2017

Biological Process: Characterization, Establishment of Shelf Life and Optimizing Method Transfers

Syama S. Adhibhatta
University of Rhode Island, syamaa@my.uri.edu

BIOLOGICAL PROCESS: CHARACTERIZATION, ESTABLISHMENT OF SHELF LIFE AND OPTIMIZING METHOD TRANSFERS

BY

SYAMA S ADHIBHATTA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2017

DOCTOR OF PHILOSOPHY DISSERTATION

OF

SYAMA S ADHIBHATTA

APPROVED:

Dissertation Committee:

Major Professor Jay Wang

Godi Fischer

Valerie Maier Speredelozzi

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2017

ABSTRACT

Biological processes are complex and can be modeled using combination of linear and non-linear models. During the cell culture process, cells may change or evolve and it is extremely important to understand the variability of the process to manufacture consistent product. In order to maximize the output from the process, process parameters need to be characterized and optimized. Data from process involves linear and non-linear patterns and some of the parameters are auto-correlated. One of the objectives of study was to compare Unsupervised Dimesional Reduction methods to Supervised machine learning algorithms applied to biopharmaceutical manufacturing process data and suggest a new 2-stage approach including a combination of unsupervised and supervised algorithms for better predictability. Analytical methods are used to measure the quality of the product. Main objective of methods transfer is to avoid release of product that does not meet specifications as well as avoid rejection of good product. The effect of sample size for establishing analytical method equivalency and comparison of statistical methods during assay transfers was performed and criteria for out of specification risk mitigation was recommended. Shelf life of a biopharmaceutical product is typically based upon the stability data. Factors that impact stability of a product and shelf life were studied in detail using multiple statistical models and criteria for choosing the appropriate model was recommended.

ACKNOWLEDGMENTS

I owe my deepest gratitude to my thesis advisor Dr. Jay Wang without whose support and encouragement my thesis could not have been completed. He was always available for his valuable suggestions and insights. I am extremely grateful for his endless patience in guiding me throughout my research and writing this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D study.

I am thankful to the faculty members of the Department of Industrial and Systems Engineering for the excellent instruction and advice I have received over the years. I would like to deeply acknowledge the members of my dissertation committee: Dr. Godi Fischer, Dr. Valerie Maier Speredelozzi, Dr. David Freeman and Dr. David Taggart for their support and advice.

I would like to extend my gratitude to Dr. Ali Cinar at the Illinois Institute of Technology, Chicago for providing me with the simulation software without which I would not be able to complete part of my research.

I am extremely thankful to my loving wife (and my proof reader) Santosha Adhibhatta without whose loving support and encouragement, I would not have been able to complete my PhD program. Her love and support carried me through the roughest times. She has been extremely supportive of my studies for which I express my heartfelt thanks. My heartful and loving thanks to my son Abhiram Adhibhatta for being my motivation for this project.

I express my heartfelt thanks to my parents Adhibhatta Suryanarayana Sastry and Adhibhatta Lakshmi for their moral and emotional support. I am extremely grateful to my father for instilling a confidence in me to pursue my higher studies and his support financially. My mother's endless understanding, patience and support is something I would cherish forever. I offer a special word of thanks to

my brothers Aditya Adhibhatta and Chandrasekhar Adhibhatta who have given me moral support throughout.

My heartfelt thanks to my father-in-law, Subrahmanyam Ammu and my mother in law Syamala Ammu who have given me the encouragement and support when needed.

Finally my endless thanks to all my relatives, friends and colleagues. I would like to give a special thanks to my friends Sivaram Vongala and Swarna Deepika Yadavalli who have been there for me and my family throughout.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Drugs manufactured from bio-technology based processes have been important in treating chronic and rare diseases such as cancer and enzyme deficiency disorders. Biopharmaceuticals are a major portion of pharmaceutical market. It is now commonly reported that a growing percentage, now over 40%, of all pharmaceutical industry research and products in development are now biopharmaceutical rather than small molecules. Progress in science have pushed biopharmaceutical manufacturing into another domain of unpredictability as of late. The FDA has approved at least 10 large molecules (biologics) in each of the past five years. Furthermore, over 900 biologics were in development as of February 2013, suggesting that there will be increased need for commercial production of biologics in the coming years as seen in Figure 1[1]. Biologic medicines such as vaccines are complex molecules made by or from living cells and are often infused or injected.

The emergence of biologics, which are complex than small molecule medicines and are frequently gotten from living cells, has introduced incredible difficulties to manufacturers. These molecules require multiple steps that entail the use of robust technology to ensure purity, consistency, and quality. Research and development, clinical and commercial manufacturing of these products require a good understanding of chemistry, manufacturing and controls. Biopharmaceutical processes are typically comprised of multiple processes operated in batch/continuous mode to produce proteins, and have complex biological mechanisms that result in non-linear and time-variant process dynamics. Manufacturers place a high emphasis on ways to improve the consistency and predictability of processes over time to ensure product quality. Biopharmaceutical manufacturing process consists of

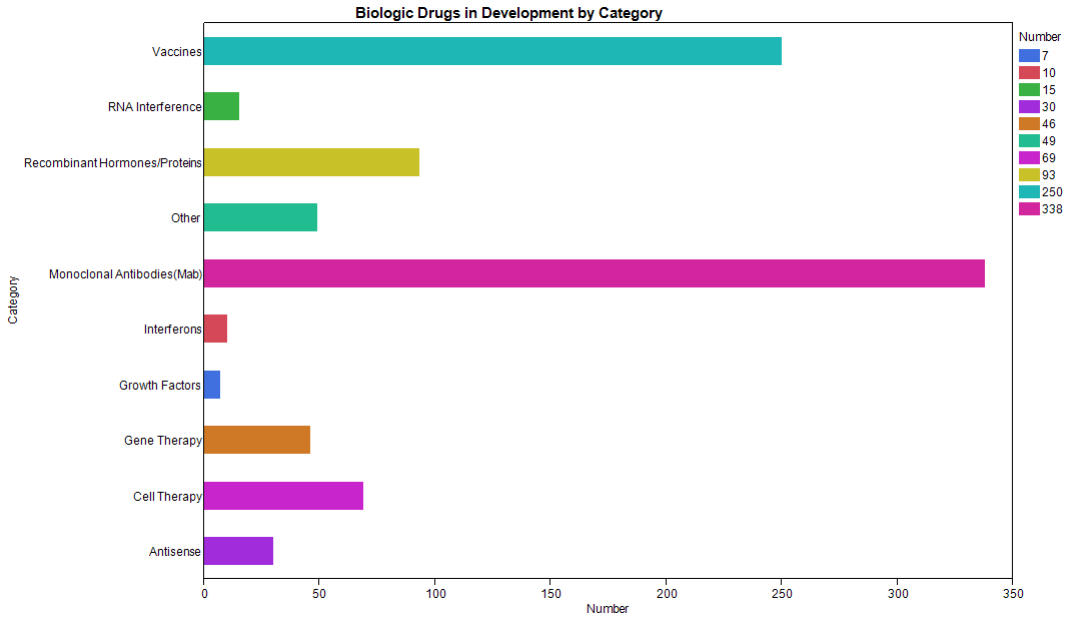


Figure 1: Biologic drugs in development by category[1]

cell culture(upstream) and Purification(Downstream) operations. The upstream process is defined as the entire process from early cell isolation and cultivation, to cell banking and culture expansion of the cells until final harvest (termination of the culture and collection of the live cell batch). Upstream process consists of inoculum, media development, scale up of processes and optimization of growth so that cells can grow and protein can be extracted. The downstream manufacturing process purifies the protein from cell mass transferred and processed to meet purity and quality requirements. Downstream processing consists of cell disruption, purification section and a polishing section.

Modeling the relation between input parameters, process conditions to the output attributes require advanced analytical tools and complex models. After the manufacturing of therapeutic proteins, shelf life or the stability time period need to be established. Stability testing for biologics consists of more quality attributes and stability profiles vary from product-to product and lot-to-lot, result in identifying patterns and interpreting results more difficult. The measuring assays for biologics

are complex and may be more variable, making it difficult to interpret results. The purpose of stability testing is to provide an understanding on how the quality of an active substance or finished product varies with time under the influence of a variety of environmental factors such as temperature, humidity, and light, and to establish a re-test period for the active substance or a shelf life for the finished product and recommended storage conditions. Stability testing thus evaluates the effect of environmental factors on the quality of the drug substance or a formulated product which is utilized for prediction of its shelf life, determine proper storage conditions and suggest labeling instructions. Moreover, the data generated during stability testing is an important requirement for regulatory approval of any drug or formulation [3].

Major reason for stability testing is the concern for the well-being of the patient suffering from the disease for which the products is designed. The product may degrade and lose its activity due to storage conditions and other factors resulting in failure of therapy. Other reasons include ensuring the reputation of the producer by guaranteeing that the item will hold stable for use as for all practically important qualities for whatever length of time that they are available. Other benefits of stability studies at the developmental stage or of the marketed products are to provide a database that may be of value in selection of adequate formulations, excipients and container closure systems for development of a new product, to determine shelf life and storage conditions for development of a new product, preparation of registration dossier, to substantiate the claimed shelf life for the registration dossier and to verify that no changes have been introduced in the formulation or manufacturing process that can adversely affect the stability of the product [3, 4]. Shelf life is commonly estimated using two types of stability testing: real-time stability tests and accelerated stability tests. In real-time stabil-

ity testing, a product is stored at recommended storage conditions and monitored until it fails the specification. In accelerated stability tests, a product is stored at elevated stress conditions (such as temperature, humidity, and pH). Degradation at the recommended storage conditions can be predicted using known relationships between the acceleration factor and the degradation rate.

Analytical methods are used to measure the quality of the product. Typically, similar to a manufacturing process, they are developed in a pre-commercial setting and need to be transferred to commercial manufacturing. Main objective of method transfer is to avoid release of product that does not meet specifications (safety impact) often considered as Type I error as well as avoid rejection of good product (business impact Type II error). Because of these potentially high-impact implications, an assay cannot be used by the new lab until transfer is successfully completed. Method transfer activity is the procedure of building up a qualified diagnostic test methodology that begins in another facility. It documents that method execution is comparable across sites. The assay transfer process involves robust studies from the receiving and sending labs to establish equivalence. Similar to method validation, analytical method transfers provide a snapshot of point in time and should not be the single means to approve usage of the method. For a successful method transfer, the criteria need to be statistically and scientifically meaningful. Different statistical methods including significance tests, absolute acceptance criteria and equivalence tests are described and compared with detailed examples. The processes may be dynamic and things might change over time. Analytical method transfers should be used along with method validation or verification and continuous assay monitoring to establish robustness of the method. Based upon the assessment of analytical procedure, statistical equivalence tests can be used to establish equivalence between sending and receiving sites. Equivalence

tests include both practical equivalence limits and also take in to consideration of statistical risks.

1.2 Statement of the Problem

The biologics supply chain consists of several processes, as shown in Figure 2.

