavariate Data Analysis Basic Principles and Applications (Part I), L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, and S. Wold, Umetrics

13. PAT toolkit, The Leading Software for the PAT Initiative, Umetrics

14. Multivariate Analysis of Quality An Introduction, Harald Martens and Magni Martens, Wilfy Publications, February 2001

15. CFR - Code of Federal Regulations Title 21, FDA.gov

16. General Principles of Software Validation; Final Guidance for Industry and FDA Staff, Document issued on: January 11, 2002

17. Guideline on General Principals of Process Validation May, 1987 (REPRINTED MAY 1990)

18. Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application, August 2003

19. SIMCA User Guide, Software version 13.0.3

20. Springer Series in Statistics, I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer publications

21. A tutorial on Principal Components Analysis, Lindsay I Smith, February 26, 2002, 12-20

22. http://www.eso.org/sci/data-processing/software/esomidas//doc/user/98NOV/volb/node211.hmml

23. http://www.unt.edu/rss/class/mike/6810/Principal%20Components%20Analysis.pdf, 09/14/13 at 07:20am

24. Analysis of Multiblock and Hierarchical PCA and PLS Models, Journal of Chemometrics, J. Chemometrics, 12, 301-321 (1998)

25. http://origin-ars.els-cdn.com/content/image/1-s2.0-S0169409X06000883-gr3.jpg, accessed on 09/07/13 at 02:00pm

26. http://www.nature.com/nrgastro/journal/v5/n6/images/ncpgasthep1125-f3.jpg, accessed on 09/07/13 at 01:50pm

27. Online Batch/Fed-Batch Process Performance Monitoring, Quality Prediction, and Variable-Contribution Analysis for Diagnosis, Cenk Undey, Sinem Ertunc, and Ali Cinar

28. Nelson, P.R.C., Taylor, P.A., MacGregor, J.F. Missing data methods in PCA and PLS: Score calculations with incomplete observations, Chemom. Intell. Lab. Syst. 1996, 35(1), 45-65

29. Neogi, D.; Schlags, C.; Multivariate statistical analysis of emulsion batch process, Ind. Eng. Chem. Res. 1998, 37, 3971-3979

30. http://origin-ars.els-cdn.com/content/image/1-s2.0-S0952197607001133-gr1.jpg, accessed on 10/26/13 at 9:55am

31. http://statistics.ats.ucla.edu/stat/spss/output/spss_output_pca_5.gif, accessed on 11/03/13 at 5:40am

32. http://www.powershow.com/view/11cf54-ZTQyO/The_Least_Squares_Principle_powerpoint_ppt_presentation, accessed on 11/05/13 at 6:30am

33. http://en.wikipedia.org/wiki/ISO_9000, accessed on 11/05/13 at 10:41am

34. Process Analytical Technology , Edited by Katherine A. Bakeev, 13-37, Blackwell Publishing

35. A Beginner's Guide to Partial Least Squares Analysis, Michael Haenlein (Department of Department of Electronic Commerce, Otto Beisheim Graduate School of Management), Andreas M. Kaplan (Department of Media Management, University of Cologne)

36. Using PAT for Real Time Monitoring and Quality Assurance in Chemical Development, Pharmaceutical Review, June 2010

37. http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ucm103453.pdf

38. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/QualitySystemsRegulations/MedicalDeviceQualitySystemsManual/ucm122439.htm

39. PAT in Biologics Book (Eds. C.Undey, D. Low, J. Menezes, M. Koch) and the QbD in Biopharmaceuticals Book (Eds., A, Rathore)

MANUSCRIPT I: APPLYING QbD AND PAT IN BIOLOGICS MANUFACTURING, AND VALIDATION OF PROCESS MONITORING TOOL

Prakash Bedre (Principal Validation Engineer, Amgen Inc.); Cenk Undey (Director, Process Development, Amgen Inc., Clinton Chichester (Professor, University of Rhode Island)

ABSTRACT

Conventional pharmaceutical and biopharmaceutical manufacturing processes are accomplished using batch processing coupled with laboratory testing of off-line samples to determine the quality of the product prior to release. This approach is successful and can continue to be employed. However; due to recent technological developments, significant opportunities are available for improving the biological manufacturing processes. These opportunities include innovation in process development, process analysis, and process control by applying and using Quality by Design (QbD), Process Analytical Technology (PAT) and statistical tools. The regulatory agencies have given overwhelming support to these concepts throughout the last decade. They took initiative to support these innovation and technological developments by revising and approving guidance documents such as cGMPs for the 21st Century: A Risk Based Approach [1], PAT framework [2], International Conference on Harmonization (ICH) Q8, Q9, Q10 [3, 4, 5], Food and Drug Administration's (FDA's) 2011 guidance for Process Validation [6] and lastly ICH Q11 in 2012 [7]. These guidelines are continuously setting new industry trends as well as continuing to raise expectations. This study was focused on the use of a multivariate statistical data analysis tool for real-time process monitoring and its validation test cases to support good manufacturing practice (GMP) decisions. It also discusses case studies to demonstrate how a batch can be monitored using multivariate (MV) models to quickly identify out of trend results or batch failures real-time. Data from one biologics commercial manufacturing process is used for the creation of MV models to highlight the industrial application of the tool. The business benefits of implementing real-time multivariate statistical analysis in a GMP environment are also discussed.

45

INTRODUCTION

Most of the existing manufacturing processes for biological product at scale are not designed for real-time process monitoring. In order to enable real-time process monitoring capabilities, modifications to existing facilities need to be modified by installation of PAT with incorporation of QbD principles. In addition to technological modifications, various computerized systems must be implemented or modified for acquisition of large quantities of process data in real-time. Software applications to analyze this data and the statistical methods to process and generate the results are additionally required. This entire data acquisition, analysis and processing system is termed as "Real-Time Statistical Process Monitoring System".

A typical biologics manufacturing process is comprised of a series of upstream and downstream unit operations. The upstream manufacturing process (a.k.a. fermentation process) consists of a master cell bank vial thaw, cell line expansion in a shake flask to shaker bottles or bags, seed bioreactors, production bioreactors and centrifugations for harvesting protein. The downstream manufacturing process (a.k.a purification process) consists of various chromatography columns, viral inactivation, ultra-filtration/diafiltration (UF/DF), and viral filtration for purification and separation of therapeutic proteins prior to final fill-finish operations.

Each unit operation is comprised of multiple phases. Each phase is operated by multiple process parameters or variables. These parameters are categorized into input and output parameters. The input parameters are evaluated in the operating space and characterization studies and are maintained within the known operating ranges to achieve the desired output. However, the output parameters (a.k.a performance parameters) have a pre-set acceptance criteria to ensure that the process delivers consistent results every time.

For every biological process batch, there are many process variables measured during the course of production. It is important to make sure that each variable is operating within it operating range to ensure process performance consistency and product quality..

The suggested PAT framework using combination of following PAT tools includes:

I. Multivariate (more than one variable) data acquisition and analysis
II. Process and endpoint monitoring and control tools
III. Continuous improvement and knowledge management tools

The multivariate statistical process monitoring system efficiently monitors many variables at the same time by utilizing multivariate charts. The system also explains how these variables are changing in correlation with performance variables.

The goal of this study was to demonstrate how above stated PAT tools can be utilized to ensure continued process verification as outlined in FDA's 2011 process validation guidelines [6] for making critical manufacturing decisions in real-time.

In order to collect data for the multivariate data analysis (MVDA), modern process analyzers such as pH, temperature, agitation, dissolved O_2, CO_2, and cell density probes must be installed so that the process information can be gathered at regular intervals. Various software applications store and maintain this process data into databases enabling extraction of meaningful and critical process information from this data.

This study focused on the use of multivariate analysis tool as outlined in PAT framework by using SIMCA software to create multivariate models for real-time process monitoring. The data collection and data mining process is a critical step which required installation of multiple software interfaces for linking the software databases, database modifications, and creation of trigger tags, timers, batch tags and monitoring markers. The three dimensional data extracted from the databases must be unfolded and saved in a specific format so that it can be used by the SIMCA software for the creation of batch evolution and batch level models.

Various MV models are generated using the successful performance batches from the historical databases. Two popular and commonly used MVDA methods such as principal components analysis (PCA) and partial least squares (PLS) were employed to demonstrate the use of PAT tool [2, 9, 10]. Then new batches were tracked against these

models to ensure process consistency and detect deviations or process failures in real-time.

This study proved that PAT tools can be used to achieve continued process verification which meets the lifecycle approach in FDA's process validation guideline [6] and adaption of new mantra that the "Process validation should not be viewed as a one-off event. A lifecycle approach should be applied linking product and process development, validation of the commercial manufacturing process and maintenance of the process in a state of control during routine commercial production" [15].

MATERIALS AND METHODS

MATERIALS

SIMCA Software v13.0.3 [8]: The SIMCA software version 13.0.3 was utilized for the creation of statistical MV models. This software was developed by Umetrics.

This software was used to create a design of experiments and multivariate data analysis. This tool transforms data into information, which can be seen in the form of color coded graphical control charts to enable the process analyst to make correct decisions in real-time.

Historical data from consistent successful process batches (batches which have minimum deviations) were used from a biological manufacturing process. The data was extracted from the historical databases by making configurations, tags, scan rates, and compression settings to the source system. This data was pre-treated and organized in the appropriate format prior to importing into the SIMCA software for the creation of multivariate models.

METHODS

This study was conducted by using the data from a commercial biological manufacturing facility. The commercial biological manufacturing process was enabled with various modern process analyzers and was equipped with both a distributed control system (Delta-V) system and a plant data historian to collect the process data. The MVDA tool was linked to the process databases to acquire the process data. Modifications were made to the existing databases for the appropriate collection of data from the unit operations. The case studies presented were focused on one upstream unit operation (i.e., bioreactor) and one downstream unit operation (i.e. UF/DF).

The data set used for the bioreactor and UF/DF unit operations, to develop empirical models for multivariate monitoring purposes, were gathered from one of the existing products. The data were modified (normalized for propriety reasons) as necessary prior to using it for the creation of MV models in SIMCA.

The bioreactor and UF/DF unit operations were monitored using the parameters listed in Table 1 and Table 2. These unit operations were connected to the DeltaV system to collect the process data. The process database was connected to both DeltaV and a plant data historian and the historical and current batch data was saved continuously. The plant data historian was configured with the correct tags to enable advanced monitoring [11]. Along with online data collection, cell viability and viable cell density data were also collected from off-line measurements to check the process performance at every twenty-four hour interval during the course of the unit operation. The configuration of trigger tags, timers and batch tags were made as required to get all the relevant batch and continuous process data from the historical databases.

The on-line data was collected at 15 minutes intervals for the bioreactor unit operation and at 10 second intervals for the UF/DF unit operation. The data from fifteen batches were used for the creation of the MV model and control charts for the real-time statistical process monitoring. The historical batches which had lowest number of deviations and alarm conditions were selected as an input for the MV model creation to represent inherent process variability also known as common cause variation. The goal for this rational subgroup selection criterion was to create a reference MV model. The MV model created the average and ± 3 standard deviation confidence limits for the control charts about the scores and raw variable trajectories. Two new batches were projected over this model for testing the effectiveness of the real-time monitoring application. An artificial disturbance was introduced into one of the two new batch data generated by making changes to certain variables in order to test if the MV model could detect the changes. This study was used to prove if the real-time multivariate system can monitor the state of control for the new batches, detect any process failures due to sensor malfunction and end point monitoring.

A similar concept of creating MV models and process monitoring in real-time was utilized for all the unit operations. The deployment of this project into cGMP environment requires qualification and validation of the computerized systems per CFR Title 21 [12], Software Validation Guidance [13] and Part 11 Guidance [14]. The

validation of the real-time process monitoring system was performed enabling the system use in the cGMP environment to achieve continued process verification meeting FDA's process validation lifecycle approach.

ASSUMPTIONS

I. SIMCA software version 13.0.3 from Umetrics was utilized for the creation of statistical multivariate models. Therefore, the calculations performed for the generation of multivariate models in SIMCA software were correct because SIMCA was a commercial-of-the-shelf (COTS) application.

II. The SIMCA graphics and control charts calculated by the software were accurate.

III. The data used for MV model creation was collected from the validated commercial manufacturing process by creating batch identifiers, monitoring markers from historical database.

IV. The data collection and data mining method for each unit operation to feed into the SIMCA software for the creation of MV models is identical.

MV MODEL CREATION

In this section, the MVDA methods such as PCA and PLS were used to create batch evolution model (BEM) and batch level model (BLM) for the bioreactor and UF/DF unit operations for a legacy biological manufacturing process. The data set used for the bioreactor and Ultra –Filtration and Di-Filtration (UF/DF) unit operations to develop the empirical models for multivariate monitoring purposes was gathered from an existing biologics commercial manufacturing facility.

Data selection criteria:

The bioreactor and UF/DF unit operation are equipped with multiple online sensors as shown in Tables 1 and 2. This state of the art data collection capability enables real-time monitoring of the unit operations. The data collected from these sensors were preprocessed and arranged in a specific data format to import into the SIMCA software for the creation multivariate statistical process models for data acquisition and analysis.

MV for Bioreactor:

The dataset used in model building contained data for N=17 batches. Out of these, fifteen batches were selected for the creation of MV model. The batch selection criteria were to include a little variability among the batches used for the UF/DF process. The main objective of the case study was to create a MV model so that it could be used as a reference to monitor new batches in process and identify whether those batches are in multivariate control. Fourteen variables (J=14) were monitored and the data was collected at fifteen minute interval, giving a total of $\sim K = 279$ time points. The total duration of the unit operation was 68 hours, 30 minutes. The bioreactor unit operation was assumed as a single phase process.

MV for UF/DF: The dataset contained data for N=17 batches. Out of these, fifteen batches were selected for the creation of a MV model. The batch selection criteria were to include a little variability among the batches used for the UF/DF process. The main objective of the study was to create a MV model so that it could be used as a reference to monitor new batches in process and identify whether those batches were in multivariate control. Nineteen variables (J=19) were monitored and the data was collected at ten

second interval, giving a total of K = 817 time points. The total duration of the unit operation was 2 hours: 31 minutes: 10 seconds. The UF/DF unit operation was comprised of three phases which include concentration, diafiltration and recovery. Purpose of this unit operation is to concentrate and buffer exchange via tangential flow filtration. Following concentration and diafiltration, the product pool is recovered by filtering it through 0.2μ membrane filter.

Table 1: Bioreactor Variables Monitored

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Dissolved O_2	%	Operating	Input	Dissolved oxygen is monitored as a measure of cell culture performance
Culture pH	pH units	Operating	Input	pH affects final viable cell density
Air flow	sLPM	Operating	Input	Air flow is monitored as a measure of cell culture performance
O_2 flow	sLPM	Operating	Input	O_2 flow is monitored as a measure of cell culture performance
CO_2 flow	sLPM	Operating	Input	CO_2 flow is monitored as a measure of cell culture performance
Bioreactor level	L	Operating	Input	Bioreactor level is monitored to maintain consistent volume
Agitation	RPM	Operating	Input	Agitation maintains a homogenous solution and oxygen transfer to the cells
Vessel pressure	psig	Operating	Input	Pressure influences mass transfer and mitigates contamination
Temperature	°C	Operating	Input	Temperature (controlling probe) excursions can impact final viable cell density and viability
Culture duration	days	Operating	Input	Culture duration affects final viable cell density and is the maturity variable
Temperature probe difference (A-B)	°C	Operating	Input	Temperature probe difference (A-B) is monitored to detect equipment drift or malfunction
VCD (Viable Cell Density)	10^6 cells/mL	Performance	Output	VCD is monitored as a measure of cell culture performance.
Viability	%	Performance	Output	Viability is monitored as a measure of cell culture performance.

Table 2: UF/DF Variables Monitored

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Feed pressure	psig	Operating	Input	Feed pressure is monitored to detect excursions in pressure
Retentate pressure	psig	Operating	Input	Retentate pressure is monitored to detect excursions in pressure
Permeate pressure	psig	Operating	Input	Permeate pressure is monitored to detect excursions in pressure
Transmembrane pressure (TMP)	psig	Operating	Input	TMP is monitored to detect any excursions in pressure
Feed flow	LPM	Operating	Input	Feed flow is monitored to detect any excursions in flow rate
Retentate flow	LPM	Operating	Input	Retentate flow is monitored to detect any excursions in flow rate
Permeate flow	LPM	Operating	Input	Permeate flow is monitored to detect any excursions in flow rate
Feed Retentate DP	psig	Operating	Input	Feed retentate pressure is monitored to detect differential pressure
Permeate flux	L/hr/m2	Operating	Input	Permeate flux (permeate flow rate normalized by membrane area) is monitored to detect excursions in flow rate
Permeate UV (ultra violet)	AU	Operating	Input	Permeate UV is monitored to detect product loss during UF/DF II
Permeate conductivity	mS/cm	Operating	Input	Permeate conductivity is monitored to ensure product conductivity targets
Concentration factor	N/A	Operating	Input	Concentration factor is monitored to ensure concentration targets are met
Diafiltration Factor	L	Operating	Input	A minimum number of diavolumes are required to meet pH and conductivity specifications for the UF/DF II Pool

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Feed flow process totalizer	L	Operating	Input	Feed flow totalizer is monitored to ensure concentration and diafiltration targets are met
Permeate flow process totalizer	L	Operating	Input	Permeate flow totalizer is monitored to detect any excursions in the total permeate volume
Retentate flow process totalizer	L	Operating	Input	Retentate flow totalizer is monitored to ensure concentration and diafiltration targets are met and is the maturity variable
Permeate control valve	%	Operating	Input	Permeate control valve is monitored as a controller output
Retentate control valve	%	Operating	Input	Retentate control valve is monitored as a controller output
Step yield	%	Performance	Output	Step yield is monitored as a measure of UF/DF performance.

The data obtained from the above variables was pre-processed prior to use in the creation of the MV model. Multivariate methods are maximum variance projection methods. A variable with a large variance is more likely to be expressed in the modeling than a low-variance variable. In order to give equal weight to all the variables, the data from the variables required standardization. The unit variance scaling [16] was chosen for these case studies. For example, DO_2 values vary from ~ 30 to 90 ranges whereas Air flow values vary from 0 to 1 for bioreactor. If we do not perform scaling then the DO_2 variable will have very high variance and it will have an impact on the model as compared to Air flow. Therefore, the MV model without data scaling may not be accurate.

Two levels of batch monitoring were employed; the BEM and BLM were implemented by unfolding the three-way matrix into two-way matrix as shown in Figure 1a and 2b [17].

BEM: The goal of BEM was to develop a model of the desired batches and monitor new batches against this model to determine if they were evolving within the confidence limits. The data generated by a batch process was arranged in data blocks as shown in Figure 1a and 2b [17]. The batches were depicted as "I", variables depicted as "J" and time points depicted as "K". In order to execute BEM, the three way batch data table was unfolded in such a way that the direction of the variable was preserved as shown in Figure 1a [17]. The resulting two-way matrix then had "I *J" rows and K columns. Each row contains data points X_{ijk} from a single batch observation (batch I, time K and variable J). The regression was made against local batch time, the resulting PLS scores reflected linear t_1, t_2, and t_3 relationships to local time batch time. A PCA on the three way data matrix showed how the individual observations related to each other [10].

Each batch had a different completion time resulting in an unequal number of measurements [18]. SIMCA software resolved this issue by auto generating a dummy y variable reflecting relative local batch time which was a dummy variable called "$Time", to align the unequal batches to have the same length. The value of $Time predicted by the PLS model was used to indicate how far a batch has evolved [8]. The fifteen batches were used for the creation of the reference PLS model of X data matrix versus local time ($Time).
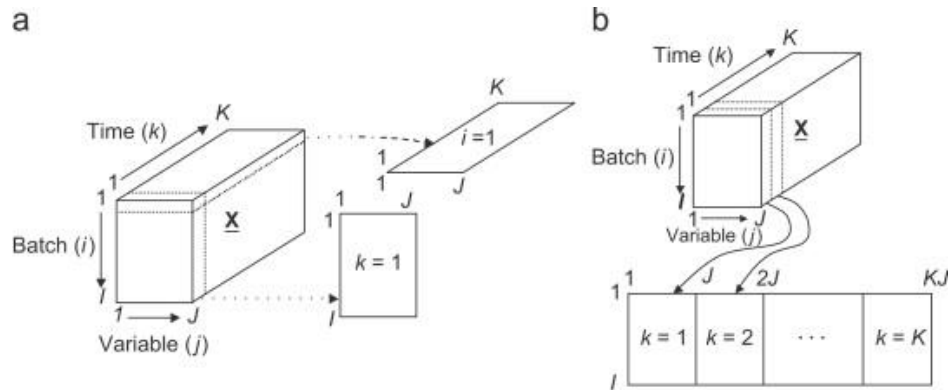
Figure 1a: Unfolding Three Way Data Matrix [Source: 17]. The three-way table of historical batch process data comprises "I", J variables and K time points. In the BEM, this three-way data table was unfolded by preserving the variable direction. This gave a two-way matrix with I*K rows and J columns. Each row contained data points X_{ijk} from a single batch evolution. Figure 1b: The three-way table of historical batch process data comprises "I", J variables and K time points. In the BLM, this three-way data table was unfolded by preserving the batch direction. This gave a two-way matrix with I rows and J*K columns. Each row contained data points from one single batch.

BLM: In batch level modeling, all the data are from input variable matrix (X) and output variable matrix (Y) is available. Therefore, the data from the whole batch is used to create a model. The aim of the batch level model is to verify whether the new batch is within multivariable control. The data generated is unfolded in such a way that the direction of the batch is preserved. The resulting two-way matrix then has "I*K" rows and J columns as shown in Figure 1b [17]. Another important objective of batch level model is to understand how Y (output) variable is influenced by the X (input) variables [10].

Diagnostics and Interpretation Controls Charts for the Bioreactor Unit Operation MV model:

BEM Model: The new BEM model created for the bioreactor unit operation shows that the model is fits well. The original variables when projected on these new components,

results into scores which are new co-ordinates that gets plotted as t_1, t_2, t_3 to a local batch time ($Time). This PLS model shows how the X-matrix input variables relate to each other in respective score plot and the loading plot Figure 2 and 3. The first principal component (PC) explains the maximum variation in X and then subsequent PC's explain the remaining variation in a descending order. The general expectation is that at least 75% to 85% of the variation must be accounted for by a good model [10]. This is because the scores of BEM model are used at the BLM to predict the output variable.

Scatter Plot t_1 vs. t_2 (Bioreactor):

The score plot is a map of the observations. Figure 2 displays the score plot for t_1 vs. t_2 scores. These are the new variables summarizing the X-variable matrix. The scores are orthogonal, (i.e., completely independent of each other). The score t_1 (first component) explains the largest variation of the X space, followed by t_2. The scatter plot of t_1 vs. t_2 is a window in the X space, displaying how the X observations are situated with respect to each other. This plot in Figure 3 shows the possible presence of outliers, similarities, and other patterns in the data.

The score plot shows that there is a strong behavioral similarity among the variables from all the batches. Each batch starts in the top left hand quadrant and ends in the top right quadrant.
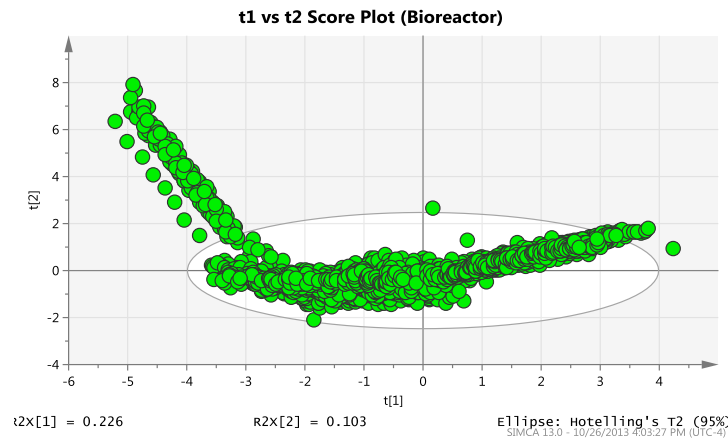

Figure 2: The Score Plot (BEM) for the bioreactor unit operation. The plot is created using the scores of first two principal components. The vertical axis

depicts t_2 scores and the horizontal axis depicts t_1 scores. The score plot shows that all the batches are aligned properly and fitting on 95% confidence ecllipse.

Loadings Plot W*C [1] vs. W*C [1]:

The loading plot shows this model can be furthered analyzed by Figure 3. It reveals that the response variable ($Time) is positively correlated with O_2 flow and Culture duration. The response variable is less correlated with Air Flow. There is a strong similarity between the behaviors of temperature, agitation, pressure and the bioreactor level. Other variables like CO_2 flow, and bioreactor level are negatively correlated.
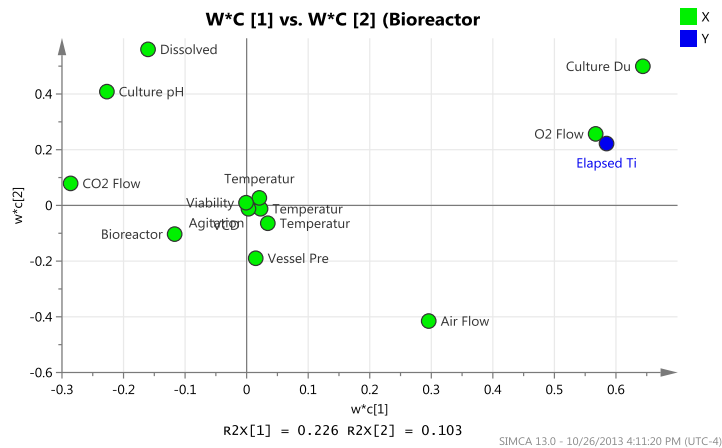

Figure 3: The Loading Plot (BEM) for the bioreactor unit operation. The plot is created using the loadings of first two principal components. The vertical axis depicts 'w*c [2]' loadings and the horizontal axis depicts 'w*c [1]' loadings. The plot shows how variables are correlated to each other.

The scores batch control chart displays that the selected score values (t_1) over time for all the 15 batches. The chart depicts the average batch (green) and the ±3 standard deviation (red). Figure 4 and 5 show the score contributions for scores t_1 and t_2. The t_1 score plot demonstrates that all the batches start with low scores and then increase steadily until the termination. Whereas, for t_2 scores, all of the batches move steady and end the same way. All of the fifteen reference batches behave well between the ±3 standard deviation and around the average for both t_1 and t_2 scores.

Figure 4: The PC 1 Score Batch Plot for the bioreactor unit operation. The fifteen batch plot for PC 1 show that all batches are ending in a similar fashion within the ±3 standard deviation confidence limit.
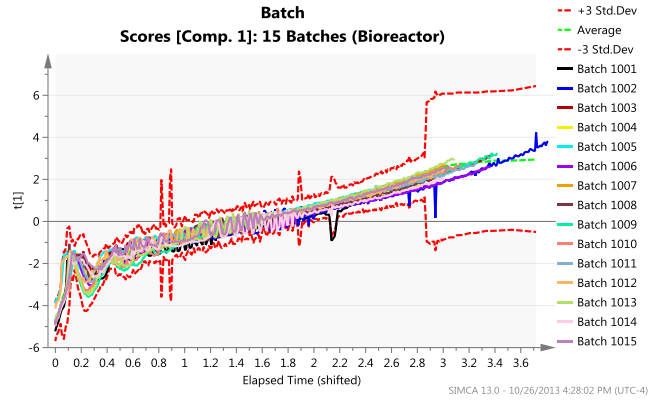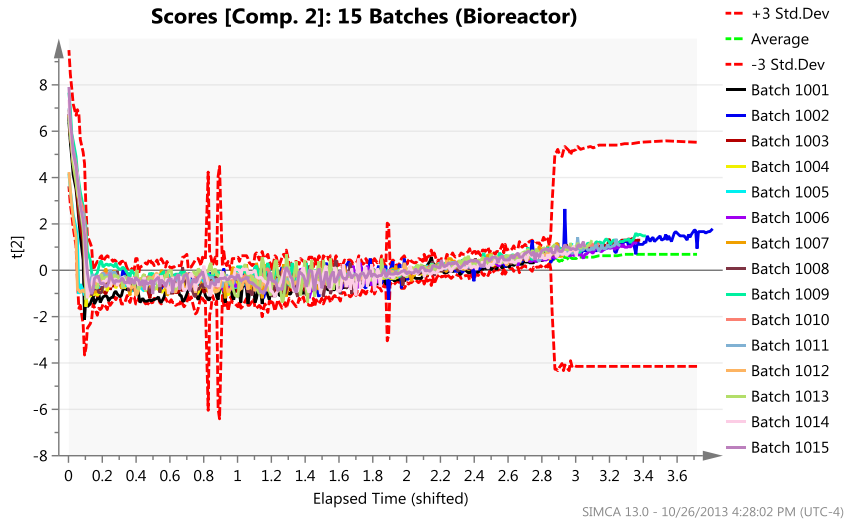

Figure 5: The PC 2 Score Batch Plot for the bioreactor unit operation. The fifteen batch plot for PC 2 shows that all batches are ending in a similar fashion within the ±3 standard deviation confidence limit.

Hotelling T^2 chart in figure 6 demonstrates that all the data from all the variables are within the score dimension.
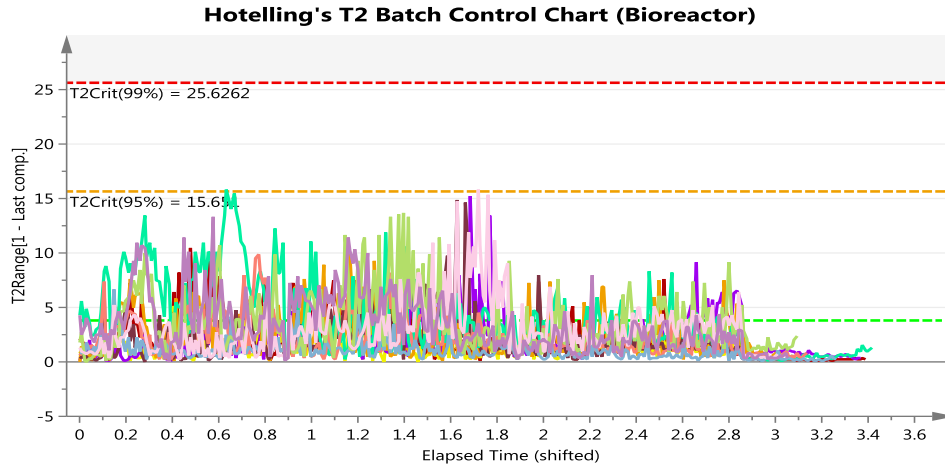
Figure 6: Hotelling T2 Plot for the bioreactor unit operation. The plot shows that all data from all variables is within the score dimension.

BLM Model: In batch level modelling, the entire batch data from all input and output variables are used to create the PLS model. In order to accomplish BLM modelling, the three-way batch data table is unfolded as shown in Figure 2b. The data is unfolded in such a way that each row in the data table represents one whole batch as shown in Figure 2b [17]. The unfolded data from X matrix (input variables) is regressed with unfolded data from the Y matrix (output variables). The data are arranged in such a way that the direction of the batch is preserved. The three-way data is arranged in two way matrix which had N*KJ rows (batches) and column (variables). There are two output variables (i.e. viable cell density and cell viability). The fifteen batches used for the creation of the reference PLS model of X data matrix versus Y data matrix can now be used to classify new batches still under development as to whether they are in multivariable control and see how Y variables are influenced by X variables[10]. The bioreactor batch level PLS model as shown in Figure 7, demonstrates that all batches are evenly scattered and are within the 95% confidence interval ellipse. Batch# 1001 is on the ellipse and is within the 95% confidence limit. The diagnostics charts, Hotelling's T^2 and DModX, charts as shown in (Figure 8 and 9) were well with within the confidence limits for all fifteen batches.
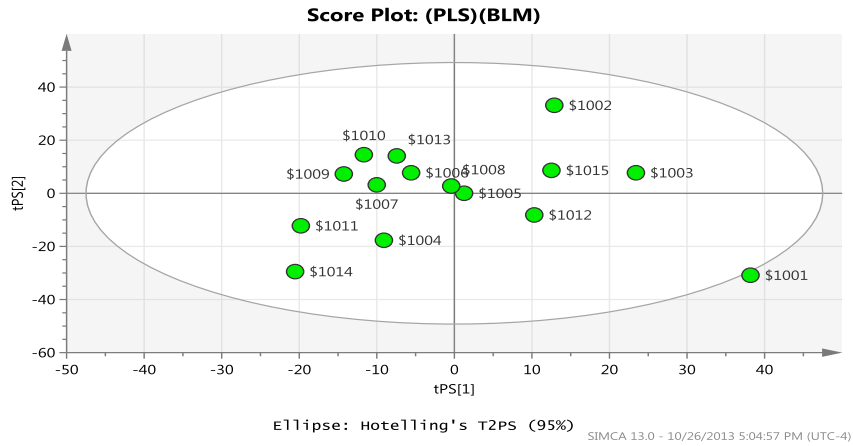
Figure 7: The Batch Level Score Plot for 15 Batches. The plot displays that all batches are well within the 95% ecllipse.
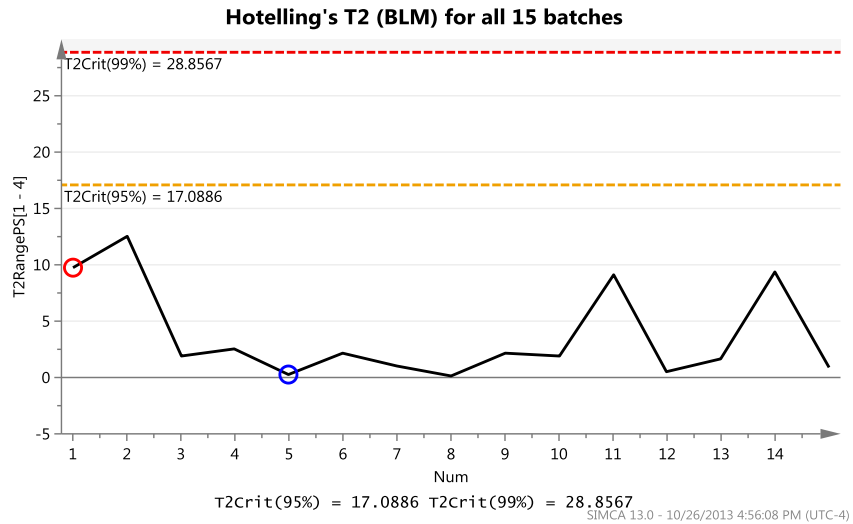

Figure 8: The Hotelling's T^2 Batch Level Plot for 15 batches. The plot displays that all batches are well within the 95% confidence limit.
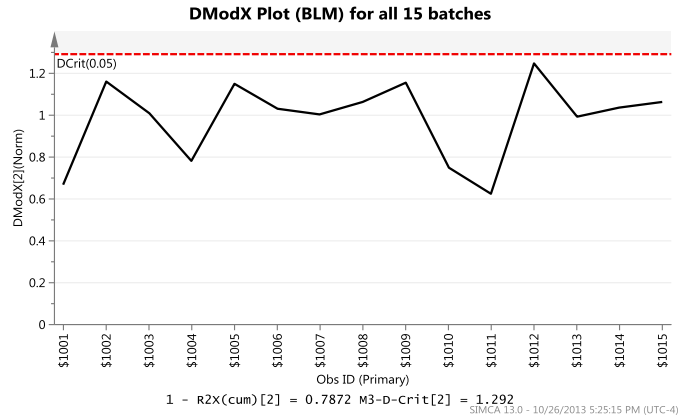
Figure 9: The DModX Batch Level Plot for 15 batches. The plot displays that all batches are well within the critical limit of 0.05.