Figure 2: Biologics Supply Chain

Bio-technology manufacturing processes are complex and consist of multiple processing operations with numerous parameters including inputs and their correlation structures impact the final output or productivity of the process. The first two stages shown in Figure 2 are the manufacturing processes.

For the purpose of this study, research will be focused on upstream (cell culture) part of the manufacturing process. Cell culture is the goal of using recombinant technology to make cells convert one product into another. In most cases, this is either a feedstock like glucose or aminoacids like glutamine. Compared to chemical synthesis used in small molecule drugs where the ingredients tend to be consistent, cells may change or evolve during the process and it is extremely important to understand the variability of the process to manufacture consistent product. Several parameters include inputs (Ex: pH, Oxygen, Glucose, Amino acids, Temperature, pressure, agitation etc.,) that can be changed or adjusted during the course of the process to achieve consistent quality. Statistical correlations are computed to identify relation between inputs and output. Correlations

were classified as time independent and time linked. Time independent correlations are static i.e., they exist throughout the bioreactor run regardless of phase or time period. Time independent correlations form the basis for observation wise (also referred to as variable wise) modeling and are useful for identifying relationship that are continuously present in direction and magnitude between process parameters. Time linked correlations are dynamic i.e., they exert an impact at one period of time and not another. Time linked correlations form the basis for batch wise modeling and are very useful for highlighting time point in a bioreactor run when one parameter exerts an influence on another parameter with variable direction and/or magnitude.

Models were constructed by using multiple techniques: parsimonious curve-fitting, partial least squares and support vector machine algorithm. Data from process involves linear and non-linear patterns and some of the parameters are auto-correlated.

The third step in the Biologics supply chain is quality control. Assays that measure the quality of the product need to be transferred from development/validation lab to production lab and sometimes the assays have to be transferred to contract manufacturing labs. The purpose of the transfer is to demonstrate that the two labs are comparable to avoid re-validation of the assay. Comparability and equivalency between sending and receiving labs and the number of replicates that required for establishing equivalence need to be optimized for resource efficiency and reducing producers risk. Based upon current practices, means are compared between the receiving lab results, to the average results of the sending site, considering sending site as the reference value. Comparability should be established based upon both accuracy (comparing the means between the sites) and also the precision (comparing the variances) criteria. Some of the current statistical approaches for

establishing the comparability of accuracy criteria include using students t-test for identifying any statistical mean difference between the sites or using equivalence test/Two-one sided t-test(TOST) that takes into consideration any practical significance. Determining the appropriate acceptance criteria is critical to decide the appropriate study design and sample size required for method transfers. Literature review suggests number of approaches are proposed for analytical method transfer. This study will provide more details on establishing the acceptance criteria and sample size required for successful method transfer optimizing patient and producers risk.

The process of establishing shelf life for new products and extension of shelf life for existing products will be studied based upon the current regulations and statistical techniques. This is the final step in the supply chain in Figure 2. Based upon International Conference on Harmonisation(ICH) Q1E [5], for quantitative attributes that are used for stability determination, linear and nonlinear regression and statistical modeling through poolability tests are used for determining the estimated shelf life of a drug product. According to these guidelines, results from at least three stability registration batches are obtained at pre-determined storage times. A linear regression model can be used for common intercept and slope, separate intercepts and common slope or separate intercepts and separate slopes for shelf life determination. ICH guidelines suggest a fixed batch methodology to establish a shelf life. This approach assumes that batches used in establishing shelf life are representative of the products distribution in terms of the products manufacturing process. This approach only considers within-batch variability into consideration and batch to batch variability is ignored. The random effects measure the batch to batch variability. Based upon the ICH guidelines, different factors will be included in the study. These factors are poolability of data, worst case lot and fixed and

random effects for establishing and extension of shelf life[6]. Criterion for selecting an appropriate model(Full/Random/Mixed) will be defined and comparison of shelf life using existing guidance and proposed methodology will be included for justification of extension.

1.3 Significance of the study

Biological processes are complex and can be modeled using combination of linear and non-linear models. During the cell culture process, cells may change or evolve and it is extremely important to understand the variability of the process to manufacture consistent product.In order to maximize the output from the process, process parameters need to be characterized and optimized. Statistical methodologies can be applied to manufacturing process data to analyze the correlations for process optimization. Data collected from biological manufacturing process is vast and auto correlated. The data cannot be analyzed by linear regression methods due to the nature of time varying and autocorrelation. Due to high dimensionality of the data, few methods such as Principal component analysis(PCA) and Partial least squares (PLS) were chosen for dimensional reduction. PCA and PLS methodologies belong to classification of unsupervised algorithms. Dimensionality reduction is based upon the inherent structure in the data, in an unsupervised manner (A model is prepared by deducing structures present in the input data) to summarize or describe data using less information.

The same dataset will also be analyzed by supervised machine learning algorithms like Symbolic Regression (SR) and Support vector machines (SVM). Supervised learning can be explained by building a model through iterative training process until the model achieves desired level of accuracy for predictions. The present study compares machine learning supervised learning models, unsupervised dimensional reduction techniques and suggests an alternate method of combining

supervised and un supervised algorithm (2-stage) for better prediction purposes. These predictions can be used for scheduling purposes for downstream processing and also for capacity planning and scheduling optimization.

Shelf life of a biopharmaceutical product is typically based upon the stability data. The shelf life of a pharmaceutical product is the maximum time at which the response of a stability limiting characteristic for all parameters (or other unit) in the batch does not exceed the specification limit. ICH Q1E provides general guidance for establishing the shelf life based upon average of the results, but does not provide any guidance on individual results. Also, there is no guidance provided on extending shelf life of a product based upon historical data.ICH approach estimates shelf life indirectly by interval estimate on the mean. According to the guidelines, shelf life for each individual batch is computed and a minimum value is considered as the shelf life for the product. The approach might not address the future batches and does not provide clarity around product meeting specifications. Previous approaches considered batch as fixed factor and proposed approach will consider batch as random effect to address future batches.

The process for establishing or extending shelf life depends upon a number of factors and hence it is essential to understand each of the factors in detail in order to determine the appropriate statistical methodology. Some of the important factors that need to be considered are poolability of lots, determination of Worst case Lot (WCL) and the type of model (Full/Random/Mixed). Hence, it is required to conduct a detailed study of the effect of these factors and then recommend the best model based on defined criteria, which is the goal of this work.

Often, assay methods need to be transferred between development and validation to commercial laboratories and also between contract manufacturing laboratories. The accuracy and precision of the sending and receiving labs should be compa-

rable and should not have statistical significant differences.There are no official regulatory guidelines that exist for assay transfers.FDA guidance for validation of bioanalytical methods which is assay transfer between labs is considered as partial validation.However, no guidance regarding the protocol design, number of replicates/samples required establishing equivalence and comparability is present in neither agency guidelines nor ICH documentation [7]. Hence a detailed study of the effect of sample size for establishing equivalency, comparison of methods for establishing goal posts for equivalency and comparing accuracy, and precision will be studied for practical applications.

1.4 Objectives and Deliverables of the study

- **Objective 1**: The First objective of this study was to compare Unsupervised Dimensional Reduction methods to Supervised machine learning algorithms applied to biopharmaceutical manufacturing process data and suggest a new 2-stage approach that includes a combination of unsupervised and supervised algorithms for better predictability.

  Deliverable: Comparison of two unsupervised dimensional reduction methods (PCA, PLS) to supervised machine learning methods (Symbolic regression, Support Vector regression) and suggest a Scores Based Symbolic Regression methodology applied to penicillin fermentation process data.

- **Objective 2**: Second objective of the study was to identify and recommend appropriate statistical methodology for establishing shelf life and extension of shelf life for biopharmaceutical products.

  Deliverable: Study the impact of factors like poolability of lots, determination of Worst case Lot (WCL) and the type of model (Full/Random/Mixed) on shelf life determination. Selection of model using defined criterion will be recommended.

- **Objective 3**: Third objective of the study was to study of the effect of sample size for establishing equivalency and comparison of methods during assay transfers and to recommend criterion for out of specification risk mitigation. **Deliverable**: Analysis of sample size required at different levels of power and Type I error will be performed. Optimal samples required for method transfer will be recommended based upon balancing power and producers risk. A better approach for analytical method transfers is proposed, that involves process capability calculations that can avoid potential out of specification results.

1.5 Structure of the study

Chapter two contains a background and literature review regarding bioreactor process optimization, shelf life determination and extension of shelf life, analytical method transfers and proposed criteria for establishing acceptance criteria.

Chapter three discusses the methodology considered and used in the study. It describes the research methods, data collection and data analysis procedure.

Chapter four presents the results of the study and the practical implementation of the findings.

Chapter five presents the conclusions of the study including its limitations. Suggestions for further research is also discussed in this chapter.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

Data from Biologics manufacturing processes are available since the start of manufacturing drugs like penicillin, but tools and methodology to monitor and analyze have only matured only in recent years. This has been brought about by advances in detecting technology for capturing and storing multidimensional data and faster computing technologies that has enabled complex transformations and manipulations of the large datasets quickly and economically. Before reviewing the literature, we discuss various statistical methodologies that were used for monitoring and analyzing biologics processes, stability data analysis and method transfers.

2.1 Statistical Methodologies for analyzing Process data

2.1.1 Partial Least Squares (PLS)

PLS is a regression methodology that is most suitable for data with high auto correlation between cause variables. Partial Least Squares can be applied to batch data where data can be divided into inputs (X variables) and outputs (Y variables). The methodology works by selecting factors of input variables in a sequence which successively maximizes the explained covariance between the input and output variables. Based upon a matrix of input variables, X (of size $m \times n_x$, where m is the number of observations and n_x is the number of input variables), and output data, Y (of size $m \times n_y$, where n_y is the number of output variables), a factor of the input data, t_k (length m), and effect data, u_k (length m), is evaluated, such that:

$$X = \sum_{k=1}^{n_p<n_x} t_k p_k^T + E \text{ and } Y = \sum_{k=1}^{n_p<n_x} u_k q_k^T + F \tag{1}$$

The vectors t_k are mutually orthogonal. These vectors and the u_k vectors are selected to maximize the covariance between each pair (t_k, u_k). E and F are errors and p_k and u_k are referred to as loading vectors as explained by O. Marjanovic et al., [8]. Linear regression is performed between the t_k and the u_k vectors, to produce the inner relationship, such that:

$$u_k = b_k t_k + \epsilon_k \tag{2}$$

where b_k is a regression coefficient, and ϵ_k refers to the prediction error. PLS provides a mechanism: a model to select appropriate scores, latent variables/Principal components and length of the model.Typically, relatively few principal components can explain most of the variation in the input variables that impacts out variables. Required number of principal components can be selected by comparing the amount of variation explained and how well the model can predict future observations. Cross-validation can also be applied to obtain optimized model.Process Data encompassing multiple batches with associated variables changing with time can be envisioned as a three-dimensional cube with batch, variable, and time represented by terms i, j, and k, respectively. Three dimensional spaces can be unfolded to two dimensional space in three ways as shown by Lee et al. [2].