Diagnostics and Interpretation Controls Charts for MV model for UF/DF:

BEM Model: The new BEM model created for UF/DF unit operation shows that the model is fitted well for concentration, diafiltration and recovery phases. The original variables when projected on these new components, results in scores which are new variables that gets plotted as t_1, t_2, t_3 to a local batch time ($Time). This PLS model shows how the X-matrix input variables relate to each other in both a respective score plot and the loading plot (Figure 10) for concentration phase.
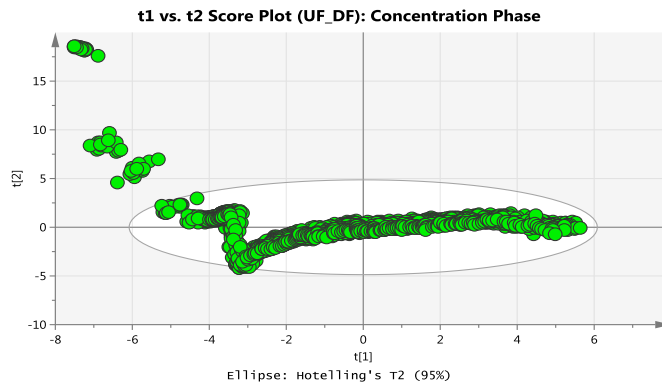

Figure 10: The Score Plot for Concentration Phase for the UF/DF unit operation. The score plot is created using the scores of the first two principal components. Vertical axis depicts t_2 scores and horizontal axis depicts t_1 scores. The score plot shows that all the batches are aligned properly and fit the 95% confidence limit ecllipse.

64

Loadings Plot W*C [1] Vs. W*C [1]:

The loading plot demonstrates that this model can be further analyzed by interpretation of Figure 11. It reveals that the response maturity variable ($Time) is positively correlated with the retentate flow process totalizer, permeate flow, process totalizer, feed flow and concentration factor. These variables steadily increase with time. All other variables maintain somewhat steady state and are reasonably co-related with each other except for retentate pressure which is negatively correlated as it shows a decreasing trend with time.
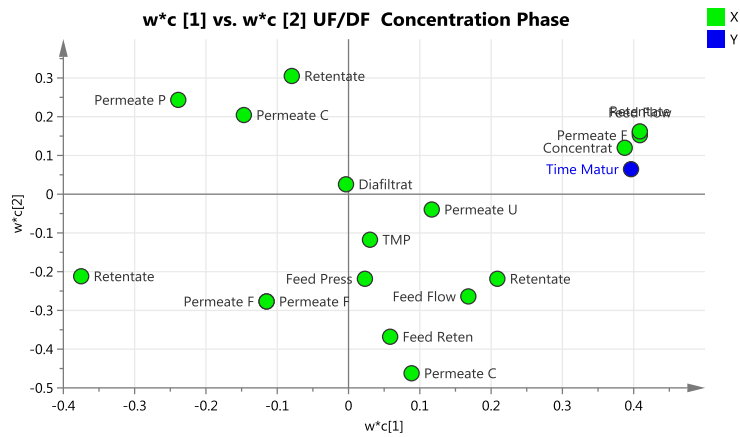
Figure 11: The Loading Plot for UF/DF. The BEM plot is created using the loadings of first two principal components. The vertical axis depicts 'w*c [2]' loadings and the horizontal axis depicts 'w*c [1]' loadings. The plot shows how the variables correlate to each other. Variables like retentate flow process totalizer permeate flow process totalizer, feed flow, and concentration factor are strongly correlated with the maturity variable. Retentate flow control negatively correlates with the maturity variable.

The scores batch control chart displays the selected score value (t_1) over time for all fifteen batches. The chart also shows the average batch (green) and the ±3 standard deviation (red). Figure 12 shows the score contribution for t_1 scores. The t_1 score plots demonstrate that all batches start with low scores and then steadily increase until termination. All of the fifteen reference batches behave well between ±3 standard deviation and around the average for first PC.

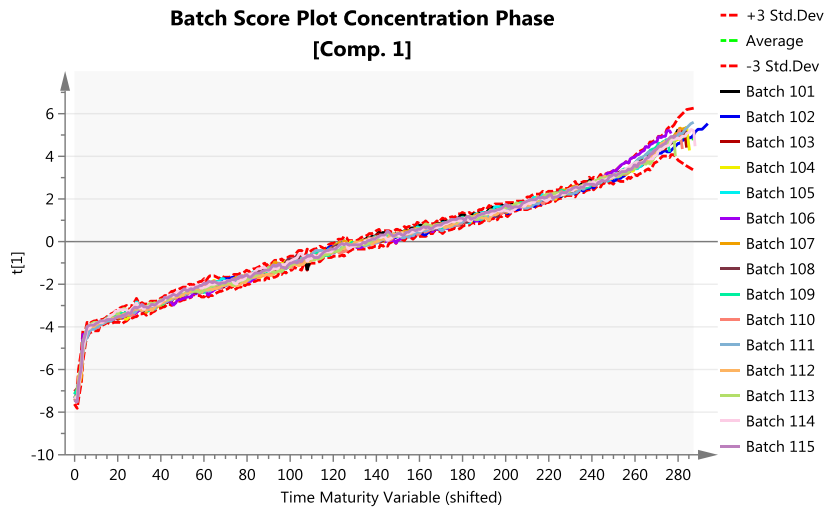Figure 12: The PC 1 Score Batch Plot for the UF/DF unit operation. The plot shows that all fifteen batches are ending in a similar fashion within ±3 standard deviation confidence limit.

BLM Model: In batch level modelling, the entire batch data from input variables and output variables are used to create the PLS model. In order to accomplish BLM modelling, the three-way batch data table is unfolded as shown in Figure 2b [17]. The data is unfolded and arranged as explained in the bioreactor BLM section. In the case study for UF/DF, the output variable is step yield. The fifteen batches used for the creation of the reference PLS model of X data matrix versus Y data matrix can now be used to classify new batches that still under development and demonstrate how Y variables are influenced by X variables[10]. The UF/DF batch level PLS model as shown in Figure 13, using the first two components demonstrate that all the batches are evenly scattered and are within the 95% confidence interval ellipse. When the diagnostics charts like Hotelling's T^2 and DModX are evaluated, charts, (Figures 14 and 15), are well with within the confidence limits for all fifteen batches.

66

Score Plot (BLM) for UF/DF

Ellipse: Hotelling's T2 (95%)
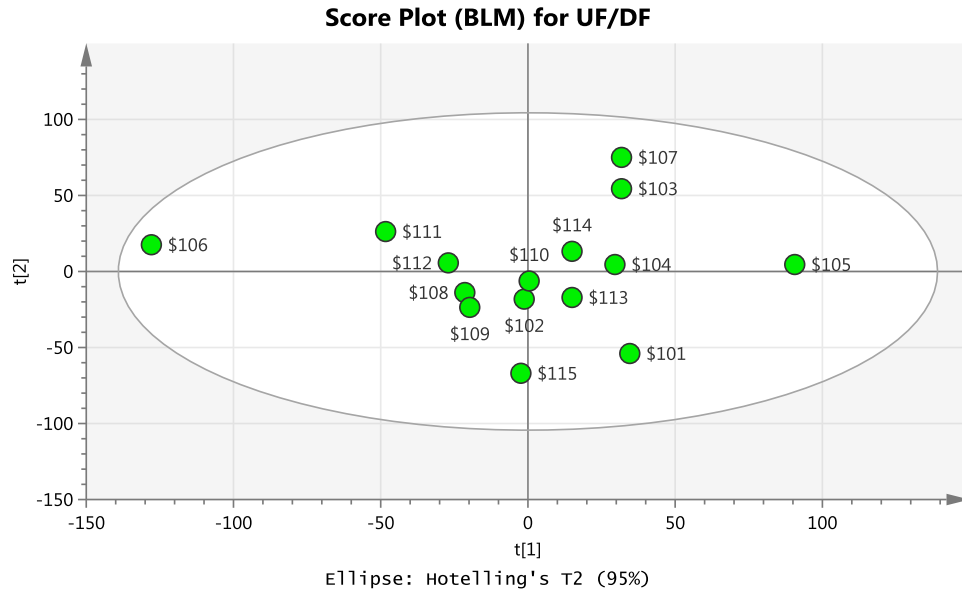
Figure 13: The UF/DF Batch Level Score Plot for 15 batches. The plot shows that all batches are well within the 95% ecllipse.

Hotelling's T2 (BLM) for all 15 batches

T2Crit(99%) = 20.8338

T2Crit(95%) = 12.216

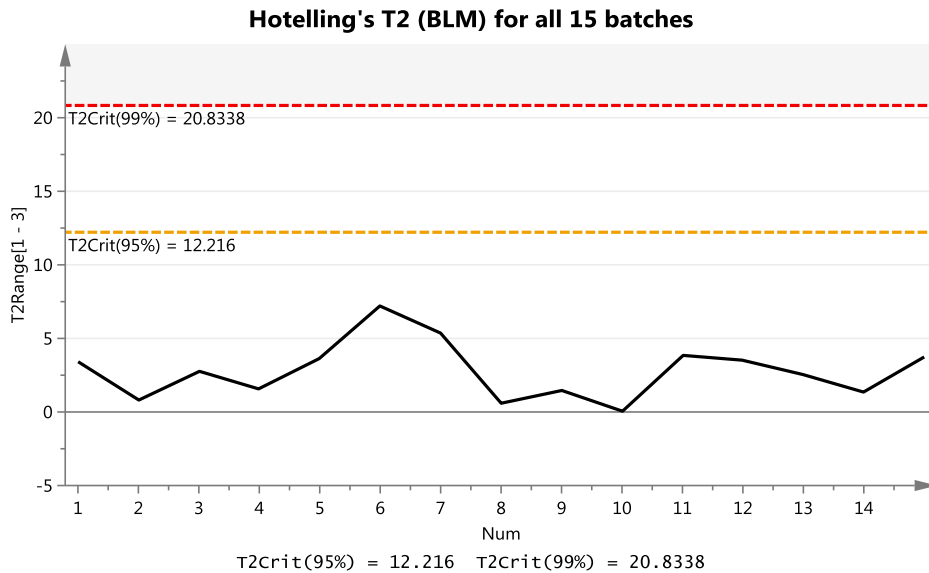T2Crit(95%) = 12.216   T2Crit(99%) = 20.8338

Figure 14: The UF/DF Hotelling's T^2 Batch Level Plot for 15 batches. The plot demonstrates that all batches are well within the 95% confidence limit.

67

DModX Plot (BLM) for all 15 batches

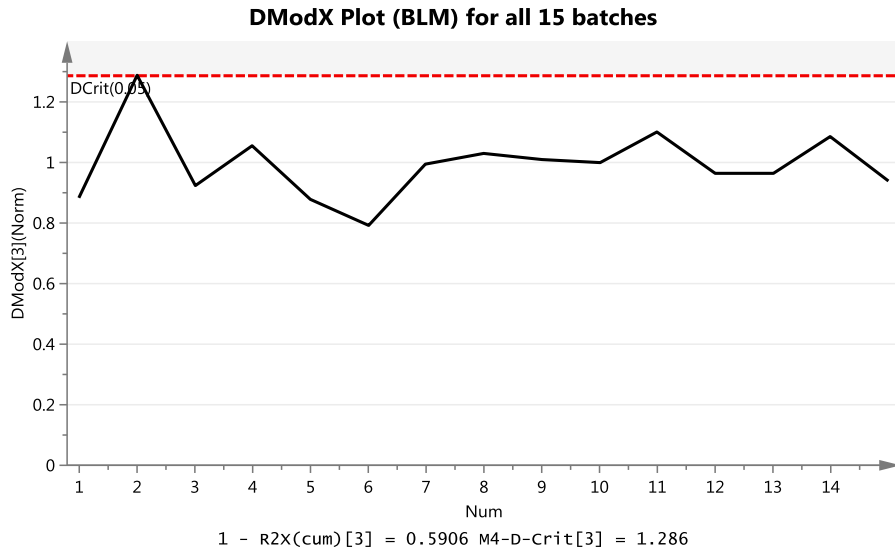1 – R2X(cum)[3] = 0.5906 M4-D-Crit[3] = 1.286

Figure 15: The UF/DF DModX Batch Level Plot for 15 batches. The plot demonstrates that all batches are well within the critical limit of 0.05.

In the previous section, the multivariate BEM and BLM models were created by using fifteen good batches. The control limits and the averages were established using the data from the model batches. The following sections will demonstrate how these models can be used to monitor the new batches in process. If the new batch is a good batch then it is expected to evolve with in the confidence limits. This can be seen on the control charts. At the same time, if for any reason, there is a deviation then the cause and the source of the deviation can be tracked using the control charts. Introduction of this tool into the GMP manufacturing facility equipped with on-line data collection technologies can monitor every new batch in real-time to provide continuous improvement data as well as a scientific knowledge management opportunities and additionally meeting the FDA's third stage, continued process verification life cycle approach.

In order to test the real-time process monitoring system to ensure continued process verification, two new batches (Batch# 1016 and 1017 for the bioreactor unit operation and batch# 116 and 117 for the UF/DF unit operation) were selected for each of the bioreactor and UF/DF unit operations. One new batch selected for the study was a good batch and the second new batch was deliberately modified (after the fact for simulation purposes) by making changes to feed pressure and feed flow to determine if the changes

68

can be detected by the real-time process monitoring system. Case studies were used as an example to demonstrate how the real-time process monitoring system can be used in the manufacturing process for continued process verification, detect process failures in the new batches due to sensor malfunction or process related failures and root cause analysis and process control.

CASE STUDY 1: MONITORING STATE OF CONTROL FOR A NEW
BIOREACTOR PROCESS BATCH

In this example, a new batch (batch# 1016) was projected on the MV model created using
fifteen good batches to see if this batch was running in the state of control. The batch
score plot in Figure 16 shows that the batch# 1016 shown in dark red color was well
within the 95% confidence limit. The score plot shows that batch# 1016 was a good
batch.
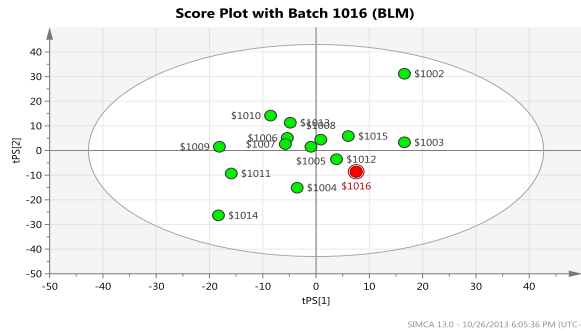
Figure 16: The Batch Score Plot for bioreactor batch# 1016. Batch# 1016 was
projected on the model with other batches. Batch# 1016 was highlighted in red.
The score plot shows that batch# 1016 was a good batch because it was well
within the 95% confidence limit.

The batch control score plot for batch# 1016 in Figure 17 shows that the batch was within
± 3 standard deviation. Therefore, it was a good batch.
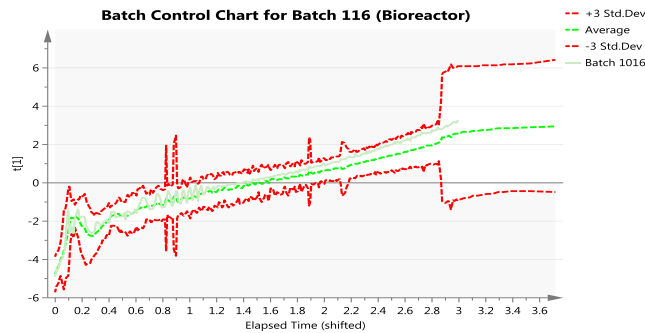
Figure 17: The Batch Control Plot for bioreactor Batch 1016. It shows that it was
within ± 3 standard deviation. Therefore, it was a good batch.

CASE STUDY 2: DETECT SENSOR MALFUNCTION FAILURE FOR NEW
BIOREACTOR PROCESS BATCH

The intent of this test was to determine if a sensor malfunction can be detected by the
MV model. The faulty sensor can show the incorrect values while the batch was
evolving. The pH and temperature raw data from the batch# 1016 was modified to show
that the pH and temperature sensors were faulty. The pH and temperature raw data were
deliberately changed at several time points in the excel datasheet prior to importing it into
the SIMCA application to project on the MV model. The new batch (batch# 1017) was
then projected on the MV model. The batch score plot in Figure 18 and batch control
score plot in Figure 19 clearly show that batch# 1017 was outside the model space.
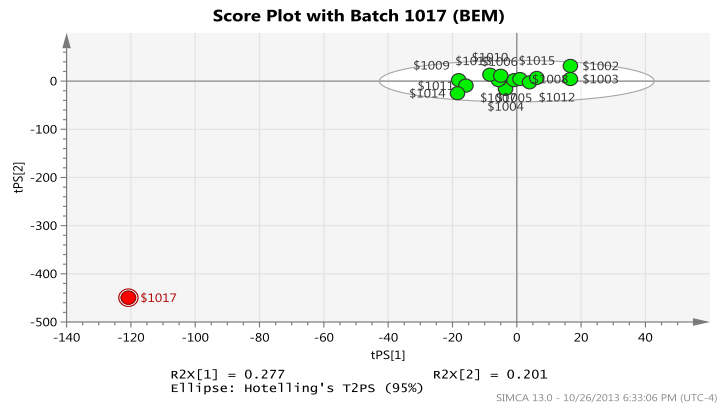
Figure 18: The Batch Score Plot for bioreactor batch# 1017. Batch# 1017 was
shown in red and was outside of the 95% confidence limit.
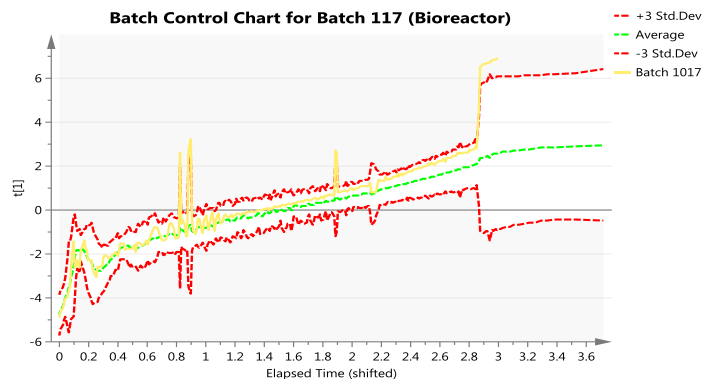
Figure 19: The Batch Control Plot for bioreactor batch# 1017. The control plot
shows that it was going out of $\pm$ 3 standard deviation confidence limits at several
places.

The variable batch plots for culture pH in Figure 20 and for temperature in Figure 21 demonstrate the time points when the batch was out of confidence limits.
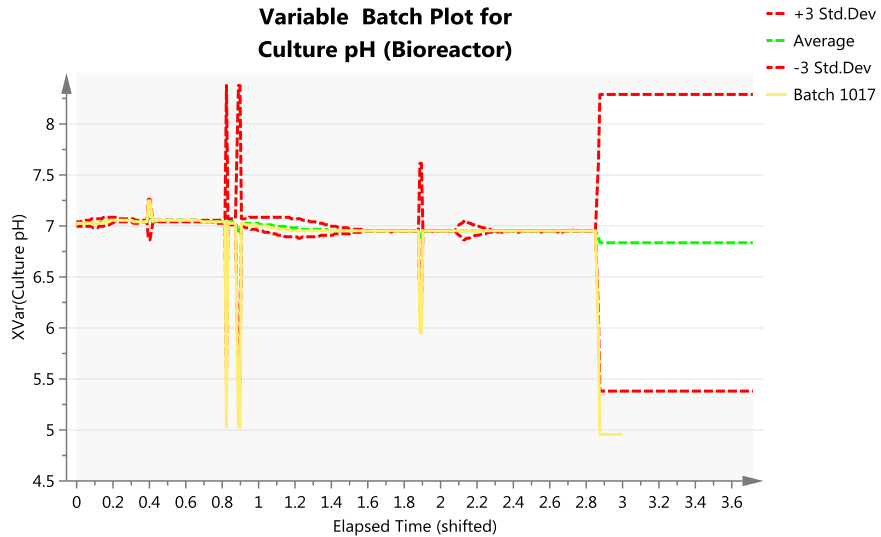
Figure 20: Culture pH Variable Batch Plot for batch# 1017. The plot shows the exact time when the batch was outside of the ± 3 standard deviation confidence limit.
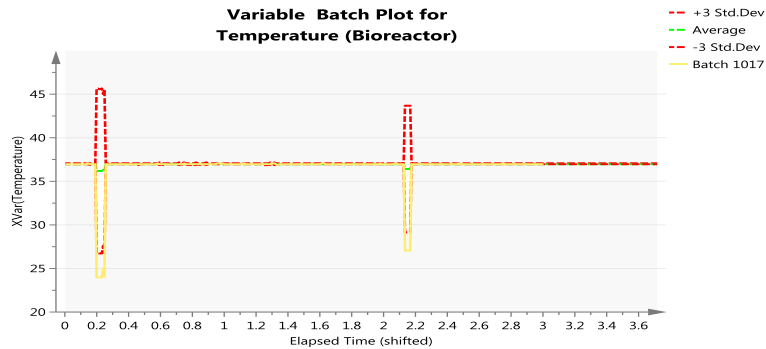
Figure 21: Temperature Variable batch Plot for batch# 1017. The plot shows the exact time when the batch was outside of the ± 3 standard deviation confidence limit.

Hotelling's T^2 and DModX charts in Figure 22 and Figure 23 also show that batch# 1017 depicted in yellow color was outside of the ± 3 standard deviation and Dcrit (0.05) shown in red. The variable batch plot demonstrates that the time points when the pH and the temperature variables were outside of the confidence limits.

72
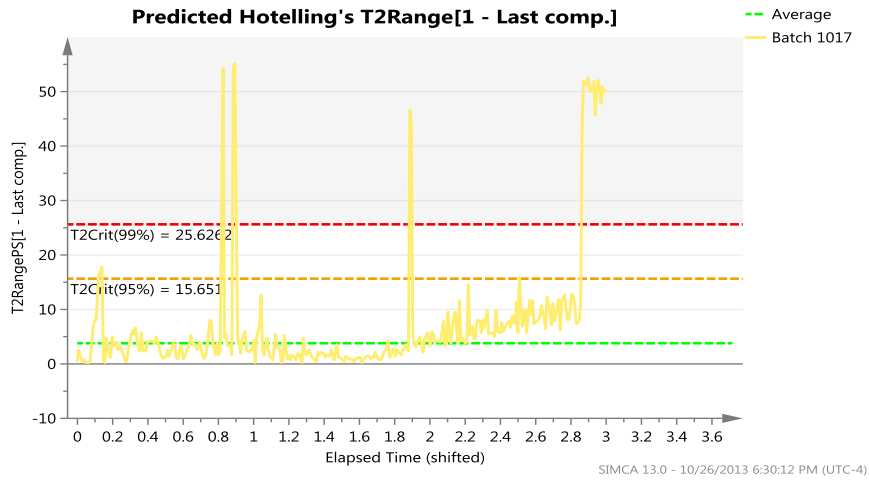
Figure 22: Hotelling's T^2 Plot for 1017. The plot shows the time when the batch was outside of the 95% critical limit.
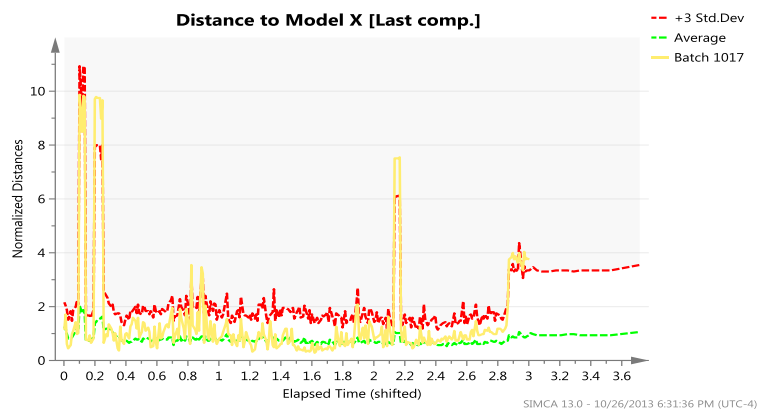
Figure 23: DModX Plot for 1017. The plot shows the exact time when the batch was outside of ± 3 standard deviation confidence limit.

CASE STUDY 3: MONITOR THE STATE OF CONTROL FOR A NEW UF/DF
BATCH

In this example, a new batch (batch# 116) was projected on MV model created using
fifteen successful batches to determine if this batch was running in a state of control. The
batch score plot in Figure 24 shows that the batch# 116 shown in dark red color was
moving within the ± 3 standard deviations shown in red and around the average shown in
green.
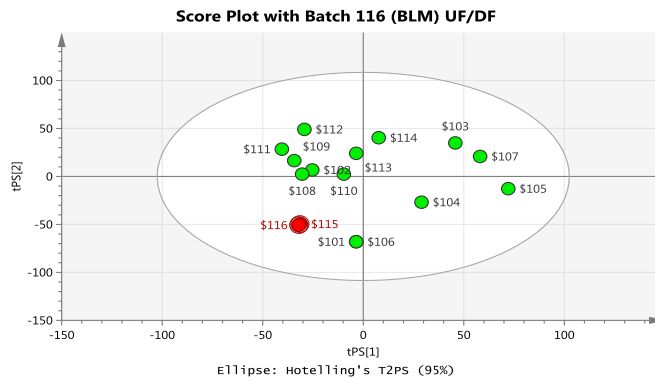
Figure 24: Batch Score Plot for UF/DF batch# 116. The plot was projected on the
model with other batches in the model and highlighted in red. The score plot
shows that batch# 116 was a good batch because it was well within the 95%
confidence limit.

Batch control plot for batch# 116 in Figure 17 shows that the batch was within the
confidence limits. Therefore, it is considered a good batch.
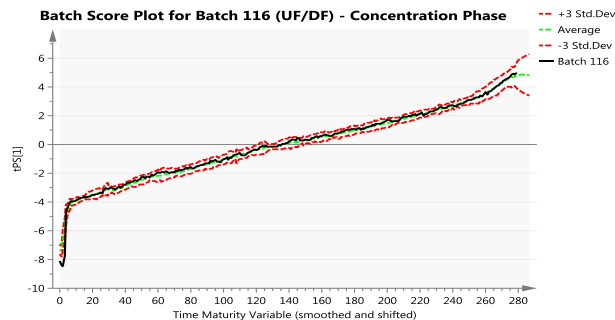
Figure 25: The Batch Control Plot for UF/DF batch# 116. The plot shows that
batch 116 was within ± 3 standard deviation. Therefore, it was a good batch.

74

CASE STUDY 4: DETECTION OF SENSOR MALFUNCTION FAILURE FOR NEW
IN A UF/DF BATCH

The intent of this test was to demonstrate if a sensor malfunction can be detected by the
MV model. The faulty sensor can show the incorrect values while the batch was
evolving. The feed pressure and feed flow raw data from batch# 116 was modified to
show that the feed pressure and feed flow were faulty. The feed pressure and feed flow
raw data were deliberately changed at multiple time points (after the fact for simulation
purposes). The new batch# 117 was then projected on the MV model. The batch
contribution score plot in Figure 26 clearly showed that batch# 117 was going outside of
± 3 standard deviation confidence limits at several places. The batch variable contribution
plot for feed pressure in Figure 27 showed the exact time point where the variable was
outside of the of ± 3 standard deviation confidence limits.
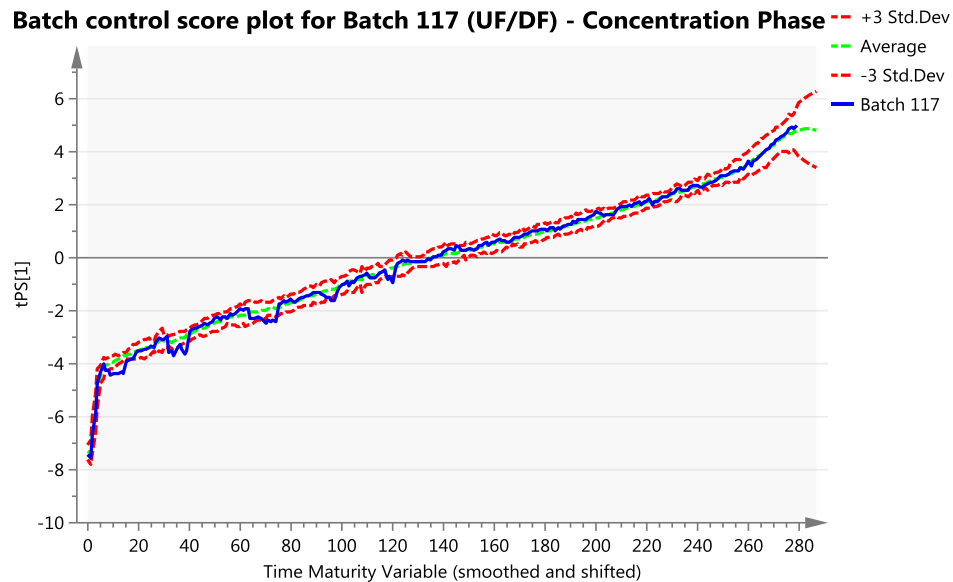
Figure 26: Batch Control Score plot for UF/DF batch# 117. The plot shows
batch# 117 is going outside of ± 3 standard deviation at several places.

The variable batch plots for feed pressure in Figure 27 and for feed flow in Figure 28 showed the exact time points when the these two variables were outside of ± 3 standard deviation confidence limits.
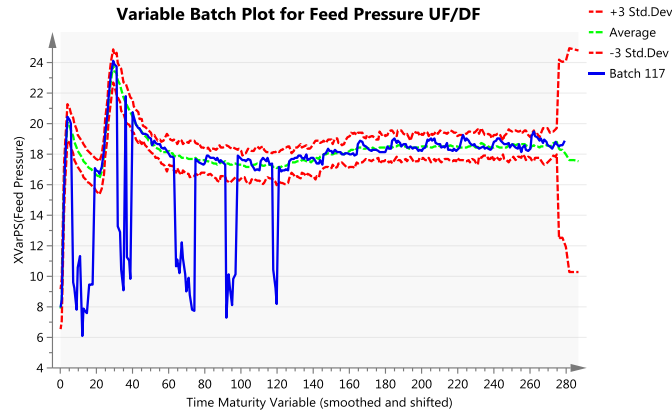
Figure 27: Feed Pressure Variable Batch Plot for 117. The plot shows the time points when the feed pressure exceeded the ± 3 standard deviation confidence limits.
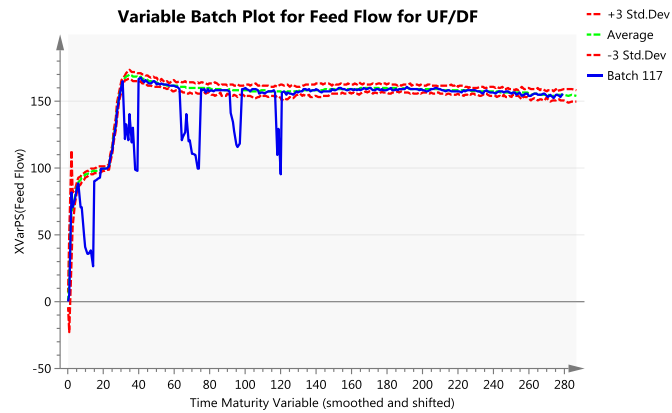
Figure 28: Feed Flow Variable Batch Plot for 117. The plot shows the time when the feed flow exceeded the ± 3 standard deviation confidence limit.

VALIDATION OF MULTIVARIATE STATISTICAL MONITORING SYSTEM

An existing commercial manufacturing facility can be PAT enabled by installation of state of the art on-line sensors, data management system and data analyzing computer hardware and software systems. This vast amount of data collected from a ten second interval to fifteen minute intervals from different unit operations is saved in databases. The entire system consists of multiple data servers, network components, software interfaces, and software applications for multivariate analysis. In this study, the SIMCA software was used for the multivariate model creation which was connected to the network via several related software interfaces. All together this entire system becomes an automated computerized system. As per FDA's software validation guidance [13], prior to using any computerized system in a cGMP environment, it must be qualified and validated. The cGMP guideline outlines that "Any software used to automate any part of the device production process or any part of the quality system must be validated for its intended use, as required by 21 CFR §820.70(i)". Validation is necessary to establish documented evidence to provide a high degree of assurance that the system will consistently operate according to pre-defined requirements and design specifications [12].

The SIMCA software is Part 11 compliant software. In order to utilize it in the cGMP environment, it must undergo validation per Computer Validation Guidelines [13] and Part 11 guidelines [14]. For the validation of a computerized system, several documents are generated, executed and approved by appropriate stake holders in the GMP facility. All validation related documents and the test cases for the validation of real-time process monitoring system are generated per software validation and Part 11 guidance.

The following section outlines the documents and test cases generated for the validation of the real-time process monitoring system.

<u>Validation Plan</u>

The purpose of the Validation Plan (VP) was to define the overall validation approach, roles and responsibilities, required deliverables, test strategy, and key decisions related to the validation. The VP provides the basis for scheduling of the validation activities and documentation was required in order to meet the acceptance criteria specified in the Requirements Specification and Design Specification documents. It also outlines test documents such as Installation Qualification (IQ), Operational Qualification (IQ), and Performance Qualification (PQ) for documenting the installation verifications of the modifications/changes to historical databases, software interfaces, database configuration, server configurations, network connections and SIMCA software. The VP also outlines the requirements specifications (RSs) and design specifications (DSs), operation and performance testing documents, requirement traceability matrix (RTM), and validation summary reports (VSRs).

When any validation project is rolled out on a large scale in the GMP environment, IS, automation, project management, process development, manufacturing, validation and quality assurance (QA) teams are involved. Therefore, it is very critical to outline the roles and responsibilities of every group for tracking and successful implementation of these tasks.

<u>Validation Strategy</u>

Validation of the MVDA system was implemented to establish documented evidence that the system will consistently provide reliable and robust data for monitoring of manufacturing purposes. The main deliverables that specify and document the testing of the system were RS, DS, IQ, OQ, and PQ documents. The RTM and VSR were written at the end of validation testing. The RTM demonstrated that the requirements were mapped to the RS/DS and to the corresponding test cases in the validation protocols. The VSR was generated to outline the completion of validation deliverables.

A risk management process outlined in Figure 32, GAMP-5 [19] is utilized while determining the extent of testing performed to validate the MVDA system.

Figure 29: Five 'Risk Management Process' steps [Source: 19] shows how the risk assessment is performed to determine the extent of validation testing required.

There are two basic classes of software testing: black box testing and white box testing [20]:

- Black box testing (also called functional testing) is testing that ignores the internal mechanism of a system or component and focuses solely on the outputs generated in response to selected inputs and execution conditions.

- White box testing (also called structural testing and glass box testing) is testing that takes into account the internal mechanism of a system or component.

A combination of black box and white box testing methods were used to test the real-time process monitoring system which was focused on the functional requirement of the system and was based on external characteristics of the program being tested. Validation used the following types of testing:

Intrinsic: Features are qualified intrinsically as other features are tested. For example, transition function is intrinsically tested when two consecutive steps of a sequence are allowed to run without stopping.

Inspection: Features are qualified by ensuring their presence. For example, file lists and other documentation can be confirmed by ensuring their existence and accuracy.

Structural testing: Features are qualified by testing individual components as specified in the RD (Requirements and/or Design document). Testing will ensure that each requirement stated in the RD is made to execute during testing and that each requirement stated in the RD performs its intended function.

Functional testing: All the hardware used for this system is a standard hardware, which is subject to IQ to verify the installation and connection to components. Software category 3 will be subjected to the validation process to ensure it meets the requirement specifications and design specifications. Testing will ignore that the internal mechanism or structure of a system or component and focuses on the outputs generated in response to selected inputs and execution conditions.

In March of 1997, FDA issued final part 11 regulations that provide criteria for acceptance by FDA, under certain circumstances, of electronic records, electronic signatures, and handwritten signatures executed to electronic records as equivalent to paper records and handwritten signatures executed on paper [1].

SIMCA is compliant with 21 CFR Part 11 (Electronic Records). "Umetrics quality systems for software development and validation can be audited". The audit trail is administrator-controlled and check-sum protected. SIMCA-4000 is OPC certified by the OPC Foundation. [21]. Therefore, the test cases related electronic record was performed to ensure the Part 11 compliance.

Validation Limitations and Assumptions:

All test cases assumed that the components, systems and services of servers were operating as expected. The test cases also assumed that the complex calculations performed during the generation of multivariate models in SIMCA software were correct and accurate because SIMCA is COTs software. Therefore, the multivariate statistical calculations were not verified. Testing conducted to verify one unit operation from the upstream (bioreactor) and one unit operation from downstream (UF/DF) for data flow from historical databases to SIMCA software for MV model creation was assumed to work exactly the same way for every unit operation.