1. Batches $\times$ times for each specific variable (observation-wise unfolding)

2. Variables $\times$ time for each specific batch (batch-wise unfolding)

3. Batches $\times$ variables at each specific time (time-wise unfolding)

Each method is useful for addressing a variability. Observation-wise unfolding(also referred to as variable-wise unfolding) can be used to obtain information about the variability among the batch variables. This approach is useful for uncovering time independent relationships between variables. Batch-wise unfolding facilitates the

analysis of the variability among batches. This is a particularly useful method when applied to understanding how process variability impacts process outcome. The two unfolding methods depicted above are commonly used for batch process monitoring and have been used extensively in this study and is depicted in Figure 3. The third unfolding method time-wise unfolding is useful for analyzing variability

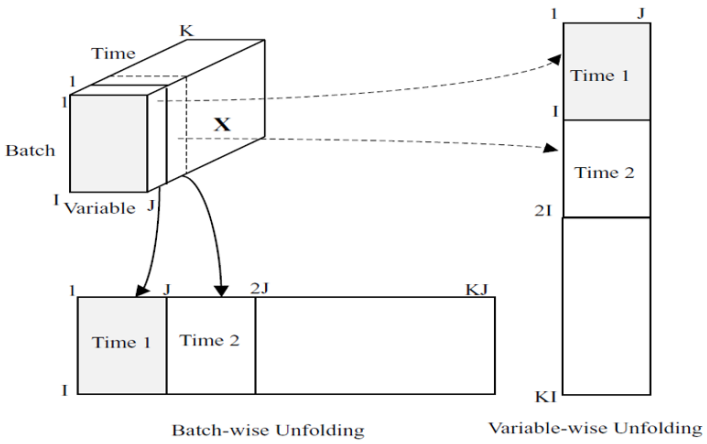


Figure 3: Depiction of batch-wise and variable-wise data unfolding [2]

among samples. Time-wise unfolding is more sensitive to detecting variations and deviations from normality in process monitoring. It is not as useful for system dynamic modeling involving variable to variable relationships or batch to batch variation and has not been used in this study.

2.1.2 Symbolic Regression

Symbolic regression is an established method based on evolutionary computation for searching the space of mathematical expressions while minimizing various error metrics. Unlike traditional linear and nonlinear regression methods that fit parameters to an equation of a given form, symbolic regression searches both the parameters and the form of equations simultaneously. Initial expressions are formed by randomly combining mathematical building blocks such as algebraic operators $(+, , , \times)$, analytical functions (for example, sine and cosine), constants,

and state variables. New equations are formed by recombining previous equations and probabilistically varying their subexpressions. The algorithm retains equations that model the experimental data better than others and abandons unpromising solutions. After equations reach a desired level of accuracy, the algorithm terminates, returning a set of equations that are most likely to correspond to the intrinsic mechanisms underlying the observed system.

Eureqa software will be used for application of symbolic regression to the proposed study of penicillin manufacturing process. Symbolic Regression has been described in detail by V. Aryadoust [9]. Once the operators are selected, symbolic regression uses evolutionary algorithms (EAs) that are similar to Darwins evolution theory [10, 11]. The process starts with initialization phase, where first generation of mathematical functions start fitting the data and solutions are randomly generated. Potential solutions fit is evaluated comparing to one or more fit indices[12]. Based on the results of the fit indices, most fit solutions are chosen as primary solutions and modified further to reproduce new solutions for use in successive iterations also called as generations. Evolutionary algorithm based symbolic regression must maintain diversity in generations of solutions. If not, the process will end in pre-mature convergence [13]. Stagnation based on convergence of solutions without an optimal solution point is called a local optimum and impacts the fit and precision of solutions. Multiple technical softwares were developed to avoid pre-convergence and Eureqa is one of them and used in the current research. The software uses three optimizations: Cross over, mutation and agefitness pareto optimization.

In crossover optimization, two or more primary solutions are chosen to reproduce one or more secondary solutions. Eureqa uses a one-point crossover: a random point on each primary solution is taken, and divided into sections before

and after the crossover point. All these sections are combined to two secondary solutions and replace sub optimal solutions from the previous generations [14, 11].

Mutation is another optimization technique where a single solution is partially or entirely modified to generate a better solution [12, 15]. Cross over is a more commonly used technique compared to mutation. According to Michalski and Ryszard [16], 50 % functions go through crossover, but only 1% follow mutation. The technique helps in preventing pre-mature convergence by diversifying the solutions.

The third optimization used in Eureqa to avoid premature convergence is age-fitness pareto optimization, where evolving solutions are selected based upon age of the solution and fit to the data [17, 18]. Age-fitness pareto optimization selects the low aged and fittest solutions, and over iterations try to minimize error, improve fit and maximize predictive capability. Schmidt and Lipson [18] indicated that this optimization technique outperforms other available techniques such as deterministic crowding and age-layered optimization. Multiple statistics can be used to evaluate the effect of cross-validated solutions , but Eureqa uses Mean absolute error, Mean squared error, Correlation coefficient and R^2 goodness of fit.

Multiple fit statistics are used to assess the efficacy of cross-validated solutions, as follows:

1. Mean absolute error (MAE): Assuming that error of measurement follows a double exponential distribution, MAE estimates the difference between predicted and observed values. The closer the MAE to zero, the higher the precision.

2. Mean squared error (MSE): Like MAE, MSE estimates the difference between the predicted and observed values. Unlike MAE, however, MSE assumes that error of measurement is normally distributed. Lower MSE indices indicate

higher precision.

3. Correlation coefficient (R): ranging between 0 and 1, the R index quantifies the correlation between observed and predicted values. Values above 0.7 indicate significant correlation between model-estimated and actual output.

4. R^2 goodness of fit: The R^2 index indicates the percentage of output (values of the dependent variable) that can be explained by the input or independent variables. It ranges between 0 and 1, with values closer to 1 indicating higher predictive power.

Eureqa assigns each operator in the solutions a numerical value indicating its complexity. For example, addition and negation have a value of 1, whereas logistic and step functions have a value of 4. The total complexity of each solution is the sum of the complexity values of the operators used in that solution. As less complex models with low errors of measurement are desirable, sometimes the researcher has to make a trade-off between complexity and fit by choosing less complex models over more complex models with slightly better fit statistics [13]. This decision can also be based on the sensitivity of each input variable, that is estimated according to its frequency of occurrence across all solutions, as well as the sensitivity or relative impact of each input variable on the output within each model.

2.1.3 Support Vector Machines(SVM)

SVM is an excellent method aiming to optimize finite data points based on statistical learning theory (SLT) [19, 20] advanced by Vapnik in 1990s. It adopts the structure risk minimization (SRM) principle and avoids the complex computing kernel function of low dimensions instead of dot-matrix of high dimensions space. The main idea of SVM is described by Peng et al. [21] and documentation is provided by Mathworks [22]. This is further discussed below.

Linear SVM Regression: Primal Formula

Suppose we have a set of training data where x_n is a multivariate set of N observations with observed response values y_n. To find the linear function $f(x) = x^{'}\beta+b$, and ensure that it is as flat as possible, $f(x)$ is found with the minimal norm value ($\beta^{'}\beta$). This is formulated as a convex optimization problem to minimize

$J(\beta) = \frac{1}{2}\beta^{'}\beta$,

subject to all residuals having a value less than ϵ; or in equation form

$\forall n : |y_n - (x_n^{'}\beta + b)| \leq \epsilon$. It is possible that no such function $f(x)$ exists to satisfy these constraints for all points. To deal with otherwise infeasible constraints, slack variables ξn and $\xi * n$ are introduced for each point. Including slack variables leads to the objective function, also known as the primal formula [23]:

$J(\beta) = \frac{1}{2}\beta^{'}\beta + C\sum_{n=1}^{N}(\xi_n + \xi_n^{'})$,

subject to

$\forall n : y_n - (x_n^{'}\beta + b) \leq \epsilon + \xi_n$.

$\forall n : (x_n^{'}\beta + b) - y_n \leq \epsilon + \xi_n^{*}$.

$\forall n : \xi_n^{*} \geq 0$

$\forall n : \xi_n \geq 0$

The constant C is the box constraint, a positive numeric value that controls the penalty imposed on observations that lie outside the epsilon margin (ϵ) and helps to prevent overfitting (regularization). This value determines the trade-off between the flatness of f(x) and the amount up to which deviations larger than ϵ are tolerated. The linear ϵ-insensitive loss function ignores errors that are within ϵ distance of the observed value by treating them as equal to zero. The loss is measured based on the distance between observed value y and the ϵ boundary. This is formally described by

$$L_\epsilon = \begin{cases} 0, & |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon, & \text{otherwise} \end{cases}$$

Linear SVM Regression: Dual Formula

The optimization problem previously described is computationally simpler to solve in its Lagrange dual formulation. The solution to the dual problem provides a lower bound to the solution of the primal (minimization) problem. The optimal values of the primal and dual problems need not be equal, and the difference is called the "duality gap." But when the problem is convex and satisfies a constraint qualification condition, the value of the optimal solution to the primal problem is given by the solution of the dual problem. To obtain the dual formula, a Lagrangian function is constructed from the primal function by introducing non-negative multipliers α_n and α_n^* for each observation x_n. This leads to the dual formula, where we minimize

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^{'}x_j + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) \quad (3)$$

subject to the constraints

$\sum_{n=1}^{N}(\alpha_n - \alpha_n^*) = 0$

$\forall n : 0 \leq \alpha_n \leq C$

$\forall n : 0 \leq \alpha_n^* \leq C$

The β parameter can be completely described as a linear combination of the training observations using the equation

$$\beta = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)X_n \quad (4)$$

The function $f(x)$ is then equal to

$$f(x) = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)(X_n X) + b \quad (5)$$

Karush-Kuhn-Tucker (KKT) complementarity conditions are optimization constraints required to obtain optimal solutions. For linear SVM regression, these conditions are :

$\forall n : \alpha_n(\varepsilon + \xi_n - y_n + X_n^{'}\beta + b) = 0$

$\forall n : \alpha_n^*(\varepsilon + \xi_n^* + y_n - X_n^{'}\beta - b) = 0$

$\forall n : \xi_n(C - \alpha_n) = 0$

$\forall n : \xi_n^*(C - \alpha_n^*) = 0.$

LSSVM is a modified version of SVM that was described by Suykens et al.[24] and utilizes the equality constraints to replace the original convex quadratic programming problem. In LSSVM for function estimate, the following optimization problem is considered:

$$\underset{\omega,b,\xi}{min} J(\omega, \xi) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{k=1}^{N}\xi_k^2, \gamma > 0 \tag{6}$$

subject to the equality constraints:

$y_k = \omega^T.\phi(\xi_k) + b + \xi_k, k = 1, 2, ...N$ where γ is the regularization parameter, ϵ_k is the error between the real output and estimated value at the k_{th} sample point. Similar to the approach of ordinary least squares (OLS) and PLS, SVM also finds a linear relation between the repressors (input variables, X) and the dependent variables (y). The cost function (the function that is minimized to obtain the best regression model) consists of a two-norm penalty on the regression coefficients, an error term multiplied by the error weight, C, and a set of constraints. Using this cost function, the goal is to simultaneously minimize both the coefficients size and the prediction errors (function smoothness and accuracy). The ideology of the LS-SVM method is very close to that of SVR, but in this case the more usual sum of the squares of the errors is minimized, and no ϵ-based selection is made between samples. This is common to all least square methods. This can make the final model more accurate and less computationally expensive.

2.2 Establishing stability/shelf life

The purpose of stability testing is to provide understanding on how the quality of an active substance or finished product varies with time under the influence of a variety of environmental factors and to establish a re-test period for the active substance or a shelf life for the finished product and recommended storage conditions. Major reason for stability testing is the concern for the well-being of the patient suffering from the disease for which the products is designed. The product may degrade and lose its activity due to storage conditions and other factors and thus leading to failure of therapy. According to International Conference Harmonisation(ICH) Q1E guidelines [5], appropriate statistical methodologies should be used for stability data analysis. The significance of the analysis is to calculate with high confidence, expiry date during which a quantitative attribute will maintain with specifications for all future batches manufactured, and stored under similar circumstances. Linear regression/regression analysis is considered an appropriate statistical methodology for establishing shelf life. Based upon the ICH Q1E guideline, an appropriate approach to retest period or shelf life estimation is to analyze a quantitative attribute (e.g. assay, degradation products) by determining the earliest time at which the 95% confidence limit for the mean intersects the proposed acceptance criterion. If a parameter decreases with time, the lower one-sided 95% limit should be compared to the acceptance criterion. If a parameter increases with time, the upper one-sided 95% confidence limit should be compared to the acceptance criterion. If a parameter either increases or decreases, or whose direction of change is not known, two-sided 95% confidence limits should be calculated and compared to the upper and lower acceptance criteria. The document provides guidance for data analysis for a single batch, one-factor and multi-factor, full-design, bracketing design, matrix studies including testing for poolability of

batches. As per the guideline, an extrapolation of stability data assumes that the same change pattern will continue to apply beyond the period covered by long-term data. The correctness of the assumed change pattern is critical when extrapolation is considered. Statistical methods can be applied to test the goodness of fit of the data to the assumed line or curve. No such internal check is possible beyond the period covered by long term data. Thus, a retest period or shelf life granted on the basis of extrapolation should always be verified by additional long-term stability data as soon as these data become available.

Analysis of covariance is the most commonly used statistical methodology for stability data analysis. Analysis of covariance (ANCOVA) is an extension of linear regression and an important kind of multiple regression that involves two predictor variables: one continuous (e.g. time) and one categorical (e.g. batch of material). Similar to simple linear regression, simple ANCOVA fits straight lines to response measurements over time: one line for each level (i.e., batch) of the categorical variable. A key objective of ANCOVA is to determine whether the profiles for all batches are best described as having a common-intercept-common-slope (CICS) model, a separate-intercepts-common-slope (SICS) model, or a separate-intercepts-separate-slopes (SISS) model. As per ANCOVA, models are chosen based on comparing slopes or intercepts. US food and Drug administration recommends a p-value < 0.25 for significance for a model to estimate shelf life.

2.3 Analytical Method Transfers

Analytical method transfer is a process that involves transferring analytical method from a sending lab to receiving lab to ensure that a method is executed similar at both labs. Analytical method transfer is a key component of technology transfer between process development laboratory to commercial manufacturing. Regulatory agencies require that transfer between labs are documented and equiv-

alence is established. The objective of the method transfer is to ensure that results obtained by receiving lab are reliable and comparable to sending laboratory to meet the specifications of a product and address any bias between the labs. Multiple agencies proposed recommendations for transferring analytical methods including ISPE (International society of pharmaceutical Engineers), WHO (world health organization), United States Pharmacopia(USP). All these recommendations are not explicit and vary across agencies.

ISPE provided initial guidance on analytical transfers in 2003 as a part of its technology transfer guideline [25]. According to the guideline, the scope of a transfer if defined as follows: The Receiving Unit can routinely reproduce the transferred product, process, or method against a predefined set of specifications as agreed with the Sending Unit. Per the guideline, focus was on well-defined documented process and should include knowledge transfer. It provides recommendation on experimental design and acceptance criteria for different type of assays and minimum number of batches required for establishing equivalency [26].

WHO published a guideline in 2011 [27], similar to ISPE guideline that includes outline of the transfer process, responsibilities of sending and receiving units and documentation requirements. USP$<$ 1224 $>$ transfer of analytical procedures does not describe details of experimental designs and acceptance criteria, but provides general requirements for the transfer process. Different types of transfer methodologies were described. They are Comparative testing, Covalidation, Complete or partial (re-)validation and Transfer waiver.

In the pharmaceutical industry, there is a continuing need to evaluate bias or changes on test results to improve the test process performance. It is important that any bias or change will have negligible effect on test results for a characteristic of a material. Statistical hypothesis testing offers a rigorous, objective approach

to distinguishing truly significant differences in measurements from noise. The most familiar statistical test is the conventional Students t-test, which has been used extensively for the evaluation of bias or changes. In a two-sample t-test, two numerical test results are considered not equivalent when there is a significant difference between the means of two. However, a nonsignificant result implies only that equivalency cannot be ruled out. Consequently, the risk of incorrectly concluding equivalence can be very high. Another common criticism of the conventional t-test is that the strict definition of equivalence in Students t-test may result in the detection of differences that are not functionally or practically meaningful and do not identify the practically meaningful differences. An Equivalence Test is a statistical technique used to demonstrate the equivalence between two test results based on the practical significance. Two test results are considered equivalent when the mean difference is within prescribed limits, termed equivalence limits, for which smaller differences are considered practically equivalent. The equivalence limits are usually predetermined based on the potential impact and risk based on scientific knowledge and clinical relevance.

Schuirmanns two one-sided test (TOST) [28] is used to conduct the equivalence testing on numerical data from two sources. The standard practice for the TOST is provided by ASTM International [29]. Two sample TOST is the statistical methodology used to establish equivalency between the labs.

Benefits of TOST compared to 2-sample t test were described by Limentani, Giselle et al. [30]. The two-sample t test allows comparison of the mean values of two data sets by the calculation of the test statistic.

$$T = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{7}$$

The null hypothesis of the two-sample t test is that the mean values of the two data sets are equal. Although it is an appropriate test for proving that two data sets are

different, it has some problems using it for equivalence. One of the problems is that an increase in pooled variance and decrease in n results in a smaller T-value and make it difficult to prove that mean values are not equal. Other problem is that the results of two-sample t test may lead to conclude that a statistically significant difference exists between the mean values when the magnitude of the difference is of no practical importance. This need to be justified with an explanation when the difference is minor and has no practical significance. TOST is most suitable for equivalence testing compared to two sample t test.

2.3.1 Two Sample TOST Statistical Hypotheses

The two sample TOST determines the equivalence of the means of test results from two testing processes. The data from two populations are assumed to arise independently from normally distributed populations having distinct means, denoted as μ_1, μ_2, and a common standard deviation, denoted as σ.

Unlike t-test, the null hypothesis of TOST is that the two population means are not equivalent and the alternative hypothesis is that two means are equivalent. The test is conducted by performing two separate one sided t tests on the difference of two population means with the predetermined limit E in both directions.

The two null hypotheses H_0 and corresponding alternative hypotheses H_a for the two t tests are set up as follows:

	Hypothesis on left hand side	Hypothesis on right hand side
Null Hypothesis	$H_{01} : \mu_2 - \mu_1 \leq -E$	$H_{02} : \mu_2 - \mu_1 \geq E$
Alternative Hypothesis	$H_{a1} : \mu_2 - \mu_1 > -E$	$H_{a1} : \mu_2 - \mu_1 < E$

If both sides of the null hypotheses are rejected, it is asserted that $-E < \mu_2 - \mu_1 < E$ and the two population means are said to be equivalent; otherwise, the two are deemed non-equivalent.

2.3.2 Equivalence Limit

The value E, termed equivalence limit, is selected as the worst-case difference between the two observed means from two populations with equivalent population means. It should be predetermined for each parameter under test based on scientific knowledge and potential risk. Higher risk should allow only small practical differences, thus a tight equivalence interval (-E,E) is required. Conversely, lower risks should allow larger practical differences which leads to a wider interval (-E,E). Scientific knowledge, product experience and clinical relevance should be evaluated when justifying the risk.

2.3.3 Testing Procedure

The two sample TOST procedure is carried out using the data sampled from the two populations. Let $\overline{X}_1$ and S_1, and $\overline{X}_2$ and S_2 denote the means and the standard deviations from sample 1 and sample 2, respectively. Note that for the assumption of common variance to be valid, the S values of the two data sets should be similar. The equivalence limit E, the confidence level $1-\alpha$, and sample sizes are determined based upon power and producer's risk.

Let the mean difference $D = \overline{X}_2 - \overline{X}_1$, and its standard error $S_D = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$, where the pooled standard deviation $S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}$, with the degrees of freedom $f = (n_1+n_2-2)$. An appropriate variance test, such as Levenes, should be used to further evaluate significant differences in measurement precisions. If the sample size from two populations are equivalent, i.e., $n_1 = n_2 = n$, then the standard error $S_D = S_p\sqrt{\frac{2}{n}}$, where $S_p = \sqrt{\frac{S_1^2+S_2^2}{2}}$. There are two operationally identical methods to test the equivalence. One method is to test the joint hypotheses that the mean difference is not as large as the upper value of a specified range and not below the lower bound of the specified range of equivalence. The t-statistics are $t_1 = \frac{E+D}{S_D}$ and $t_2 = \frac{E-D}{S_D}$,

for hypotheses on left-hand side and right-hand side, respectively. Both null hypotheses are rejected when $t_1 > t_{1-\alpha,f}$ and $t_2 > t_{1-\alpha,f}$, where $t_{1-\alpha,f}$ is the upper $(1-\alpha)th$ quantile of the Students t-distribution with f degrees of freedom. It is equivalent to use the p-values calculated by t statistics under the null hypotheses. The null hypotheses are both rejected if the p-values are less than α on both sides.

If both hypotheses are rejected, it is asserted that $-E < \mu_2 - \mu_1 < E$ and the two population means are said to be equivalent; otherwise, the two are deemed non-equivalent.

Another method, which is easier to apply and understand, is to construct a two- sided $100(1-2\alpha)\%$ confidence interval for the difference between two means and compare it with the equivalence limit. The reason the confidence interval is $100(1-2\alpha)\%$ and not the usual $100(1-\alpha)\%$ is because this method is tantamount to performing two one-sided tests. For instance, using a 90% confidence interval yields a 0.05 significance level for testing equivalence.