Requirement and Design Specifications [13]

Every computerized system and software is developed or designed based on its intended use. While designing the system, the developers must know the specific requirements of the user (a.k.a 'User Specific Requirements'). The designed system also must meet certain inherent capabilities for it to function and meet the user specifications (a.k.a 'Functional Specific Requirements'). In order to validate the computerized system and software in the cGMP environment, these URSs and FRSs were required to be tested.

There can be many different kinds of requirements (e.g., design, functional, implementation, interface, performance, or physical). Software requirements are typically derived from the system requirements for those aspects of system functionality that have been allocated to software. Success in accurately and completely documenting software requirements is a crucial factor in successful validation of the resulting software. There are also many different kinds of written specifications (e.g., system requirements specification, software requirements specification, software design specification, software test specification, software integration specification, etc.). All of these documents establish "specified requirements" and are design outputs for which various forms of verification are necessary [13].

The list of URS and FRSs were taken from the software manuals provided by the Umetrics and other hardware providers. They were documented in a 'Requirement Specification' and 'Design Specification' documents.

Validation Installation Qualification, Operational Qualification and Performance Qualification (IQ/OQ/PQ) [13]:

Validation is establishing documented evidence that system or software is installed as per the designed specifications; it operates as per the functional requirement specifications and performs per the user specifications [22].

The real-time process monitoring system was validated per the traditional validation life cycle approach outlined in the Good Automation Manufacturing Practices (GAMP) guidance [23]. As outlined in the Figure 33 [25], IQ, OQ and PQ documents were written with specific test cases as listed in Table 3. The each test script in the validation protocol was executed to ensure that it meets the expected results.
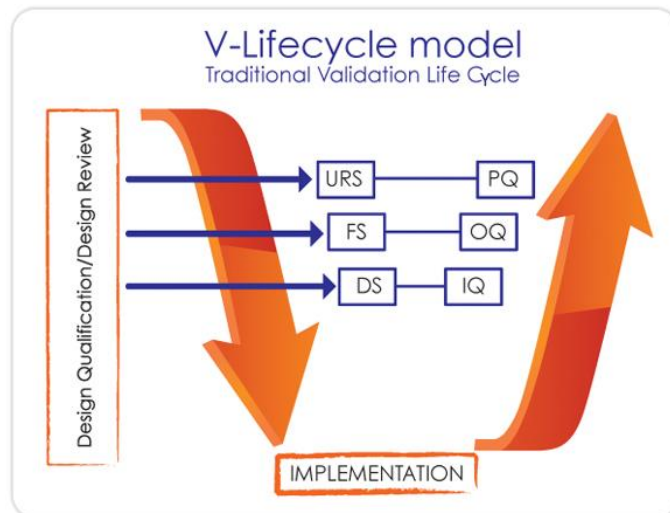

Figure 30, V- Lifecycle model, [Source: 25] shows that the validation approach utilized for software related systems. The user and function specification requirements are tested in the validation protocol and are tracked in the requirement traceability matrix.

Table 3, Test Cases in the Validation Protocol [13, 14, 19]

Test Case #	Description	Test Intention
1	Change Control Status verification (IQ)	Prior to implementing any GMP system, the change control needs to be initiated
2	Server Qualification verification	Ensure that servers are configured and qualified to meet the functional
3	Hardware & Software Installation verification (IQ)	Ensure that all the hardware, software interfaces, database interfaces, Operating system, and maintenance software's are installed and configured correctly,
3	Hardware configuration verification (IQ)	Ensure that all the hardware meet the processor speed, memory, graphic and driver settings
4	Creation and installation of MV model and Projects (IQ)	Ensure that all the MV models are created and saved correctly at a secured location
5	Documentation verification	Ensure that all the hardware, software related manual, drawings are documented and attached or achieved.
6	Pre-requisite documentation	Ensure that all the OQ pre-requisite documentation such as VP, RS and DS are approved prior to initiating OQ and PQ testing
7	Standard Operating Procedures (SOP)	Ensure that all the SOP's are created for the operation of the GMP system.
8	Start up and shutdown verification (OQ)	Each system component (server, network PC's, network devices are connected properly and go through flawless reboot process in case of power outage or routine start and shut down process
9	Logical Security for the Operating system and the Software's (OQ)	Ensure that the servers and the PC's which are used to operate the system have the restricted access so that unauthorized users cannot modify or delete the secured data folders/files or modify the recording time.
10	Logical Security for the Software's (OQ)	Ensure that the SIMCA software has the restricted access so that unauthorized users cannot modify or create new folders/files or modify existing folders/files
11	Password Security (OQ)	Ensure that the users can create unique passwords with specific length, alpha-numeric combinations and allow only certain number of attempts. Only the application or IS administrator is allowed to add the users or reset the passwords.

Test Case #	Description	Test Intention
12	Authority & Privilege levels (OQ)	Ensure that there are different authority levels and the user permissions for each group so that each unique group can maintain different functions while using the software
13	Operational Sequence and data flow (OQ)	Ensure that all the required connections are installed and operate correctly in the intended sequence.
14	Control Charts analysis and visualization of graphical plots verification(OQ)	Ensure that the PCA and PLS MVDA control charts for BEM and BLM can be seen correctly with set confidence limits and average.
15	Data Export and Data Management verification(OQ)	Ensure that when the data is exported or saved, it does not alter the information.
16	Remote Access verification (OQ)	Ensure that the MVDA software can be accessed remotely without any flaw.
17	Time Synchronization verification (OQ)	Ensure that all the clocks are synchronized on the servers and the PC, so that the there is no error during data transfer.
18	Backup and Restore verification (OQ)	Ensure that all the data can be back-up and restored in case of disaster.
19	Alert limits and Action limits settings verification (OQ)	Ensure that the alarm limits are configured correctly and they show the appropriate alarm conditions
20	Audit Trail verification (OQ)	Ensure that the MVDA system is enabled with audit trail. The audit trail is human readable and the entries do not overwritten.
21	End to End performance verification (PQ)	Ensure that system meets the performance specification over the period of time

Performance Qualification (PQ) [12, 13, 22]:

The performance verification of the real-time process monitoring system was performed after the IQ and OQ testing was complete. The intent of the PQ was to ensure that the system performed according to the expectations and was able to monitor the process in real-time. The test scripts were written to test one unit operation from upstream and one unit operation from downstream from start to end as shown below.
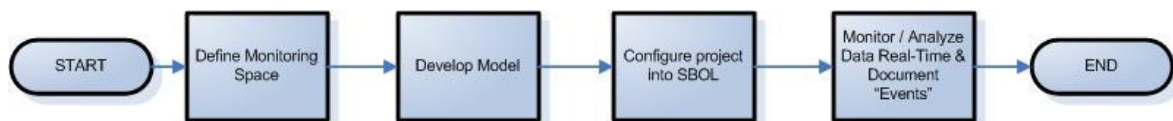
Requirement Traceability Matrix (RTM) [13, 19]:

The RTM was generated to map the functional testing of the real-time process monitoring system in validation documents (IQ/OQ/PQ) to the corresponding RS and DS specifications. This mapping helped to ensure that the requirements were met and traced to the appropriate qualification document(s). All requirements were verified and were traced to the test activity to prove that each requirement had been met.

Validation Summary Report (VSR) [13, 19]:

The VSR summarized the deliverables, validation activities, test results and deviations encountered during validation of the system. This document was generated at the end of the validation campaign to summarize a qualification conclusion that the real-time process monitoring system was validated and is suitable for using in the GMP environment.

Summary

The PAT enabled facility can generate data at the desired intervals. When these technologies are combined with Multivariate statistical methods can analyze the data to give meaningful information. Upon validation of the entire system can be used for real-time process monitoring to achieve the FDA's CPV requirements.

BUSINESS BENEFITS

In an ideal situation and complete implementation of lifecycle approach using QbD and PAT tools [2] can offer several tangible and intangible benefits to the biopharmaceutical manufacturers. The benefits of this system include detecting raw material and equipment related process variability to real-time lot release as outlined below. The real-time process monitoring at every manufacturing step, will result in tremendous benefits to the manufacturers and regulatory agencies throughout the lifecycle process [10, 24].

Each of the benefits listed below are associated with significant financial savings ultimately, cost saving and financial gains to meet the product life cycle requirements is the objective along with meeting the regulatory expectation.

Operating Space: Leveraging scientific understanding and process knowledge helps process scientists to establish an operating space. The use of QbD and PAT can expedite technology transfer and stage 1 activities ultimately resulting in a faster scale up.

Scientific Knowledge: The cost benefit is through knowledge that helps in setting the accurate confidence limits and operating space to an optimum level so that the process does not have to be modified frequently. It can save a great deal of time, money and resources in the long range and avoid process modifications at regular intervals.

Early fault detection: There are multiple cost benefits of early fault detection. It can help identify the exact cause of the failure and save time during investigation. This is linked to timely release and patient supply. If early fault detection and the cause of failure are identified quickly, then the decision of corrective measures or decision to stop the batch can be made. This can save further processing costs and the next batch can be started quicker. This can also reduce the equipment downtime.

Additional PPQ Runs: If the process is modified within, the operating space there may not be a need for additional PPQ runs which may reduce additional regulatory review and

approval. The traditional approach required change impact assessment, re-validation and refilling to regulatory agencies. With a QbD and PAT approach, every change made within the operating space is backed by scientific justification and monitored by real-time process monitoring. Therefore, the cost of revalidation and refilling is significantly reduced or eliminated. The scientific and data driven justification for every modification can save time during investigations.

Elimination or reduction of manual sample handling [34]: With traditional approach, routine samples are taken during the entire manufacturing process to ensure that the process is in control. This requires a great deal of time and resources. With the implementation of the real-time process monitoring tool, each and every critical and key process parameter is monitored in real time. This can significantly reduce or eliminate the need for off-line sampling. This is a major cost saving benefit.

Quicker investigation Time: Every process deviation requires thorough investigation and supporting data to justify the cause and effect. This is an extremely time consuming process which can delays batch release. With the real-time process monitoring tool, the identification of the issue and readily accessible data driven justification can reduce the investigation time. This can result in ensuring quicker lot release.

Consistent Product Quality: With a real-time process monitoring tool, it is ensured that every batch is consistently meeting the quality requirement. This can help establish the assurance and confidence with regulatory agency and patients.

Real-time Release: The real-time process monitoring tool can assist in maintaining the patient supply and managing the inventory. Consistent product quality with minimal variability and higher yield results in a higher return on investments.

CONCLUSION

In the traditional approach, set points and operating ranges for process parameters are defined. The control strategy is based on the demonstration of process reproducibility and testing to meet the established acceptance criteria. There are certainly flaws in the traditional approach which needed to be improved with an enhanced approach. The enhanced approach is backed by risk management studies, scientific knowledge, and process understanding.

The latest guidelines such as PAT framework [2], FDA's 2011 process validation guideline [6] and Q11 guideline for development and manufacture of drug substances by FDA [7], are eliciting the same message that the innovative technologies can used in drug manufacturing processes.

In this study, the use of one of the PAT tools for process monitoring showed how a state of control is achieved and process failures could detect batch discrepancies or sensor malfunctions. The study was conducted using the data from existing biologics manufacturing process to demonstrate the industrial application of the tool. The study outlined the validation of a process monitoring system to show that this tool could be used in the GMP environment. Even if adapting this tool requires an initial investment, it can be applied easily with appropriate management support. It definitely offers significant enhancement to process understanding, process monitoring, and scientific thoroughness in decision making. It significantly enhances qualitative and quantitative performances and cost savings. The use of multivariate process monitoring tool provides an opportunity to improve control of monitoring the process real-time so that issues can be addressed quickly.

There are multiple benefits of implementing PAT tools in the drug development, validation and manufacturing phases. In the development phase, it can provide thorough scientific knowledge and process understanding to achieve stage 1 – process design. In stage 2 – process qualification stage, it can help determine and justify the number of PPQ batches required for process validation. In stage 3 – the continued manufacturing stage, it

can help gain confidence and assurance in real-time that the batch is moving in the right direction. It may also reduce or eliminate the off-line testing [24].

If the QbD and PAT tools are applied to new products then it can help establish solid justification for the number of batches for PPQ prior to process validation campaign.

If the PAT tools are applied to an existing product then every batch can be monitored in real-time just like a process validation batch. The early fault detection can help in assuring that the processes are running at the optimum level within the operating space to give maximum efficiency, consistent quality and higher yields. This may also result in lower production cost and energy consumption.

This project is expected to reduce costs by helping to better control process variability, improve yields, reduce waste, and ensure high-quality product consistently. The cost savings upon implementation of this system for the conventional manufacturing process or new processes can be calculated using significant number of batches, right first time, quality costs and other metrics. This capability not only provides financial benefit but ensures quality product and meets the regulatory expectation for continued/continuous process verification.

The outcome of this project supports that PAT can be used for the existing or new manufacturing processes to achieve the FDAs lifecycle approach meeting "Continued Process Verification".

REFERENCES

1. Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach, Final Report, September 2004

2. Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, September 2004

3. Guidance for Industry Q8, Pharmaceutical Development, ICH, May 2006

4. Guidance for Industry Q9, Quality Risk Management, ICH, June 2006

5. Guidance for Industry Q10, Pharmaceutical Quality System, ICH, April 2009

6. Guidance for Industry; Process Validation General Principles and Practices, January 2011, Revision 1

7. Guidance for Industry Q11, Development and Manufacture of Drug Substances, ICH, November 2012

8. SIMCA User Guide, Software version 13.0.3

9. Quality-by-Design, A Perspective from the Office of Biotechnology Products (OBP), http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4187B1_01_13-%20OBP-QbD.pdf, accessed on 10/20/13 at 10:41am

10. Multivariate and Megavariate Data Analysis Basic Principles and Applications (Part I), L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, and S. Wold, Umetrics

11. http://www2.emersonprocess.com/siteadmincenter/PM%20DeltaV%20Documents/ProductDataSheets/PDS_ContinuousHistorian.pdf, accessed on 9/29/13 at 11:45am

12. CFR - Code of Federal Regulations Title 21, FDA.gov

13. General Principles of Software Validation; Final Guidance for Industry and FDA Staff, Document issued on: January 11, 2002

14. Guidance for Industry Part 11, Electronic Records; Electronic Signatures - Scope and Application, August 2003

15. EMA Draft Guideline on Process Validation, EMA/CHMP/CVMP/QWP/70278/2012-Rev1, 29 March 2012

16. http://www.nature.com/nrgastro/journal/v5/n6/images/ncpgasthep1125-f3.jpg, accessed on 09/07/13 at 01:50pm

17. http://origin-ars.els-cdn.com/content/image/1-s2.0-S0952197607001133-gr1.jpg, accessed on 09/07/13 at 02:00pm

18. Online Batch/Fed-Batch Process Performance Monitoring, Quality Prediction, and Variable-Contribution Analysis for Diagnosis, Cenk Undey, Sinem Ertunc, and Ali Cinar, Ind. Eng. Chem. Res. 2003, 42, 4645-4658

19. GAMP 5, A Risk-Based Approach to Compliant GxP Computerized Systems

20. IEEE, "IEEE Standard 610.12-1990, IEEE Standard Glossary of Software Engineering Terminology," 1990

21. SIMCA Brochure (http://www.umetrics.com/kb/audit-trail-compliance-21-cfr-part-11-3)

22. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/QualitySystemsRegulations/MedicalDeviceQualitySystemsManual/ucm122439.htm

23. The Good Automated Manufacturing Practice (GAMP) Guide for Validation of Automated Systems GAMP 4

24. Using PAT for Real-Time Monitoring and Quality Assurance in Chemical Development, Pharmaceutical Review, June 2010

25. http://usolutionspr.com/wp-content/uploads/2012/10/PICTURE-v-, accessed on 10/05/13, at 4:00pm

MANUSCRIPT 2: USE OF MULTIVARIATE STATISTICAL PROCESS MONITORING TOWARDS 'CONTINUED PROCESS VERIFICATION'

Prakash Bedre (Principal Validation Engineer, Amgen Inc.); Cenk Undey (Director, Process Development, Amgen Inc., Clinton Chichester (Professor, University of Rhode Island)

ABSTRACT

This study intends to demonstrate the benefits of a process analytical technology (PAT) enabling a real-time multivariate process monitoring (RT-MSPM) system to be used during product manufacturing. The implementation of the RT-MSPM tool shows the application of FDA's 2011 Process Validation Guidance for industry to perform process monitoring for continuous process verification. The study was focused on the need of completing stage 1 - 'Process Design' and 'stage 2 – 'Process Performance Qualification (PPQ)' activities for the successful implementation of stage 3 – Continued Process Verification (CPV)'. With sufficient scientific knowledge of process design and process qualification, CPV could be implemented by using PAT tools. The article focuses on the use of multivariate data analysis methods such as PCA and PLS to create two models: batch evolution and batch level models. The data sets from historical batches for one upstream unit operation (bioreactor) and one downstream unit operation (UF/DF) are used to develop the experimental models for multivariate monitoring purposes. The study summarizes was conducted at a commercial biologics manufacturing facilities.

NEW PROCESS VALIDATION APPROACH DUE TO 2011 GUIDANCE

All drug manufacturing facilities must follow regulatory guidelines from the Food and Drug Administration (FDA or USFDA) such as 21 *CFR* Parts 210 and 211 of the cGMP regulations [8], *1987* process validation guidance [1] and other guidelines which are introduced on a regular basis. The biopharmaceutical industry has been conscientiously following the cGMP regulations to make consistent and reproducible commercial products.

Yet, due to recent technological developments and innovations, the pressing need and the pressure challenged the drug manufacturing process. From the beginning of the century, FDA and other worldwide agencies introduced new guidelines such as QbD and PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance [6], International Conference on Harmonization's Q8 for Pharmaceutical Development [2], Q9 for Quality Risk Management [3], Q10 for Pharmaceutical Quality System [4] and Q11- Development and manufacture of drug substances [5]. Regulatory agencies objective is to encourage the innovation in the drug manufacturing process [6].

In January 2011, FDA published new guidance for industry entitled Process Validation: General Principles and Practices [7]. Since it is guidance from the regulatory agency, it is legally enforceable per the Federal Food, Drug, and Cosmetics Act. The requirements are called out in 21 *CFR* Parts 210 and 211 of the CGMP regulations, more specifically in Part 211.100 (a) [8].

There had been a gap of exactly 25 years between FDA 1987 Guideline and the 2011 Guidance for process validation. The 2011 Guidance is entirely consistent with the basic principles of process validation articulated in the 1987 Guideline.
"Nonetheless, more than 25 years' worth of experience and regulatory oversight, along with the cGMPs for the 21st Century Initiative [9], prompted FDA to revisit the principles and concepts in an effort to update and clarify FDA's thinking on process validation".

Per this new guidance, manufacturers are urged to adopt the lifecycle approach in three stages [7]:

In Stage 1, process design, the commercial process is defined based on scientific knowledge gained through development and scale-up activities.

In Stage 2, process qualification, the process design is evaluated and assessed to determine if the process is capable of reproducible commercial manufacturing.

In Stage 3, continued process verification, ongoing assurance is gained during routine production that the process remains in a state of control.

Per 2011 guidance, FDA states that process validation is to be a lifecycle approach instead of being a one-time activity. The FDA's new approach is to make 'every manufacturing batch as a 'Validated' batch via 'Continued Process Validation'. The following Figure 1 outlines the FDA's new process validation expectation [10].

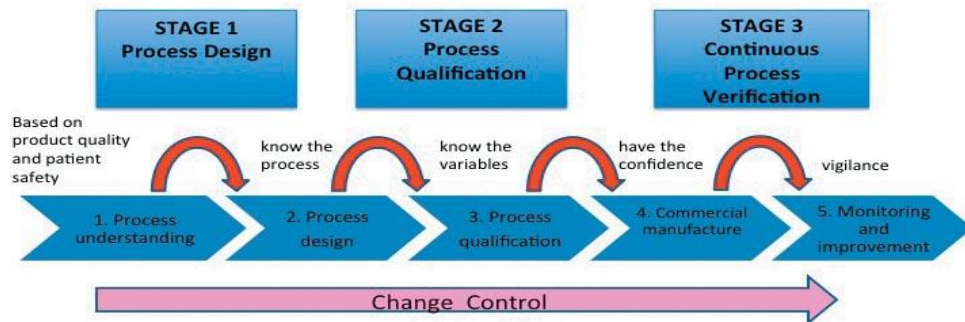Figure 1: Three stages of Process Validation. FDA's 2011 guidance requires drug manufacturing process to be a three stage process. Stage 1, stage 2 and stage 3. The stage 1 activities are required for thorough scientific knowledge and process understanding is required to establish a robust operating space. Stage 2 represents process validation and stage 3 is for continued process verification. [Source: 10]

DECODING THE REGULATORY EXPECTATIONS

<u>Stage 1 - Process Design</u>: Well Characterized Process for Thorough Understanding of Process Parameters via 'Risk Assessment'

Scientific knowledge gained during the pharmaceutical development program is critical for enhanced understanding of product quality and provides a basis for risk management and increased regulatory flexibility [11]. This stage is considered the design phase (stage 1) of the FDA's latest guidance. The initial scientific knowledge about the drug and its manufacturing process is gained prior to the validation campaign. Use of several laboratory scale, pilot scale and development runs for thorough understanding of process parameters and their critically are required. Every parameter must be evaluated utilizing quality risk management practices.

Stage 1 – Process Design states that, "The commercial manufacturing process is defined during this stage based on the knowledge gained through development and scale-up activities [7]". The first step to achieve this mile stone is establishing an operating space.

The challenge is successfully answering a set of key question which can lead to gaining scientific knowledge and thorough process understanding to establish the 'Operating Space'. The questions are: What are the physical, chemical and biological properties of the raw materials used?, What are the physical, chemical and biological properties of the API or drug substance?, What parameters are associated with the drug?, What are the operational (input) parameters?, What are the process (output) parameters, Has risk and criticality assessment of each parameter performed?, which once are critical, key or non-key parameters?, What is the target set-point for each parameter?, What is the normal operating range of each parameter?, What is the proven operating range of each parameter?, What are the limits of failure?, what is the concentration of the final drug product?, what are the attributes being test tested to analyze the API and final drug product? What are the acceptance criteria for the process parameters?, How is the acceptance criteria established?

The answers to these questions helps you understand the process limits for every parameter you need to establish the design space as shown in Figure 2 [12]. ICH guidance Q8, defines it as "the multidimensional combination and interaction of input variables and process parameters provides the assurance of quality [2]".
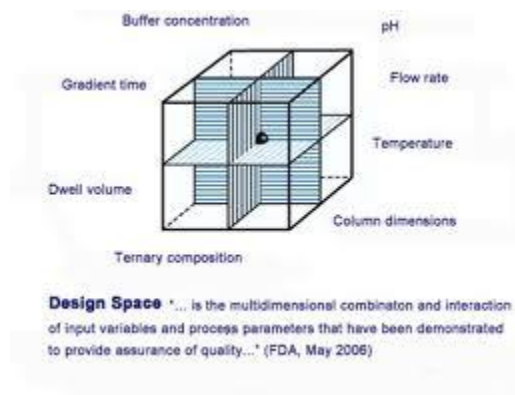
Figure 2: Design Space [Source: 12]

The scientific knowledge of operating space provides the understanding of variability in raw materials, the relationship between a process and product's critical quality attributes (CQAs), and the association between CQAs and product's clinical properties. This through understanding can help "Control the variation in a manner commensurate with risk it represents to the process and product [7]". The scientific knowledge of drug and process parameters can be achieved by conducting a design of experiments (DOE) a.k.a characterization studies. The high degree of scientific knowledge and assurance in the performance of the manufacturing process is obtained from objective information and data from laboratory, pilot, and/or commercial scale studies [7]. The results obtained from these studies define the operating space. DOE studies can help develop process knowledge by revealing relationships, including multivariate interactions between the variable inputs (e.g., component characteristics or process parameters) and resulting outputs (e.g., in-process material, intermediates, or final product) [13].

Performing the Risk Assessment [13]:

The most import task is performing a risk assessment of the operational and process parameters identified during characterization studies of the DOE. Per the 2011 guideline, 'All the parameters should be evaluated in terms of their roles in the process and impact on the product or in-process material [7]. A team of representatives from manufacturing, process development, quality assurance and validation are required to perform the assessment. A typical quality risk management model outlined in Figure 3 [3] is commonly used as described in the ICH Q9, Quality Risk management [4]. Each and every parameter identified during a characterization study is evaluated to find out what might go wrong? The likelihood of it going wrong and the consequences of it are discussed during the risk assessment. Based on the evaluation of the risk model, a score is assigned in three different categories such as low, medium and high. The risk parameters are weighted against the likelihood of occurrence, probability of detection, and severity of consequences. All three scores are multiplied to obtain a risk priority number (RPN) as shown in Figure 3 [3]. A decision is made to identify the parameter as critical, key or non-key to merit process characterization. 2011 guidance expects, 'a higher degree of control for the parameters that pose higher risk [7]. The results are documented as characterization reports to establish the operating space.
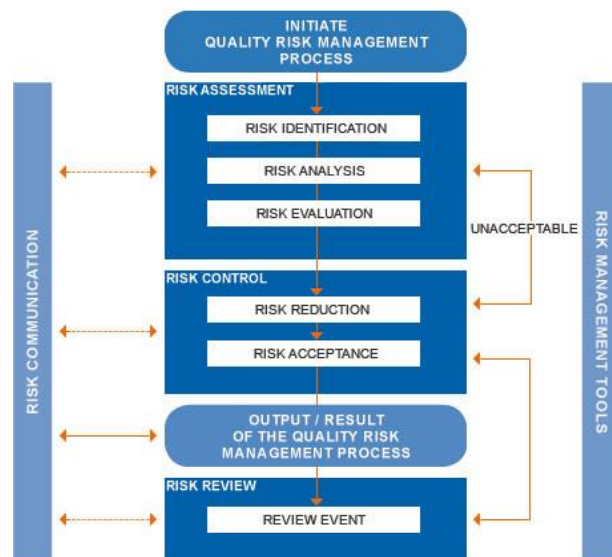

Figure 3: Risk Assessment Tool. During the risk assessment process, each and every process parameter is evaluated using a risk assessment tool [Source: 3]

98
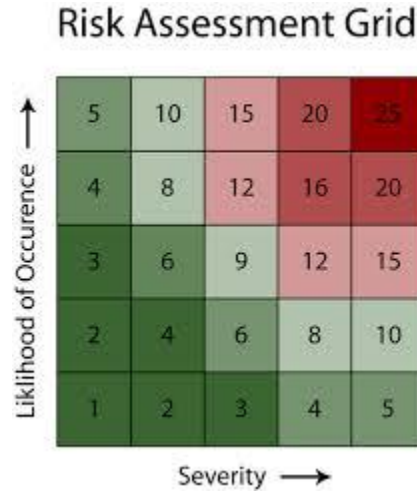
Figure 4: Risk assessment grid. The risk assessment grid is used to assign the RPN number for each and every process parameter associated with process which is evaluated and scored based on the risk likelihood and, risk severity. [Source: 14]

The operational ranges for the operational parameters and acceptance criteria for the process parameters in the design space are the basis for process validation protocols to validate the process [15]. The scientific knowledge and information gathered must be documented, and approved in accordance with the established procedure so that it can be used in the stage of lifecycle [7].

Stage 2: Process Performance Qualification (PPQ)

The goal of validating any manufacturing process is to establish scientific evidence that the process is reproducible and will consistently deliver quality products. The sufficient scientific knowledge and assurance gained during stage 1 via characterization studies, sets the stage for stage - 2 process qualifications. How do the characterization studies help in process performance qualification (PPQ)? FDA's 2011 guidance outlined that the manufacturers should [7] understand the sources of variation, detect the presence and degree of variation, understand the impact of variation on the process and ultimately on product attributes, and control the variation in a manner commensurate with the risk it represents to the process and product.

99

The scientific evidence gathered during characterization provides the appropriate level of assurance that the manufacturing system has been designed to consistently deliver a quality product to the market. The specific information obtained from the operating space such as the critical/key parameters and, control strategy to set the normal operating ranges (NOR) and proven acceptable ranges (PAR). This information derives the scientific justification for the parameters selection and to calculate and establish the control limits which serves as an input for PPQ protocol.

This phase involves evaluating the facility and equipment for its fitness for use. Utility systems and equipment are verified to be built and installed properly, and operators ensure that they operate within the intended and anticipated operating ranges. During the PPQ stage of process validation, the process design is evaluated to determine if it is capable of reproducible commercial manufacture of products [7]. The decision to distribute the product to the market is determined by the successful completion of the PPQ. The successful completion of the PPQ demonstrates that the commercial manufacturing process performs as expected.

Number of PPQ Batches:

One of the most important discussion and interpretation of FDA's 2011 guideline is about the number of batches. Until the new guidance came along, the process validation was done by performing a three-batch requirement. "...it was widely accepted throughout industry, and, indeed, implied or stated in some FDA guidance documents, that process validation was a static, three-batch demonstration event. [16]". The EU GMP Annex 15 states that "It is generally considered acceptable that three consecutive batches/runs within the finally agreed parameters would constitute a validation of the process" [17]. The 2011 guidance does not mention anywhere about the number of PPQ runs required for successful completion PPQ stage. The decision about number of PPQ batches it open for interpretation. One can interpret that 'process validation as a continuous process of collection and evaluation of data, rather than as a three-batch static event" [10].

The number of batches is not an acceptance criterion; however, the results of the data obtained from the batches are the acceptance criteria. The new definition of validation caused one industry member to state at the workshop that for the past 30 years, industry has been told that process validation is a documentation exercise. FDA expects industry to consider process validation as a scientific endeavor. That is quite a shift and 30-year habits are hard to break [10, 18, 19].

The existing products which are already in the market may have already crossed this hurdle by making three process validation (PV) batches, which got the approval of FDA for commercial manufacturing. The 2011 guidance is also applicable to these products. The FDA directive for the manufacturers of these products is to follow the life cycle approach. The legacy product manufacturers benefit from the knowledge they have already gained about the manufacturing process and the product over course of the commercial manufacturing. Use of PAT tools for these manufacturing processes can really enhance their process monitoring capabilities to achieve the stage 3 – continued process verification.

At the same time, new products and new manufacturing processes have the benefit of following the QbD and PAT principle right from the beginning to gain process understanding during stage 1 so that they can scientifically justify the number of batches required for PPQ. The manufacturers must make deliberate, rational decisions about whether their specific processes are validated and their products ready for commercial release. A manufacturing process that uses PAT may warrant a different PPQ approach. PAT processes are designed to measure in real time the attributes of an in-process material and then adjust the process in a timely control loop so the process maintains the desired quality of the output material [7].

Justification for selecting number of batches for PPQ:

The PPQ validation strategy can be used to scientifically justify the required number of batches selected for PPQ. The knowledge gained from previous molecules and during the process design stage through development, scale-up activities and engineering runs can be used to demonstrate that the current process is well characterized. A thorough process understanding has been developed by virtue of a comprehensive set of pilot and robustness studies. Operating parameter and performance parameter classifications and ranges have been determined, and a strategy for overall process control has been established. The operating and performance parameter classifications and ranges have been developed based on the studies performed as part of Stage 1 Process Design and documented in the characterization report. A manufacturing process at commercial scale is established based on this process knowledge. This information can be used to justify the required number of PPQ runs. By establishing the sufficient data, scientific justification can be provided to prove that one, two or five process validation runs are sufficient instead of three process validation runs per the traditional approach.

Sampling during PPQ batches:

The 2011 guidance also emphasized on the sampling plan, sampling points, number of samples, frequency of sampling for each unit operation and using a statistical approach for PPQ samples. The number of samples should be adequate to provide sufficient statistical confidence of quality both within a batch and between batches [7]. The use of a statistical tool and the approach is not specified but the manufacturers are expected to choose a suitable statistical tool. Homogeneity within a batch and consistency between batches are goals of process validation activities. The expectation is to use the heightened sampling and monitoring period to gain the confidence and assurance for the high risk parameters.

Stage 3 – Continued Process Verification (CPV):

The goal of the third validation stage is continual assurance that the process remains in a state of control (validated state) during commercial manufacture. A system needs to be in place for detecting unplanned departure from the process as designed during stage 1 and stage 2 [7]. Ideally, this stage should be treated as extension to stage 1 and stage 2 because all the scientific knowledge of operating space and its verification is done during process validation phases. But, this is still creating confusion because it is a new concept and the expectations stated in the 2011 guidance are vague.

The new guidance outlines that upon regulatory filing and receiving the approval for commercial manufacturing, the manufacturers maintain the same state of control as it was shown during PPQ runs to ensure that each batch is a process validation batch. When implementing stage 3, manufacturers should consider the semantic difference between the terms "continued" and "continuous". The 2011 Guidance deliberately speaks to continued process verification, which some organizations have misinterpreted to mean continuous, with mandatory enablement via PAT. The expectation is decidedly *not* that in-process or release testing required under the cGMP regulations be replaced by PAT approaches. Rather, the expectation is for ongoing, (i.e., inter and intra-batch, monitoring, and review) [18].

The 2011 guidance has been rolled out for two and half years. Most of the manufacturers are still in the process of digesting this concept. By taking the hints from the 2011 guidance, various ideas are being explored. Most of the manufacturers are in the process of modifying their procedures to adapt the CPV philosophy so that they can slowly but steadily join the band wagon. Some manufacturers are ahead in the game because they were early adaptors of QbD and PAT principles. Here are some of the examples discovered from the workshops and seminars on CPV interpretation and adaption:

Current practices for the implementation of CPV [20]:

Sampling during CPV: Instead of monitoring PPQ level sampling, only a few appropriate parameters are selected for the stage 3 sampling plan. The data is collected from these parameters until sufficient information is available to generate sufficient variability

estimates by review of historical data trends. Standard operating procedures (SOP) are put in place to define additional monitoring and the heightened sampling and monitoring period [21], list of additional parameters monitored, and statistical methods to measure Process Capability and Product Trends review [7].

Review of Supplier Audit: The incoming materials are ensured to be in a state of control. This can be demonstrated by [19] supplier audits and by verifying documents such as supplier quality agreements, certificates of analysis and raw material testing results.

MMP (Master Monitoring Plan): The drug manufacturers are planning to monitor the critical and key process parameters using pre-defined master monitoring plans for the products prior to commercial manufacturing so that the data and the data trends can be routinely monitored. Once the MMP is put in place, any change to the operating conditions (ranges and set-point), process controls post filling and regulatory inspectional commitments can be reflected in the MMP. In order to have a robust MMP, preliminary process control limits and centerlines (if applicable) must be established for all in process control (IPC) parameters and critical and key post-filling parameters being monitored at or about lot 15. The statistical process control limits and centerlines (if applicable) must be established for all IPC parameters and critical and key post-filling parameters being monitored at or about lot 30.

Monitoring Quality Systems: There are other periodic review quality systems are also being used to monitor the achieve CPV such as periodic review of post approval change control process [7], periodic review of non-conformances and defect reporting systems, verification of Root Causes and CAPA process, periodic review of validated equipment, systems and utilities at regular intervals, periodic review of monitoring for CIP and SIP cycles, periodic review, monthly and annual review of equipment and facility qualification [7], incorporating appropriate detection, control and mitigation strategies, collecting regular feedback from the process operators and quality staff on the process performance and maintaining and reviewing the product complaints data.

Statistical Evaluation and Analysis of Process Data [7, 21]

The drug manufacturers are following the FDA's suggestion of using statistics in the evaluation of data trends for analyzing the data. In order to achieve this, all IPC parameters and critical and key post filling parameters must be monitored starting from the first lot scheduled for commercial release. The product specific control limits must be established using generally accepted statistical process control practices with upper and/or lower control limits computed nominally at three standard deviation units from the mean for normally distributed data. The use of Nelson rules for supporting the making the statistical decisions is followed by many drug makers. As outlined in Nelson rules [22], all parameter results from each production lot must be examined across lots for statistically abnormal behavior. This approach shall include using the following Nelson (NEL) run rules for normally distributed parameters:

- NEL 1: One point beyond a control limit
- NEL 2: Nine points in a row on one side of the centerline
- NEL 3: Six points in a row steadily increasing or decreasing
- NEL 4: Fourteen points in a row alternating up and down

Use of PAT Tools: Lastly, due to the technological developments, many new process capabilities are available for manufacturers. They can install the on-line, at-line and in-line sensors to monitor critical and key process parameters in real-time. Installation of an electronic data collection system is a new trend in the industry, which stores batch historical batch data so that this data can easily accessed to analyze the trends between batches and within the batches. The use of multivariate statistical analysis methods for real time process monitoring [23] is also a growing trend.