The upper (UCL) and lower (LCL) confidence limits for the $100(1-2\alpha)\%$ two-sided confidence interval on the true difference are computed as follows:

$UCL = D + t_{1-\alpha,f} S_D$

$LCL = D - t_{1-\alpha,f} S_D$

where $t_{1-\alpha,f}$ is the upper $100(1-\alpha)$ % percentile of the Students t distribution with the degrees of freedom ($f = n_1 + n_2 - 2$). If the confidence interval is completely contained within the equivalence limits (-E,E), the equivalence is accepted, if not rejected.

2.3.4 Lin's concordance corelation coefficient(CCC)

Lins concordance correlation coefficient (CCC) is the concordance between a new test or measurement (Y) and a gold standard test or measurement (X). This statistic quantifies the agreement between these two measures of the same variable

(e.g. chemical concentration). Like a correlation, CCC ranges from -1 to 1, with perfect agreement at 1. It cannot exceed the absolute value of ρ, which is the Pearsons correlation coefficient between Y and X. It can be legitimately calculated on as few as ten observations. The coefficient and associated sample size formulas are presented in [31, 32, 33].

According to Lin et al.[33], n observations (Y_K,X_K) are selected from a bivariate population with means μ_Y and μ_X, variances σ_Y^2 and σ_X^2, and correlation ρ (the Pearson correlation coefficient). Here, Y represents a measure from a receiving lab and X represents the sending lab. The degree of concordance between the two measures can be characterized by the expected value of their squared difference:

$$E[(Y-X)^2] = (\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_Y\sigma_X \tag{8}$$

The concordance correlation coefficient (CCC) is defined as:

$$\begin{aligned} CCC &= 1 - \frac{E[(Y-X)^2}{E[(Y-X)^2|\rho=0} \\ &= 1 - \frac{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_X\sigma_Y}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2} \\ &= \frac{2\rho\sigma_Y\sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2} \\ &= \rho \left(\frac{2}{\frac{(\mu_Y - \mu_X)^2}{\sigma_Y\sigma_X} + \frac{\sigma_Y}{\sigma_X} + \frac{\sigma_X}{\sigma_Y}} \right) \end{aligned}$$

2.4 Literature Review

The impact of input variables to the final substrate concentration during the manufacturing process for Pencilin has been studied using time-varying and multivariate statistical analysis. Nomikos and MacGregor [34] have developed a mul-

tivariate statistical process control (MSPC) approach for monitoring batch processes, based on multiway principal component analysis (MPCA). Partial Least Squares(PLS) methodology can be applied where product quality data is available. PLS is a linear regression tool and therefore its application to non-linear, time-varying processes, such as fed-batch fermentation systems is limited. To overcome this problem, Nomikos and MacGregor [35] developed a modified PLS approach, termed multi-way PLS. MPLS uses a technique referred to as unfolding to re-arrange the cause data collected from the batch, which can be considered to be a 3-dimensional array of size $m_x \times n_x \times n_b$, where m_x is the number of samples taken during a batch, n_x is the number of cause variables that are measured and n_b is the number of batches for which data is available. This 3-dimensional array is unfolded into a 2-dimensional array, of size $n_b \times (n_x * m_x)$.

Following the approach suggested, several techniques using multivariate statistical analysis have been developed and applied to industrial batch monitoring [36, 37, 38, 39, 40, 41]. Lee et al.[2] presented a new on-line monitoring method for a penicillin cultivation process. Traditional MPCA has some limitations and application for on-line monitoring included estimating future values to the end of the batch for each batch. This will result in poor monitoring performance, since many measurements are unknown at initial stages. The approach proposed by Lee et al. [2] includes combining batch-wise unfolding and variable-wise unfolding that will solve the problem while still preserving the dynamic relations. Zhang et al. [42] demonstrated the benefits PLS approach offers for fed-batch fermentation process. The paper demonstrated that multi-way PLS can provide accurate inference of quality variables, that are often difficult to measure using on-line sensors. It also provided insight in to how PLS can be used to provide early fault detection mechanism and isolation of fault conditions within the fermenter.

Symbolic regression (SR) is utilized to find mathematical expressions of functions that can fit the given information in view of accuracy, simplicity, and generalization. Unlike linear or nonlinear regression that proficiently advance parameters in the prespecified model, Symbolic Regression tries to look for proper models and their parameters simultaneously for a purpose of getting better insights into the dataset. With no earlier learning of material science, kinematics, and geometry, some natural laws described by mathematical expressions, such as Hamiltonians, Lagrangians, and other laws of geometric and momentum conservation, can be distilled from experimental data by the Genetic Programming (GP) method on SR as described by Schmidt and Michael [43]. Symbolic regression based models are well suited for fermentation process for multiple reasons. Similar to linear and multiple linear regression models, SR models derive a quantitative relationship between input and output variables. The major advantage of SR models is their ability to deviate from standard assumptions like normality, collinearity and non-linearity and still establish model accuracy and predictability. Symbolic regression has been extensively used in different areas including financial services, aerospace, manufacturing etc. Based upon the literature review symbolic regression has not been applied to a fed batch fermentation process and will be studied extensively as part of the proposed study.

Partial Least Squares (PLS) is the most used regression technique for monitoring and prediction of Quality variables, where as application of other classification and regression techniques like Support Vector machines (SVM) are hardly known. SVMs can be very useful based upon ability to handle non-linear, global solutions and its ability to optimize high dimensional input vectors. PLS models are used often for chemometrics due to its simplicity to use and accessibility, but PLS has limitations to handle non-linear relations. Centner et al. [44] researched multiple

regression techniques and compared the application to chemometrics. Neural networks were used as an alternative to handle non-linear relations, but SVM offers more functionality to deliver a global model that is capable of dealing with high dimensional vectors. Thissen et al. compared multivariate calibration by PLS and SVR on both the high-resolution and the low-resolution Raman spectra [45]. The study concluded that application of SVR is very advantageous for spectral data, because it can perform nonlinear regression efficiently for high dimensional data sets. Furthermore, its solution is global. For this reason, SVR and PLS were compared for two spectral data sets. In the first data set, it tested if low-resolution Raman spectra could be used for the prediction of the monomer masses during a reaction. In the second data set, the NIR spectra were affected by nonlinear temperature-induced variation. For both cases, it was shown that the SVR clearly outperforms PLS in both linear and nonlinear regression. Yi and Haiqing [46] adopted least squares support vector machines (LSSVM) modeling for a penicillin fed-batch fermentation, which improved the calculated speed. However, the problem of this method is that the pre-estimation result is not ideal in the initial stages of the fermentation process. Xianfang Wang et al. [47] further studied application of Least squares support vector machines to fermentation process that include hybrid modeling. The hybrid model included modeling method by utilizing advantages of LSSVM model and kinetics model of the fermentation process to improve the limitations of single model.

David LeBlondi et al. [48] explained the application of Analysis of Covariance (ANCOVA) model to establish shelf life for pharmaceutical products. The paper illustrated different models (common slope, common intercept etc,) and how to assess model adequacy using measures such as root mean square error (RMSE), lack of fit and predicted R-square. Based upon the model selected, confidence

intervals and prediction intervals can be calculated to determine shelf life. Multiple methods to test for comparing regression coefficients were proposed in the literature to determine poolability of batches. Ruberg and Stegeman [49] and Ruberg and Hsu [50] discuss methods based on multiple comparison. Ruberg and Hsu [50] assumed a common intercept and developed a method for pooling batches based on simultaneous confidence intervals for the slopes, as compared with the worst slope.

Liu et al.[51] developed an approach based upon confidence bands. Monte-carlo simulation was used to determine the adjusted critical point in the study. Modification of this approach was further studied by Liu et al.[52] using constant-width simultaneous confidence intervals where critical point is computed from a multivariate t-distribution. Yoshioka et al.[53] and Tsong et al.[54] introduced a procedure based on the equivalence of shelf lives at a given acceptance criteria of label claim. The concept of batch equivalence was provided by Tsong et al.[54] based on mean assay content at a given target expiration date. Djira et al. [55] further provided guidance on pooling batches in stability studies. The paper explained a modification to Tsong's approach, where a delta method will approximate standard errors of the estimated shelf life. The approach proposed by Djira indicates that pooling batches at a proposed shelf life does not requires assumptions regarding equality of slopes or intercepts. Knezevic [56] explained the WHO recommendations for Vaccines stability. Later in 2006, WHO published "Guidelines for stability evaluation of vaccines" for stability testing of individual vaccines, as part of WHO Technical Report Series [57], and provided a set of general principles and a description of their application. The document provides guidance for general considerations on stability evaluations, and stability of vaccines during the manufacturing process and final product. Multiple options were recommended

to manufacturers for stability design and data analysis. One is the method based upon compliance with shelf life determinations as the time point with last measurement within the specification. Another method where expiration date is defined through statistical evaluation and extrapolation of the data. Schofield [58] provided guidance on long-term and accelerated studies. Accelerated studies not only support release specification, but also can be used to support evidence to recalculate product expiry when an unintended scenario happens. Schofield provided multiple scenarios with evaluation of shelf life using single lot, multiple lots, comparability of degradation rates, and establishment of shelf life from worst-case lot. Komka et al. presented modification of traditional approach for shelf life determination using data from long-term studies [59]. ICH guideline does not include interval-to-interval variability and reproducibility effect is not estimated separately from other variance components. Norwood [60] incorporated date of analysis in the stability model, but parameterization was not adequate. The model developed by Komka [59] improves Norwoods model applying a crossed design between the batch and date of analysis instead of nested error. The model proposed that storage time belongs to degradation rate and date scale combines interval -to-interval reproducibility. The choice of the model affects the slope of the degradation line; thus, the shelf life of the drug product.

Assays that measure the quality of the product need to be transferred from development/validation lab to production lab and sometimes the assays have to be transferred to contract manufacturing labs. The objective of the method transfer is to ensure that results obtained by receiving lab are reliable and comparable to sending laboratory to meet the specifications of a product and address any bias between the labs. Regulatory agencies, such as the US Food and Drug Administration (FDA), requires that analytical method transfers to be documented to

ensure receiver is qualified to perform the methods on a routine basis. Rozet et al. [7] provided recommendations on evaluation of method transfers, statistical evaluation and experimental design of transfer. The paper only provides general principles for transfer of certain assays and proposes a minimum sample size. No recommendations on how to decide about the acceptability of the transfer or on how to select an optimal sample size are provided. PDA Technical report [61] provides guidance on analytical method transfers. Acceptance criteria should be established and justified for the allowed difference(s) between the originating and receiving laboratories prior to the transfer. Acceptable differences between laboratories for the method performance characteristics of quantitative methods such as Accuracy and Intermediate Precision should be estimated based on historical data and/or previous AMV protocols/reports with respect to the specifications. Don Chambers et.al [62]recommends establishing analytical method equivalency, the choice of samples, acceptance criteria, data evaluation, and considering documentation. C. Agut et al.[63] proposed an updated transfer methodology. The proposal was based on the principles of risk management, this updated methodology brings a panel of strategies custom-tailored to the context from prior scientific information. The paper recommends a panel of approaches based upon level of incurred risks, depending on the compound, the category of method, the criticality of related attributes and the type of the transfer. The selection of an approach should always be based on scientific knowledge and ultimately linked to the protection of the patient.