INDUSTRIAL APPLICATION OF PAT TOOL TO ACHIEVE CPV

As discussed in the previous section, the manufactures are in the process of implementing various tools to include CPV into their manufacturing processes to meet the FDA's 2011 guidance. The use of real-time multivariate statistical process monitoring system is utilized here to show the how CPV can be achieved.

Material

SIMCA Software v13.0.3 [8]: The SIMCA software version 13.0.3 is utilized for the creation of statistical MV models. This software is developed by Umetrics.

This software is used to create design of experiments and multivariate data analysis. This tool transforms data into information, which can be seen in the form of colorful graphical control charts to enable the process analyst to make the correct decisions and take the appropriate actions in real-time.

Historical data from the good batches (batches which have minimum deviations) are used from one of the well-known commercial biological manufacturing process. The data is extracted from the historical databases by making configurations, tags, scan rates, and compression settings to the source system. This data is pre-treated and organized in appropriate format prior to importing it into the SIMCA software for the creation of multivariate models.

Methods

Two popular and commonly used Multivariate Data Analysis (MVDA) Methods are principal components analysis (PCA) and partial least squares (PLS), which are used to show the use of the PAT tool described in the QbD, and PAT framework. [PAT Framework, QbD [2, 11, 24].

The commercial biological manufacturing process is enabled with various modern process analyzers and is equipped with a distributed control system to collect the process data. The real-time multivariate statistical process monitoring system is linked to the

process databases obtain the process data. Various modifications were made to the existing databases for the collection of data from the unit operations. The study is focused one upstream unit operation (i.e. bioreactor) and one downstream unit operation (i.e. UF/DF).

The propriety data were modified and normalized as necessary prior to using it for the creation of the multivariate (MV) model in SIMCA application.

The goal is to demonstrate that the following PAT tools can be utilized to ensure continued process verification is met as outlined in FDA's 2011 process validation guidelines [7]:

I. Multivariate data acquisition and analysis
II. Continuous improvement and knowledge management tools

Most of the existing manufacturing processes for biologics product at scale are not designed to inherently enable the real-time process monitoring. The real-time process monitoring (PAT software tools) and QbD principles are required for monitoring a biologics manufacturing process in "Real-Time".

The bioreactor and UF/DF unit operations were monitored against the input parameters listed in table 1 and table 2. These unit operations were connected to the distributed control system (DCS) to collect the process data [11]. The process database was connected to DCS, which saved the historical and current batch data continuously generated from the ongoing batches. The continuous plant data historian was configured with the correct tags. The cell viability and viable cell density data were also collected from off-line measurements to check the process performance at twenty-four hour intervals during the course of the unit operation. The configuration of trigger tags, timers and batch tags were made as required to get all the relevant batch and continuous process data from the historical databases.

The on-line data was collected at fifteen minute intervals for the bioreactor unit operation and at 10 second intervals for UF/DF unit operation. The data from fifteen batches were

used for the creation of MV model and control charts for the real time statistical process monitoring. The historical batches which had lowest number of deviations and alarm conditions were selected as an input for the MV model creation. The goal for this selection criterion was to create a reference MV model. The MV created the average and ± 3 standard deviation for the confidence limits for the control charts. Two new batches were selected for the analysis to find out if they were good or bad batches. One batch of the two new batches data was deliberately modified by making known changes to certain variables for testing purposes. This study was used to prove if the real-time multivariate statistical process monitoring system can achieve:

1. Real time process monitoring
2. Fault detection due to process failure or sensor malfunction
3. Root cause analysis and process control

MV Model Creation for Bioreactor and UF/DF unit Operation

Bioreactor: The dataset contains data for N=17 batches. Out of these, fifteen batches were selected for the creation of MV model. The batch selection criteria were to have little variability among the batches used for the bioreactor cell culture process. The main objective of the study was to create a MV model so that it could be used as a reference to monitor new batches as they evolve and identify good batches from the bad batches. Fourteen variables (J=14) were monitored and the data was collected at every fifteen minute interval, giving a total of ~ K = 279 time points per batch. The total duration of the unit operation was 68 hours, and 30 minutes. The bioreactor unit operation is assumed a single phase process.

MV for UF/DF: The dataset contains data for N=17 batches. Out of these, fifteen batches were selected for the creation of the MV model. The batch selection criteria were to have a little variability among the batches used for the UF/DF process. The main objective of the study was to create a MV model so that it could be used as a reference to monitor new batches as they evolve and identify good batches from the bad batches. Nineteen variables (J=19) were monitored and the data was collected at ten second interval, giving

a total of ~ K = 800 time points per batch. The total duration of the unit operation was 2 hours: 31 minutes: 10 seconds. UF/DF unit operation has three phases such as concentration, diafiltration and recovery.

The data are scaled to UV variance and unfolded by the SIMCA software prior to using for the model creation [24].

Table 1: Bioreactor Variables Monitored

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Dissolved O_2	%	Operating	Input	Dissolved oxygen is monitored as a measure of cell culture performance
Culture pH	pH units	Operating	Input	pH affects final viable cell density
Air flow	sLPM	Operating	Input	Air flow is monitored as a measure of cell culture performance
O_2 flow	sLPM	Operating	Input	O_2 flow is monitored as a measure of cell culture performance
CO_2 flow	sLPM	Operating	Input	CO_2 flow is monitored as a measure of cell culture performance
Bioreactor level	L	Operating	Input	Bioreactor level is monitored to maintain consistent volume
Agitation	RPM	Operating	Input	Agitation maintains a homogenous solution and oxygen transfer to the cells
Vessel pressure	psig	Operating	Input	Pressure influences mass transfer and mitigates contamination
Temperature	°C	Operating	Input	Temperature (controlling probe) excursions can impact final viable cell density and viability
Culture duration	days	Operating	Input	Culture duration affects final viable cell density and is the maturity variable
Temperature probe difference (A-B)	°C	Operating	Input	Temperature probe difference (A-B) is monitored to detect equipment drift or malfunction
VCD	10^6 cells/mL	Performance	Output	VCD is monitored as a measure of cell culture performance.
Viability	%	Performance	Output	Viability is monitored as a measure of cell culture performance.

Table 2: UF/DF Variables Monitored

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Feed pressure	psig	Operating	Input	Feed pressure is monitored to detect excursions in pressure
Retentate pressure	psig	Operating	Input	Retentate pressure is monitored to detect excursions in pressure
Permeate pressure	psig	Operating	Input	Permeate pressure is monitored to detect excursions in pressure
Transmembrane pressure (TMP)	psig	Operating	Input	TMP is monitored to detect any excursions in pressure
Feed flow	LPM	Operating	Input	Feed flow is monitored to detect any excursions in flow rate
Retentate flow	LPM	Operating	Input	Retentate flow is monitored to detect any excursions in flow rate
Permeate flow	LPM	Operating	Input	Permeate flow is monitored to detect any excursions in flow rate
Feed Retentate DP	psig	Operating	Input	Feed retentate pressure is monitored to detect differential pressure
Permeate flux	L/hr/m2	Operating	Input	Permeate flux (permeate flow rate normalized by membrane area) is monitored to detect excursions in flow rate
Permeate UV	AU	Operating	Input	Permeate UV is monitored to detect product loss during UF/DF II
Permeate conductivity	mS/cm	Operating	Input	Permeate conductivity is monitored to ensure product conductivity targets
Concentration factor	N/A	Operating	Input	Concentration factor is monitored to ensure concentration targets are met
Diafiltration Factor	L	Operating	Input	A minimum number of diavolumes are required to meet pH and conductivity specifications for the UF/DF II Pool

Variable Name	Unit of Measure	Variable Type	Variable Class	Variable Use
Feed flow process totalizer	L	Operating	Input	Feed flow totalizer is monitored to ensure concentration and diafiltration targets are met
Permeate flow process totalizer	L	Operating	Input	Permeate flow totalizer is monitored to detect any excursions in the total permeate volume
Retentate flow process totalizer	L	Operating	Input	Retentate flow totalizer is monitored to ensure concentration and diafiltration targets are met and is the maturity variable
Permeate control valve	%	Operating	Input	Permeate control valve is monitored as a controller output
Retentate control valve	%	Operating	Input	Retentate control valve is monitored as a controller output
Step yield	%	Performance	Output	Step yield is monitored as a measure of UF/DF performance.

The MV model created for bioreactor unit operation using fifteen good batches is shown in Figure 5a and Figure 5b by a score plot and a batch contribution plot. The MV model created a default 95% confidence limit using F distribution, which is shown by an ecllipse in Figure 5a. In Figure 5b, the batch contribution plot shows a ±3 standard deviation and averages using the data from the reference batches. The ±3 standard deviation and averages are shown in red and green. Figure 5a and Figure 5b plots show that all batches are aligned properly and ending in a similar fashion within the confidence limits.
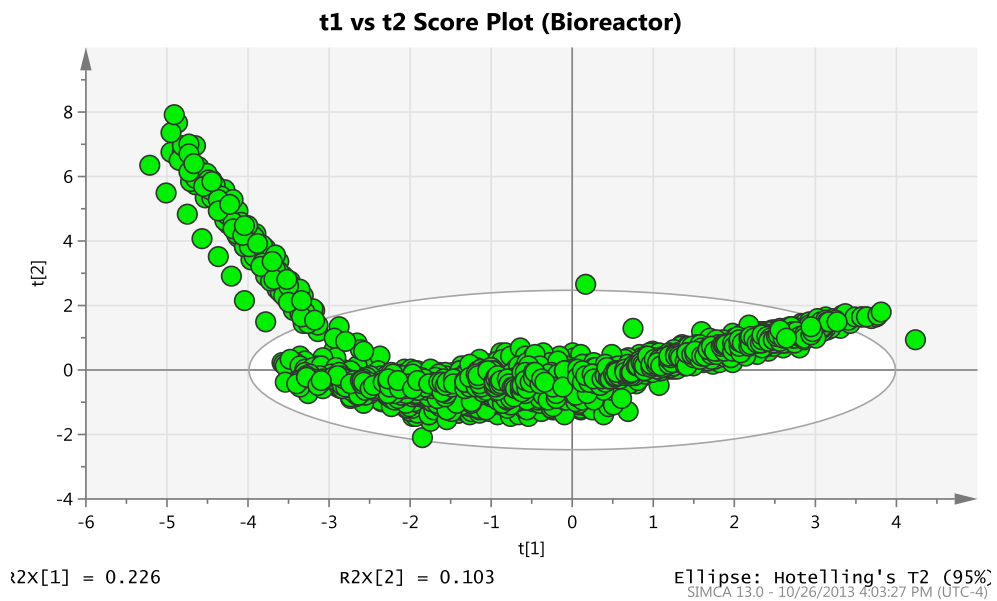

Figure 5a: The Score Plot (BEM) for bioreactor. The score plot was created for the bioreactor using the scores of first two principal components. The score plot shows that all batches are aligned properly and fitting the 95% confidence limit ecllipse.
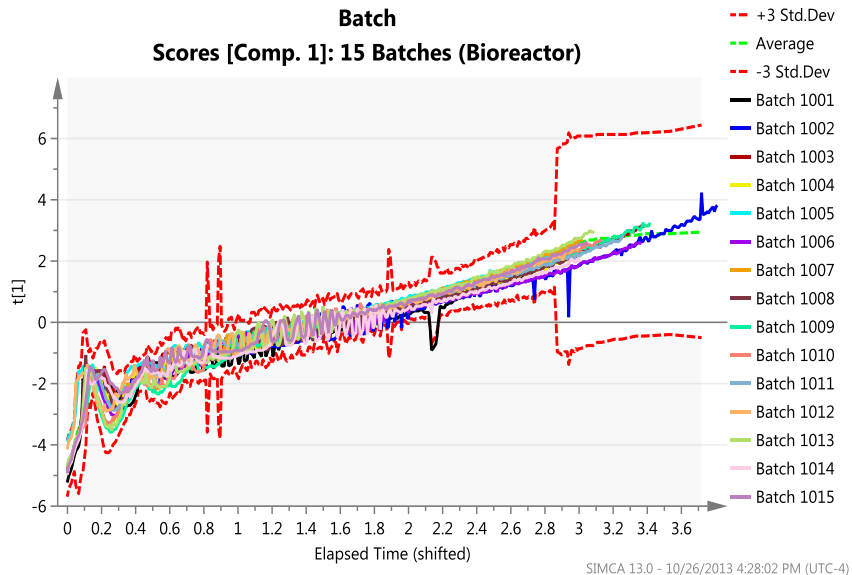
Figure 5b: The PC 1 Score Batch Plot for bioreactor. The batch score plot for the bioreactor using the scores of the first principal components on the vertical axis and the elapsed time on the horizontal axis. The score plot shows that all batches are starting and ending within a similar fashion within ±3 standard deviation confidence limit.

UF/DF unit operation has three different phases. Therefore, SIMCA created three separate batch evolution models for each phase. The UF/DF MV model for the concentration phase is shown in Figure 6a and 6b. The model is created using fifteen good batches and is depicted by a score plots and a batch contribution plot. The MV model created a default 95% confidence limit using F distribution which, is shown by an ecllipse in the Figure 5a. In Figure 5b, the batch contribution plot shows a ±3 standard deviation and averages using the data from the reference batches. The ±3 standard deviation and averages are shown in red and green. Figure 5a and Figure 5b plots show that all batches are aligned properly and ending in a similar fashion within the confidence limits.

t1 vs t2 Score Plot (UF/DF) Concentration Phase

Ellipse: Hotelling's T2PS (95%)
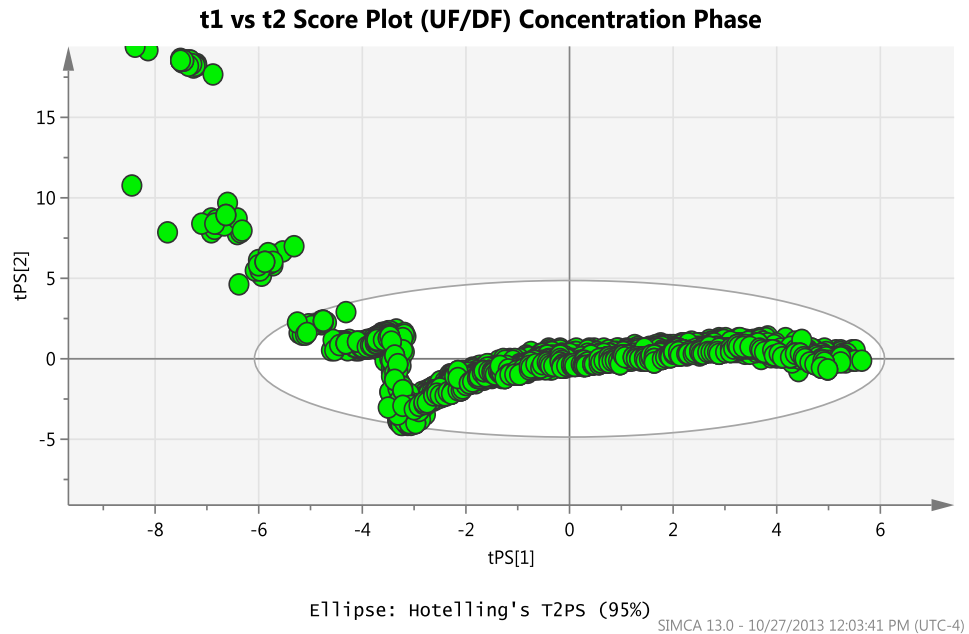
SIMCA 13.0 - 10/27/2013 12:03:41 PM (UTC-4)

Figure 6a: The Score Plot (BEM) for UF/DF. The score plot created for UF/DF using the scores of the first two principal components. The score plot shows that all batches are aligned properly and fitting the 95% confidence limit ecllipse.

Score BCC Concentration
[Comp. 1] 15 Batches (UF/DF)

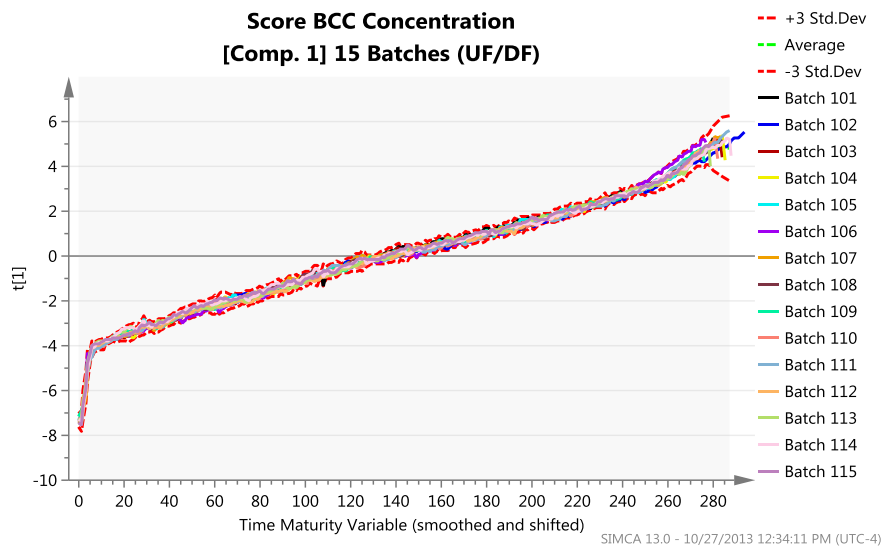SIMCA 13.0 - 10/27/2013 12:34:11 PM (UTC-4)

Figure 6b: The PC 1 Score Batch Plot for UF/DF. The batch score plot for UD/DF using the scores of the first principal components on the vertical axis and the elapsed time on the horizontal axis. The score plot shows that all batches are starting and ending within a similar fashion within ±3 standard deviation confidence limit.

115

Use of MV Model for real time process monitoring and fault detection

In this section, we can see how these models can be used to monitor the new batches as they evolve. If the new batch is a good batch then it is expected to evolve with in the confidence limits. This can be seen on the control charts. At the same time, if for any reason, there is a deviation then the cause and source of the deviation can be tracked using control charts. Introduction of this tool within the cGMP manufacturing facility which is equipped with the on-line data collection technologies can monitor every new batch in real time and provide an opportunity for continuous improvement and scientific knowledge management to meet the FDA's stage three life cycle approach.