Schuirmanns two one-sided test (TOST) [28] is used to conduct the equivalence testing on numerical data from two sources. As an alternative, Lawrence and Lin proposed methodology using concordance correlation coefficient (CCC) [32], which was originally proposed [31] to assess the reproducibility of an assay

compared to the gold standard assay. Bland and Altman [64] proposed that the Pearson correlation coefficient is not a measurement of agreement because it depends on the range of the true quantity in the sample set. Lin proposed total deviation index (TDI) to access the control of individual differences [33]. TDIπ is a product of the square root of mean squared error and a percentile of normal distribution. TDIπ approximates boundary that 100π percent of the individual differences are within the boundary. If the 95% upper confidence bound of TDIπ is no more than the predefined boundary Δ, then the assay can be transferred. Zhong and Shao [65, 66] proposed an approach that combined the mean and variances of individual difference into one statistics. Zhong et al. [67] proposed a tolerance interval approach for method transfers. The proposed two one-sided tolerance limits approach provides the same limits to the proportion below the lower specification. All the literature suggests multiple methods to establish equivalence for assay transfers, but no recommendations on how to decide about the acceptability of the transfer or on how to select an optimal sample size are provided. The proposed study will provide more details for optimal sample size based upon pre-established criteria and Producer/Consumer risks.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter details out the research methodology for the present study. The research objectives and applicable methodology to achieve those objectives as mentioned in Chapter 1 are explained here. This involved an exhaustive study of the statistical methodologies applied for biopharmaceutical manufacturing process, assay transfers and stability studies used for shelf-life determination. In addition, the study also assessed the impact of risk factors for implementation. This was followed by model validation through cross-validation.

The research methodology must be robust to minimize errors in data collection and analysis. Owing to this, simulations and historical data were chosen for data collection. This chapter describes the pilot study, data collection, and data analysis procedures of the entire study.

3.2 Research Process

The aim of the current research is to evaluate different statistical methodologies and recommend the appropriate methodology for optimizing a biopharmaceutical manufacturing process, establishing shelf life and assay transfers. The first study evaluates the manufacturing process and application of multivariate techniques to understand the process variation and predict the outputs of the process. The second study evaluates shelf life of finished products using existing guidance and potential statistical methodologies for shelf life extension. The third study evaluates establishing the acceptance criteria and sample size required for successful method transfer optimizing patient and producers risk.

3.2.1 Data Collection

The studies are aimed at optimizing biopharmaceutical manufacturing processes, stability studies and assay transfers. The data for manufacturing process is collected from a process simulator Pensim, developed by a research group in Illinois Institute of Technology, USA [68]. Pensim is a modular simulation package for fed-batch fermentation: penicillin production. The simulator is structured such that one may use different modules depending on the application needs of the user. Pensim is a simulator originally developed in MATLAB 6.0 and modularized functions were compiled into ANSI C codes and these C modules are further compiled to produce an executable stand-alone application file (A modular simulation package for fed-batch fermentation: penicillin production).

The simulation software has been used and cited in more than hundred research articles and represents the manufacturing process. The mechanistic model of Bajpai and Reuss [69] was utilized as the basis of modeling efforts in the software.The effects of environmental variables such as pH and temperature, and input variables such as aeration rate, agitation power, feed flow rate of substrate on biomass formation have been included in the model for completeness. The original model has been extended by including additional input variables such as agitation power and aeration rate. The functional relationships among the process variables are summarized below:

- X=f(X, S, CL, H, T)
- S=f(X, S, CL, H, T)
- CL=f(X, S, CL, H, T)
- P=f(X, S, CL, H, T, P)
- CO_2=f(X, H, T)

- H=f(X, H, T)

where X is the biomass concentration

S is substrate concentration

CL is the dissolved oxygen concentration

P is the penicillin concentration

CO_2 is the carbon dioxide concentration

H is the hydrogen ion concentration for pH ($[H+]$) and

T is the temperature.

All the inputs and outputs are listed in Figure 4.

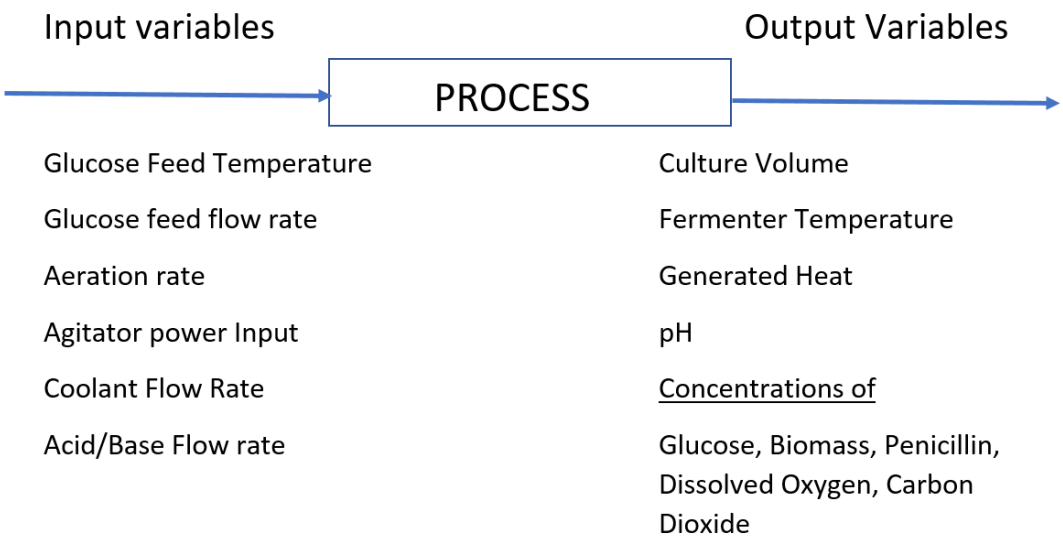


Figure 4: Process input/output structure

Two PID (Proportional Integral Derivative) controllers are used to manipulate the acid and base control values. The PID controllers were tuned for a certain range of initial conditions considered to be the normal operation as seen in Table 1.

	Value
Initial conditions	
Substrate concentration: S (g/l)	15
Dissolved oxygen concentration: CL(=CL^* at saturation)(g/l)	1.16
Biomass concentration: X (g/l)	0.1
Penicillin concentration: P (g/l)	0
Culture volume: V(l)	100
Carbon dioxide concentration: CO_2(mmol/l)	0.5
Hydrogen ion concentration:$[H^+]$(mol/l)	$10^{5.1}$
Temperature: T(K)	297
Heat generation: Q_{rxn}(cal)	0
Kinetic parameters and variables	
Feed substrate concentration: s_f(g/l)	600
Feed flow rate of substrate: F (l/h)	
Feed temperature of substrate: T_f (K)	298
Yield constant: $Y_{x/s}$ (g biomass/g glucose)	0.45
Yield constant: $Y_{x/o}$ (g biomass/g oxygen)	0.04
Yield constant: $Y_{p/s}$ (g penicillin/g glucose)	0.90
Yield constant: $Y_{p/o}$ (g penicillin/g oxygen)	0.20
Constant: K_1 (mol/l)	10^{-10}
Constant: K_2 (mol/l)	7×10^{-5}
Maintenance coefficient on substrate: m_x (per h)	0.014
Maintenance coefficient on oxygen: m_o (per h)	0.467
Constant relating CO_2 to growth: α_1 (mmol CO_2/g biomass)	0.143
Constant relating CO_2 to maintenance energy: α_2 (mmol CO_2/g biomass h)	4×10^{-7}
Constant relating CO_2 to penicillin production α_3 (mmol CO_2/l h)	10^{-4}
Maximum specific growth rate: μ_x (per h)	0.092
Contois saturation constant: K_x (g/l)	0.15
Oxygen limitation constant: K_{ox}, K_{op} (no limitation)	0

Oxygen limitation constant: K_{ox}, K_{op} (with limitation)	2×10^{-2}, 5×10^{-4}
Specific rate of penicillin production: μ_p (per h)	0.005
Inhibition constant: K_p (g/l)	0.0002
Inhibition constant for product formation: K_I (g/l)	0.10
Constant: p	3
Penicillin hydrolysis rate constant: K (per h)	0.04
Arrhenius constant for growth: k_g	7×10^3
Activation energy for growth: E_g (cal/mol)	5100
Arrhenius constant for cell death: k_d	10^{33}
Activation energy for cell death: E_d (cal/mol)	50000
Density $\times$ heat capacity of medium: ρC_p (per l $^\circ$ C)	1/1500
Density $\times$ heat capacity of cooling liquid: $\rho_c C_{pc}$ (per l $^\circ$ C)	1/2000
Yield of heat generation: r_{q_1} (cal/g biomass)	60
Constant in heat generation: r_{q_2} (cal/g biomass h)	1.6783×10^{-4}
Heat transfer coefficient of cooling/heating liquid: a (cal/h $^\circ$C)	1000
Cooling water flow rate: F_c (l/h)	0
Constant: b	0.60
Constants in K_{la}: α, β	70, 0.4
Constant in F_{loss}: λ (per h)	2.5×10^{-4}
Proportionality constant: γ (mol $[H^+]$/g biomass)	10^{-5}
Controller parameters (PID)	
pH: (base) K_c, τ_I: (h), τ_d: (h)	8×10^{-4} , 4.2,0.2625
pH: (acid) K_c, τ_I: (h), τ_d: (h)	1×10^{-4}, 8.4,0.125
Temperature: (cooling) K_c, τ_I: (h), τ_d: (h)	70, 0.5, 1.6
Temperature: (heating) K_c, τ_I: (h), τ_d: (h)	5, 0.8, 0.05

Table 1: Initial conditions, kinetic and controller parameters for nominal operation

The maximum allowable acid and base additions were set to 0.01 and 100 ml/h, respectively. The initial pH is chosen arbitrarily as 5.0 for this particular case. Due to the acid addition rate limitation, the pH reached its desired value in almost 8 h of operation, then stayed within its control limits. A set point gap of 0.05 that can be adjusted by the users prior to simulation is defined for acid flow rate controller to avoid excessive acid additions. Acid solution is only added if the pH exceeds its set point value by 0.05.

The flowsheet of the penicillin process is illustrated in Figure 5.

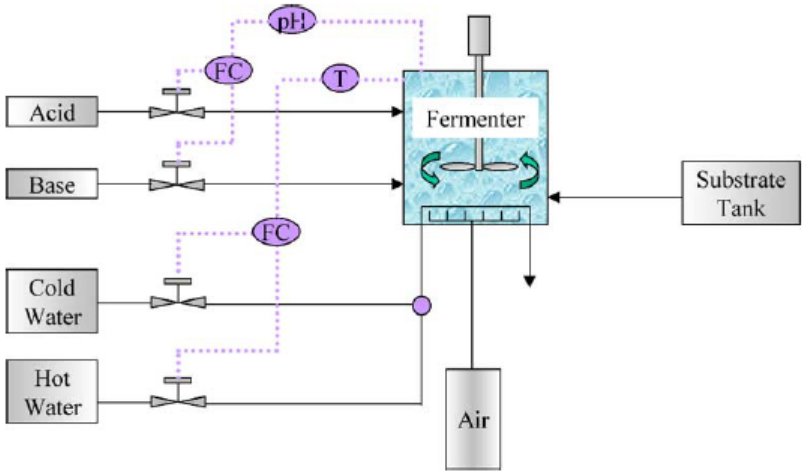


Figure 5: Flow sheet of Penicillin cultivation process [68]

Simulations are run under closed-loop control of pH and temperature while controlling glucose addition. In bioprocesses, most of the important process variables such as biomass and penicillin concentrations are analyzed off-line by the quality control laboratory resulting in a lag in process measurements.In a first phase, the fermenter is operated in batch mode. Once the substrate concentration drops below the threshold value of 0.3 g/L (after approximately 42 h of operation),

the feed flow is initiated. The fermentation is terminated after 400 h once 25 L of substrate is added. Based on the batch recipe, two main phases are identified: a batch phase of about 44 h and a fed-batch phase with a length of roughly 356 h. During the fermentation, 11 sensors record the various concentrations, flows, temperatures, and pH. Gaussian noise with the indicated standard deviation is added to each sensor to represent measurement noise. The measured signals are aligned and resampled to a length of 800 samples via indicator variables, identical to the procedure employed by Undey et al. [70]. The first 89 samples correspond to the batch phase of the fermentation, and the final 711 belong to the fedbatch phase. The transformed time signal is added as an aligned variable. The quality parameter is the penicillin concentration after batch completion [71]. A total of 150 NOC (i.e., fault-free) batches are simulated and used as historical training data. For each batch, the initial substrate concentration, biomass concentration, and culture volume are subject to small random variations to represent raw material variability. This results in a data matrix X-variables of size $150 \times 1 \times 800$ with corresponding quality matrix Y-variables of size 150×1. To assess the influence of each variable on the final quality, the set points of the five manipulated variables are adjusted after 150 h of operation in each batch.The set point changes are small to simulate actual test runs on a real installation.

The dataset collected based upon the simulation is huge and consists of 12000 rows and 17 columns. A subset of the complete dataset is included below in Figure 6.

lot	time stamp (h)	Aeration rate(fg2)	Agitator Power(pw2)	Subs. Feed Rate(f×2)	Subs. Feed Temp.(T2)	Substrate . Concentration	DO concentration	Biomass concentration	Penicillin concentration	Culture Volume	CO2 concn	pH	Temperat	Generated	flo: Acid flow rate	flo: Base flow rate	flo: Cooling/Heating water flow rate
Lot1	0	8.59886	29.93063	0	295.98969	15	1.1592	0.1	0	100	0.52995	5	298	0	0	0	0
Lot1	0.5	8.59913	29.94125	0	295.99437	14.98952	1.15715	0.10523	0	99.98755	0.46532	4.99773	298.0001	0.01138	0	0	0.07787
Lot1	1	8.59994	29.94938	0	296.00831	14.97841	1.15437	0.11073	0	99.9751	0.4895	4.99536	297.9978	0.06437	0	0	0.09162
Lot1	1.5	8.6085	29.9575	0	296.085	14.96663	1.15726	0.11651	0.00001	99.96266	0.52399	4.99456	297.9953	0.09905	0	0	0.0001
Lot1	2	8.60119	29.96688	0	296.08986	14.95414	1.15855	0.12259	0.00001	99.95022	0.49419	4.99369	297.9957	0.12553	0	0	0.0001
Lot1	2.5	8.60175	29.975	0	296.0125	14.9489	1.15554	0.12899	0.00001	99.93778	0.5549	4.99339	297.9962	0.1259	0	0	0.0001
Lot1	3	8.60256	29.98188	0	296.01594	14.93687	1.15697	0.13572	0.00001	99.92534	0.45258	4.99337	297.9968	0.21425	0	0.00001	0.0001
Lot1	3.5	8.60313	29.98625	0	296.01875	14.91201	1.15716	0.1428	0.00001	99.9129	0.55276	4.99359	297.9976	0.25671	0	0.00001	0.0001
Lot1	4	8.60394	29.99063	0	296.02094	14.89629	1.15648	0.15025	0.00002	99.90047	0.41471	4.994	297.9986	0.30137	0	0.00001	0.0001
Lot1	4.5	8.60488	29.9925	0	296.02575	14.87964	1.15789	0.15809	0.00002	99.88803	0.46286	4.99444	297.9996	0.34035	0	0.00001	0.0001
Lot1	5	8.60569	29.99188	0	296.02656	14.86202	1.15765	0.16634	0.00002	99.8756	0.4985	4.99492	298.0009	0.39729	0	0.00001	0.00104
Lot1	5.5	8.60625	29.98875	0	296.02937	14.84339	1.15506	0.17502	0.00002	99.86316	0.52224	4.99531	298.0003	0.44979	0	0.00001	0.17945
Lot1	6	8.60669	29.98688	0	296.03156	14.82368	1.15788	0.18415	0.00002	99.85074	0.4732	4.99559	297.9971	0.50449	0	0.00001	0.0001
Lot1	6.5	8.60725	29.985	0	296.03375	14.80286	1.15624	0.19375	0.00003	99.83831	0.54028	4.99577	297.9989	0.56203	0	0.00001	0.0001
Lot1	7	8.60756	29.98438	0	296.03656	14.78085	1.15822	0.20386	0.00003	99.82588	0.57639	4.99586	298.0009	0.62256	0	0.00001	0.13566
Lot1	7.5	8.60775	29.98625	0	296.03812	14.7576	1.15614	0.21449	0.00003	99.81346	0.5618	4.99587	297.9954	0.68623	0	0.00001	0.13652
Lot1	8	8.60794	29.98938	0	296.03969	14.73304	1.15584	0.22567	0.00004	99.80104	0.52388	4.99579	297.994	0.75318	0	0.00001	0.0001
Lot1	8.5	8.60788	29.99	0	296.04062	14.70712	1.15625	0.23743	0.00004	99.78862	0.47512	4.99566	297.9966	0.82259	0	0.00001	0.0001
Lot1	9	8.60781	29.99188	0	296.04094	14.67975	1.15295	0.2498	0.00004	99.7762	0.53488	4.99545	297.9995	0.89765	0	0.00001	0.08758
Lot1	9.5	8.60763	29.995	0	296.04	14.65086	1.15359	0.26281	0.00005	99.76378	0.52577	4.9952	298.0023	0.97554	0	0.00001	0.0001
Lot1	10	8.60744	29.99938	0	296.03844	14.62038	1.15331	0.27649	0.00005	99.75136	0.44813	4.99487	298.0057	1.05745	0	0.00001	0.25817
Lot1	10.5	8.60725	30.005	0	296.03687	14.58823	1.15312	0.29089	0.00005	99.73894	0.5193	4.99448	298.0031	1.14359	0	0.00001	0.66325
Lot1	11	8.60719	30.01063	0	296.03406	14.55432	1.15491	0.30687	0.00006	99.72653	0.58353	4.99402	297.9939	1.23415	0	0.00001	0.0001
Lot1	11.5	8.60713	30.015	0	296.03187	14.51857	1.1489	0.32193	0.00006	99.71412	0.5462	4.99349	297.9982	1.32935	0	0.00001	0.0001
Lot1	12	8.60694	30.01813	0	296.02906	14.48887	1.15139	0.33866	0.00007	99.70171	0.54835	4.99294	298.0028	1.42945	0	0.00001	1.00218
Lot1	12.5	8.60675	30.02125	0	296.02562	14.44114	1.15212	0.35625	0.00007	99.6893	0.54486	4.99236	297.9899	1.53468	0	0.00001	0.44011
Lot1	13	8.60656	30.02563	0	296.02094	14.39328	1.15144	0.37474	0.00008	99.67691	0.44692	4.99178	297.9856	1.64527	0	0.00001	0.0001
Lot1	13.5	8.60625	30.03	0	296.01562	14.35517	1.14831	0.39418	0.00008	99.66451	0.552	4.9912	297.9913	1.76152	0	0.00001	0.0001
Lot1	14	8.60594	30.03438	0	296.01031	14.3087	1.14976	0.41462	0.00009	99.65212	0.59389	4.99065	297.9974	1.88373	0	0.00001	0.14994
Lot1	14.5	8.60538	30.03575	0	296.00563	14.25934	1.15111	0.43611	0.00009	99.63972	0.58729	4.99012	297.9998	2.01221	0	0.00001	0.02881
Lot1	15	8.60456	30.03938	0	296.00094	14.20817	1.15085	0.4587	0.0001	99.62731	0.5476	4.98959	298.0057	2.14726	0	0.00001	0.0001
Lot1	15.5	8.60363	30.04	0	295.99687	14.15386	1.15115	0.48245	0.0001	99.61491	0.5586	4.98906	298.0131	2.28922	0	0.00002	1.16119
Lot1	16	8.60294	30.03938	0	295.99406	14.09667	1.14943	0.50741	0.00011	99.60251	0.57927	4.98851	298.0011	2.43841	0	0.00002	1.79828
Lot1	16.5	8.60238	30.04	0	295.99125	14.0365	1.14891	0.53363	0.00012	99.59012	0.49652	4.98794	297.9822	2.59511	0	0.00002	0.0001
Lot1	17	8.60156	30.03938	0	295.98719	13.97317	1.14721	0.56118	0.00012	99.57774	0.5216	4.98729	297.9912	2.7594	0	0.00002	0.0001
Lot1	17.5	8.601	30.04	0	295.98312	13.90651	1.14787	0.59014	0.00013	99.56536	0.57098	4.98661	298.0007	2.93275	0	0.00002	1.81886
Lot1	18	8.60056	30.03813	0	295.97906	13.83638	1.14543	0.62057	0.00014	99.55298	0.55789	4.98585	297.9833	3.1145	0	0.00002	0.625
Lot1	18.5	8.60038	30.03875	0	295.97375	13.76261	1.14663	0.65252	0.00015	99.54061	0.53298	4.98511	297.9815	3.30536	0	0.00002	0.0001
Lot1	19	8.60006	30.03813	0	295.96719	13.68502	1.14469	0.68609	0.00016	99.52823	0.68676	4.98438	297.9929	3.50587	0	0.00002	0.0001
Lot1	19.5	8.59963	30.035	0	295.96187	13.60338	1.14334	0.72137	0.00017	99.51585	0.50142	4.98368	298.005	3.71653	0	0.00002	0.97208
Lot1	20	8.599	30.03	0	295.9575	13.51752	1.14577	0.75843	0.00018	99.50347	0.59999	4.98297	298.0003	3.93179	0	0.00002	1.20684
Lot1	20.5	8.59838	30.02375	0	295.95313	13.42726	1.14288	0.79704	0.00019	99.4911	0.53185	4.98227	297.9934	4.1701	0	0.00003	0.0001
Lot1	21	8.598	30.01875	0	295.9475	13.33257	1.141	0.83821	0.0002	99.47872	0.6317	4.9816	298.0078	4.41406	0	0.00003	0.05328
Lot1	21.5	8.59775	30.015	0	295.94312	13.23258	1.13996	0.88114	0.00021	99.46634	0.59692	4.98094	298.0212	4.6703	0	0.00003	3.2152
Lot1	22	8.59763	30.01125	0	295.93812	13.12773	1.13882	0.92621	0.00022	99.45396	0.69392	4.98031	297.9961	4.93925	0	0.00003	2.77393
Lot1	22.5	8.59775	30.00875	0	295.93375	13.01761	1.13624	0.97349	0.00024	99.44161	0.59888	4.97972	297.9761	5.22139	0	0.00003	0.0001
Lot1	23	8.59775	30.00625	0	295.93	12.90193	1.13678	1.02312	0.00025	99.42925	0.56622	4.97914	297.9941	5.51749	0	0.00003	0.0001
Lot1	23.5	8.59788	30.005	0	295.92625	12.78835	1.13504	1.07524	0.00027	99.41689	0.66585	4.97855	298.0131	5.82838	0	0.00004	3.3154
Lot1	24	8.59775	30.00375	0	295.92312	12.65267	1.13536	1.12992	0.00028	99.40453	0.74325	4.97796	297.9911	6.15457	0	0.00004	2.30747