In order to test the real-time process monitoring system to ensure the continued process verification, two new batches were selected for both the bioreactor and UF/DF unit operations each. One of the two new batches was a good batch and the second new batch was deliberately modified by making deliberate changes to a few variables to see if they can be detected by the real-time process monitoring system.

In this case study 1, a new batch (batch# 1016) was projected on MV model created for the bioreactor process using fifteen good batches to see if this batch was running in the state of control. The batch score plot and the individual batch plot in Figure 7 showed that the batch# 1016 (shown in red) was well within the 95% control limit. The batch contribution plot shows that batch# 1016 was moving within ± 3 standard deviations shown in red and around the average (shown in green).
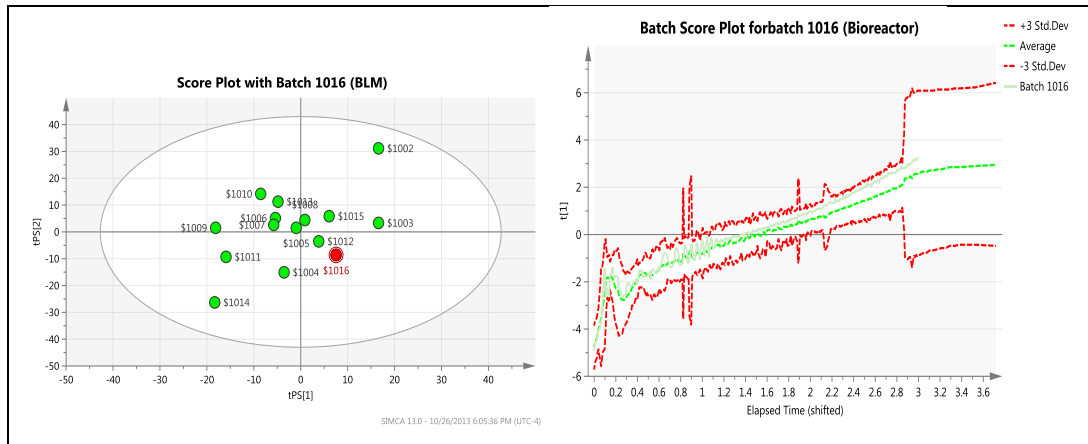
Figure 7: The batch score plot for the bioreactor batch# 1016. The plot shows that it was within the 95% confidence interval. The individual batch plot shows that the batch 1016 moved within ±3 standard deviation confidence limit.

In this case study 2, a new batch (batch# 1017) was projected on the MV model created for the bioreactor process using fifteen good batches to see if this batch was running in the state of control. The batch score plot and the individual batch plot in Figure 8 showed that batch# 1017 (shown in red) was outside of the 95% control limit. The batch contribution plot showed that batch# 1017 is outside of ±3 standard deviations at several time points. This graphical presentation of new batch in real time reveals that the batch was not a good batch and the cause of deviation to be addressed immediately.
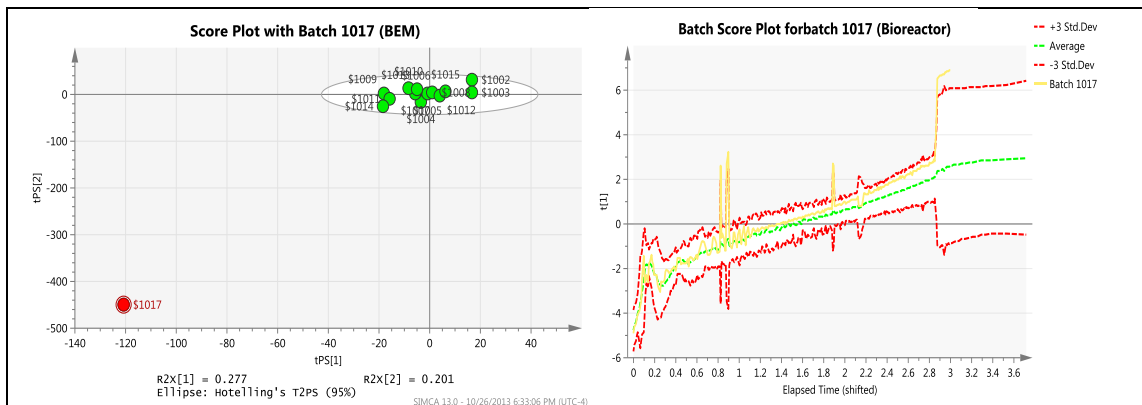

Figure 8: The batch score plot for the bioreactor batch# 1017. The plot shows that it was outside of the 95% confidence interval and the individual batch plot shows that the batch# 1017 moved outside of ±3 standard deviation confidence limit at several time points.

117

In order to find out the cause for the batch deviation, the contribution plot for each variable was evaluated. It revealed that the pH and the temperature sensors were malfunctioning. Figure 9 shows the pH and temperature batch plots with the specific time points where the batch was out of the confidence limit.
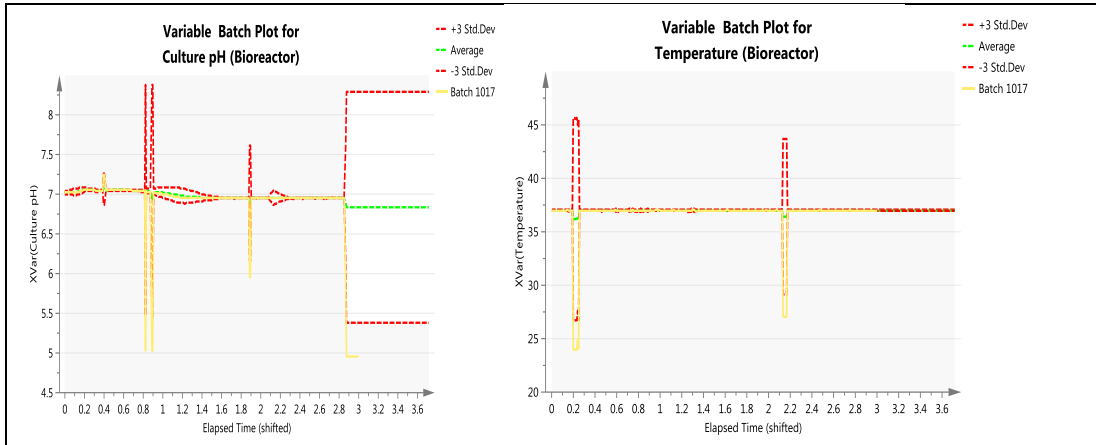

Figure 9: The Variable Batch Plots for pH and Temperature. The plot shows the exact time points where the batch was outside of ±3 standard deviation confidence limit.

In this case study 3, a new batch# 116 was projected on MV model created for UF/DF using fifteen good batches to see if this batch was running in the state of control. A batch score plot and a batch contribution plot in Figure 10 shows that the batch# 116 (shown in red) was well within the 95% control limit. The batch contribution plot shows that batch# 116 was moving within ± 3 standard deviations (shown in red) and around the average (shown in green).
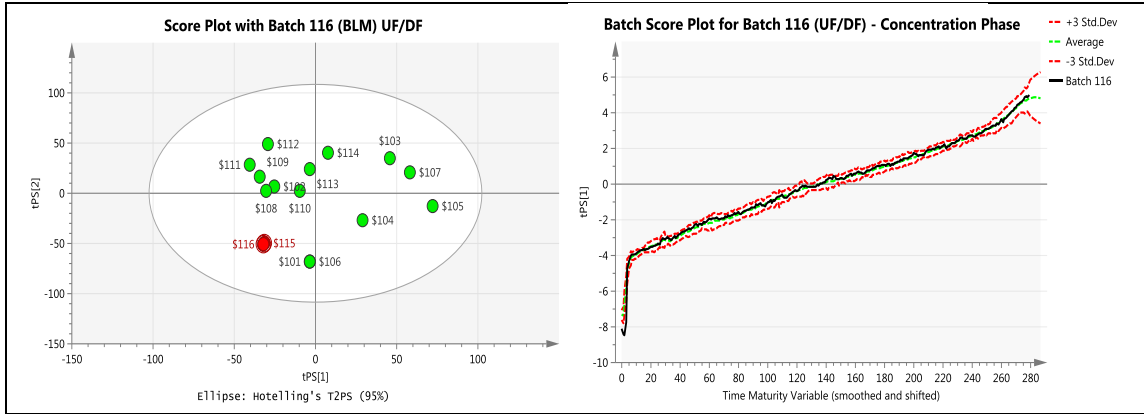
Figure 10: The Batch Score Plot for the UF/DF batch# 116. The plot shows that it was within the 95% confidence interval and the individual batch plot shows that the batch# 116 was evolving within ±3 standard deviation confidence limit.

In this example 4, a new batch# 117 was projected on the MV model created for the UF/DF using fifteen good batches to see if this batch was running in the state of control. The batch contribution plot in Figure 11 shows that batch# 117 was outside of ±3 standard deviations at several time points. This graphical presentation of new batch in real-time revealed that the batch was not a good batch and the cause of deviation to be addressed immediately.
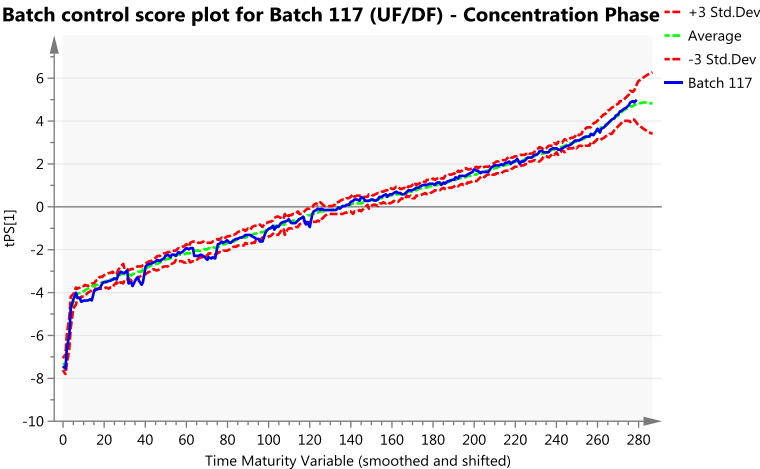

Figure 11: The Batch Score Plot for the UF/DF batch# 117. The plot shows that it was going outside of ± 3 standard deviation at several places.

119

In order to find out the cause for the batch deviation, the batch plot for each variable was evaluated. It revealed that the feed pressure and feed flow sensors were malfunctioning. Figure 12 shows the feed pressure and feed flow contribution plots with specific time points of where the batch was outside of the confidence limit.
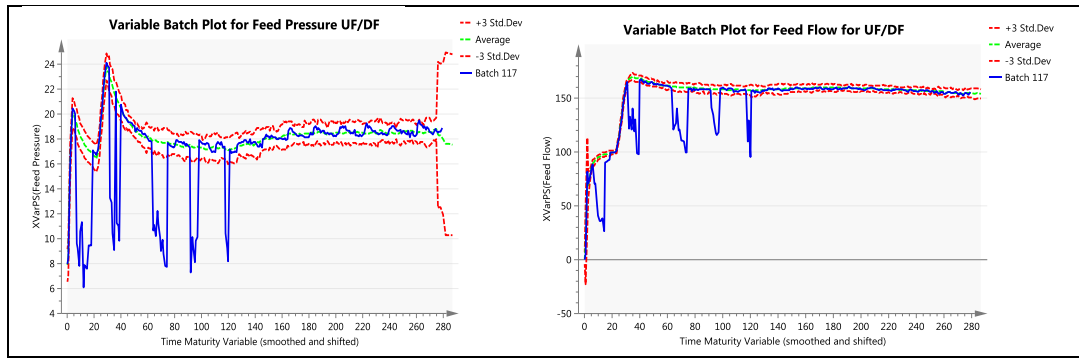

Figure 12: The Variable Batch Plot for Feed Pressure and Feed Flow for batch# 117. The plot shows the time of when the feed pressure and feed flow were outside of ± 3 standard deviation confidence limits.

Benefits of MVDA PAT Tool

The above examples show that the use of a multivariate process monitoring tool enables the real-time process monitoring for the timely fault detection and analysis. The tool also helps in detecting the root cause.

In an ideal situation complete implementation of a lifecycle approach using QbD and PAT tools, the raw material and equipment related process variability could be identified early on leading to thorough process understanding and effective process control of a particular process. The process scientist can take scientific, risk based decisions to justify the changes made within the operating space. This will result into tremendous benefits to the manufacturers and to the regulatory agencies during approval process as listed below:

Each of the following benefits can be transformed into financial benefit. There are a lot of cost savings associated with each benefit. Ultimately, cost savings in the long range to meet the product and prize demand is the ultimate objective along with meeting the regulatory expectation.

Benefits [6, 13, 25]:
1. Thorough understanding of the process makes it easy to establish an operating space.
2. Provides the scientific knowledge and classification of each and every input and output parameters
3. Early fault detection
4. Process changes within operating space may not require additional PPQ runs which may reduce the frequent regulatory review and approval
5. Scientific and data driven justification for every future modification will save time during investigations
6. Consistent and better quality product; a more robust manufacturing process
7. Consistent product quality with minimal variability and higher yield
8. Real time release

CONCLUSION

The latest Q11 [5] guideline, along with a new Process Validation guidance by FDA for process validation [7] about Continued Process Verification are eliciting the same message that "Process validation should not be viewed as a one-off event. A lifecycle approach should be applied linking product and process development, validation of the commercial manufacturing process and maintenance of the process in a state of control during routine commercial production".

With this, the regulatory agencies are encouraging the manufacturers to implement the QbD and PAT. In the traditional approach, set points and operating ranges for process parameters are defined and the control strategy based on the demonstration of process reproducibility and testing to meet the established acceptance criteria. There are certainly flaws in the traditional approach which can be improved with an enhanced approach. The enhanced approach is backed by risk management studies, scientific knowledge, and process understanding. The process knowledge and understanding gained during the process design and process qualification stages can be utilized to develop appropriate control strategies which are applicable over the lifecycle of the product.

The RT-MSPM system used in the study can be applied to any legacy manufacturing process. Even if it requires investment of resources and time, it can be applied easily with appropriate management support. It can definitely offer a significant enhancement to process understanding, process monitoring, scientific thoroughness in decision making qualitative and quantitative performances and cost savings. The use of MVDA tool provides an opportunity to have better control on monitoring the process real-time so that issues can be identified and addressed quickly.

There are multiple benefits of implementing a PAT during the drug development phase, and manufacturing phase. In the development phase, it can provide process understanding. In manufacturing phase, it can help gain the real-time monitoring and assurance that the batch is moving in a right direction. It can also eliminate off-line testing and minimize batches that are out of specification [25].

If QbD and PAT are applied to new products, a sufficient number of bench scale, development runs and engineering runs are performed to obtain sufficient data then the justification for a number of batches required for PPQ can be easily made prior to process validation campaign.

If the MVDA tool is applied to an existing product, then every batch can be monitored in real-time similar to a process validation batch. The early fault detection and any deviation from the targeted range can be detected in real-time. These capabilities not only provide financial benefit but also meet the regulatory expectation for continued/continuous process verification.

REFERENCES

1. Guideline on General Principals of Process Validation May, 1987 (REPRINTED MAY 1990)

2. Guidance for Industry Q8, Pharmaceutical Development, ICH, May 2006

3. Guidance for Industry Q9, Quality Risk Management, ICH, June 2006

4. Guidance for Industry Q10, Pharmaceutical Quality System, ICH, April 2009

5. Guidance for Industry Q11, Development and Manufacture of Drug Substances, ICH, November 2012

6. Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, September 2004

7. Guidance for Industry; Process Validation General Principles and Practices, January 2011, Revision 1

8. CFR - Code of Federal Regulations Title 21, FDA.gov

9. Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach, Final Report, September 2004

10. FDA's New Process Validation Guidance: Industry Reaction, Questions, and Challenges, By Mike Long, Hal Baseman, Walter D. Henkels, September 1, 2011)

11. Quality-by-Design, A Perspective from the Office of Biotechnology Products (OBP), http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4187B1_01_13-%20OBP-QbD.pdf

12. http://www.molnar-institut.com/HP/images/Design_space.jpg, accessed on 10/04/13, at 11:15am

13. PAT in Biologics Book (Eds. C.Undey, D. Low, J. Menezes, M. Koch) and the QbD in Biopharmaceuticals Book (Eds., A, Rathore)

14. https://encrypted-tbn2.gstatic.com/images?q=tbn:ANd9GcSMu3f6yxfsJZ4UzFl_PSbwHY6-vTOBpXyEmVstrN1X2t0h2_H, accessed on 10/12/13 at 4:00pm

15. Elements of Biopharmaceutical Production Series, Fourth Edition, Publisher Advance star scientific, BioPharm International

16. FDA, "SUPAC-IR Questions and Answers about SUPAC-IR Guidance," 1997, available at "http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm124826.htm. (See, e.g., "SUPAC-IR Questions and Answers about SUPAC-IR Guidance," describing the "first three production batches" as "validation batches;" Guide to Inspections of Oral Solid Dosage Forms Pre/Post-Approval Issues for Development and Validation, 1994, available at available at: "(http://www.fda.gov/ICECI/Inspections/InspectionGuides/ucm074928.htm, indicating that "at least three batches are needed to demonstrate consistency."). GXP

17. EU, Guidelines to Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use, Annex 15: Qualification and Validation (Brussels, July 2001)

18. FDA 2011 Process Validation Guidance: Process Validation Revisited, Paula Katz and Cliff Campbell, fda.gov

19. PDA Annual Meeting Post-Conference Workshop, "The FDA's Process Validation Guidance: Meeting Compliance Expectations and Practical Implementation Strategies" (San Antonio, TX, April 13–14, 2011)

20. Topic 2 – Stage 3 Process Validation: Applying Continued Process Validation Expectations to New and Existing Products, ISPE (*www.ispe.org/discussion-papers/stage-3-process-validation.pdf)*

21. [EMA Draft Guideline on Process Validation, EMA/CHMP/CVMP/QWP/70278/2012-Rev1, 29 March 2012]

22. http://en.wikipedia.org/wiki/Nelson_rules, accessed on 10/19/13 at 3:15pm

23. SIMCA User Guide, Software version 13.0.3

24. Multivariate and Megavariate Data Analysis Basic Principles and Applications (Part I), L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, and S. Wold, Umetrics

25. Using PAT for Real Time Monitoring and Quality Assurance in Chemical Development, Pharmaceutical Review, June 2010