Figure 6: Subset of the data used

After the manufacturing of therapeutic proteins, shelf life or the stability time period need to be established. The purpose of stability testing is to provide an understanding on how the quality of an active substance or finished product varies with time under the influence of a variety of environmental factors such as temperature, humidity, and light, and to establish a re-test period for the active substance or a shelf life for the finished product and recommended storage conditions. Ruberg and Stegeman [49] and Ruberg and Hsu [50] discuss methods based on multiple comparison for poolability of slopes and batch degradation. The first stability data is taken from Ruberg and Stegeman [49] and is shown in Figure 7.

Batch	Time(years)	Concentration (%)
1	0.014	100.4
1	0.28	100.3
1	0.514	99.7
1	0.769	99.2
1	1.074	98.9
1	1.533	98.2
1	2.03	97.3
1	3.071	95.7
1	4.049	94.5
2	0.022	100.7
2	0.118	100.6
2	0.272	100.3
2	0.566	99.9
2	1.165	98.6
2	2.022	97.6
2	3.077	96.4
3	0.025	100.2
3	0.275	99.7
3	0.547	99.2
3	0.797	99
3	1.041	98.8
3	2.058	96.4
3	2.519	96.2
4	0.066	100.4
4	0.343	100
4	0.533	99.5
4	0.802	99.3
4	1.033	99.3
4	1.538	98.2
5	0.011	100.5
5	0.31	99.8
5	0.624	99.1
5	1.063	98.4
6	0.011	100.1
6	0.31	99.5
6	0.624	98.5
6	1.063	98.4

Figure 7: Data used for stability analysis

There are six batches in this data and concentration for the batches is mea-

sured over years.

Assays that measure the quality of the product need to be transferred from development/validation lab to production lab and sometimes the assays have to be transferred to contract manufacturing labs.The purpose of the transfer is to demonstrate the two labs are comparable to avoid re-validation of the assay.Rick Lung et al. [72] discussed statistical methods for the determination of equivalence of automated test procedures. Dewe et al.[73] used total error as a decision criterion in analytical method transfer.Assay transfers include a protocol including details with scope, experimental design and statistical analysis, assay transfer criteria and conclusions from the transfer.Transfer criteria includes comparison of accuracy, precision and equivalency ranges. Data was collected for an assay transfer using the methodology in Table 2.

As seen in the table, two analysts from site A will each perform the assay nine times for the transfer samples on nine separate occasions. Two qualified analysts from Site B will each perform the assay nine times for the transfer samples on nine separate occasions.

Site	**Analyst**	**Transfer Sample**	**Subtotal**	**Total**
A	Analyst 1	9	9	18
	Analyst 2	9	9	
B	Analyst 1	9	12	18
	Analyst 2	9	9	
Subtotal		36	Total	**36**

Table 2: Methodology for data transfer

The same lots of control and transfer sample (eighteen assays for each site) will be used at both sites. Components for reagents used in the performance of the assays at each site can be from the same or different lots, but they must be stored at the testing site prior to use.

3.2.2 Data Analysis Procedure

SAS, JMP, Eureqa, SIMCA-P, Matlab software were used for statistical analyses. The data was collected as discussed in 3.2.1 and tabulated in Microsoft Excel. Required analysis was done using the statistical packages mentioned above. The statistical procedures used for the analysis of the research objectives of this study are described below.

Analysis of Bio-pharma manufacturing process data

Data for an industrial scale fed batch process was generated using Pensim Simulator. The data collected was analyzed using four different statistical methodologies : Partial Least Squares (PLS), Principal Component Analysis (PCA), Symbolic Regression and Support Vector Machines (SVM). An extension of existing statistical methodology using latent vector scores and Symbolic Regression called Scores based Symbolic Regression(SBSR) is developed and compared to the other methodologies mentioned above.

Methodology for computing SBSR is described below:

- **Step 1: Using Latent Vector Scores**: In the case of Penicillin manufacturing process, suppose n variables have been measured at p time intervals. The information can be formulated in a matrix form as:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

  where $x_1 = [x_{11}x_{12}...x_{1p}]$ is the row vector containing the penicillin concentrations measured at p time intervals for the first lot, x_2 is the row vector containing the penicillin concentrations for the second sample and so on. PCA creates new orthogonal variables (latent variables) that are linear combinations of the original x-variables. Singular value decomposition of the

matrix can be used for X:

$$X_{nxp} = U_{nxp}\Lambda_{pxp}P'_{pxp} = T_{nxp}P'_{pxp}[74] \tag{9}$$

U is the unweighted/normalized score matrix and T is the weighted /unnormalized score matrix. The new matrices contain the new variables. P is the loading matrix and the column vectors are called eigen vectors. Elements of matrix P are the loadings of original variables on each eigenvector.

- **Step 2** : Apply Symbolic Regression for the latent scores obtained from Step 1

Root mean squared error (RMSE) was calculated for each methodology to compare the prediction ability and efficiency of each regression model. Validation set was constructed as one fifth of all samples and is representative of common cause variation. Relative accuracy of each calibration model is estimated by calculating the mean average error(MAE). Leave-one-out cross validation was used to optimize the models parameters based on the root mean squared error of cross-validation (RMSECV). Cross validation is a better evaluation method compared to residuals. One of the constraints with residuals is that they do not provide predictions well for data that the validation has not already seen. One way to overcome this problem is to not use the entire data set when training a learner.Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model on "new" data. This is the basic idea for a whole class of model evaluation methods called cross validation. Leave-one-out cross validation (LOOCV) is K-fold cross validation taken to its logical extreme, with K equal to N (the number of data points in the set). That means that N separate times, the function approximator is trained on all the data except for one point and a prediction is made for that point.

As discussed previously, the Mean average error is computed and used to evaluate the model. The evaluation given by leave-one-out cross validation error (LOO-XVE) is good, but at first pass it seems very expensive to compute. Fortunately, locally weighted learners can make LOO (Leave-one-out) predictions just as easily as they make regular predictions. That means computing the LOO-XVE takes no more time than computing the residual error and it is a much better way to evaluate models. SIMCA-P software used for regression by default leaves out 1/7th of the data, which is a more stringent test. In cases of low n ($<$ say 20) you may change the number of groups used for cross validation to equal the number of observations.In this way, the cross validation will be equivalent to "leave one out". For SVM based methods prediction error was calculated. RMSECV minimization was used for optimization in all cases and for all models.

Stability Data Analysis

The statistical model used most often for comparing k batches is the analysis of covariance (ANCOVA) given by

$$Y_{ij} = \mu + \tau_i + \beta_i X_{ij} + \varepsilon_{ij}, \tag{10}$$

where Y_{ij} is the jth response for the ith batch,

τ_i is the batch effects,

β_i is the degradation rate of the ith batch,

X_{ij} is the time of the stability sample corresponding to Y_{ij},

ε_{ij} is the random error corresponding to Y_{ij}. Here ε_{ij} are independent normally distributed random variables with mean zero and common variance σ^2. The hypothesis test of interest

$$\begin{aligned} H_0 : &\quad \beta_i = \beta_j \quad \text{for all}\, i, j \\ H_a : &\quad \beta_i \neq \beta_j \quad \text{for some}\, i \neq j. \end{aligned} \tag{11}$$

If H_0 cannot be rejected, then a reduced model is used in which a common slope ($\beta_i = \beta$ for all i) is estimated for all batches in Eq 10. If the condition in Eq 11 is rejected, then the FDA Guideline suggests that subsets of batches with similar slopes may be considered together and retested as above, although no specific methodology is proposed. Ultimately, shelf-lives are computed for each subgroup of batches, and the shortest shelf-life is used for the drug product.

The sidedness of the interval also impacts the determination of shelf life. ICHQ1E guideline suggests that for an attribute known to decrease with time, the lower one-sided 95% confidence limit should be compared to the acceptance criterion. For an attribute known to increase with time, the upper one-sided 95% confidence limit should be compared to the acceptance criterion. For an attribute that can either increase or decrease, or whose direction of change is not known, two-sided 95% confidence limits should be calculated and compared to the upper and lower acceptance criteria. This will result in longer shelf life for one-sided 95% bound on mean and shorter shelf life for two-sided 95% bound on mean irrespective of specifications.

The proposed methodology studied the impact of one-sided 95% bounds for one-sided specification and two-sided 95% for two-sided specifications using multiple scenarios. This methodology evaluated the terms “Lot” and “Lot*Time” as random effects and include random effects based on Akaikes Information Criterion (AIC) to determine the model of best fit. AIC is used when control of Type I error rate is most important and choosing a model that is too simple adversely affects Type I error rate. Full model, Random Model and fixed models were constructed to compare and a recommendation will be made based upon lowest AIC.

3.2.3 Proposed Approach for establishing shelf life

The model selection will be based upon model of best fit using Akaikes Information Criterion (AIC) [75] compared to selecting the model based upon significance level. The model with smallest AIC will be recommended as the best model. AIC is a measure of the relative quality of statistical models for a given set of data. Based upon the models for a dataset, AIC can provide a basis for model selection. Burnham and Anderson [76] proposed that, chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. It is defined as:

$AIC = -2(ln(likelihood)) + 2K$ where $likelihood$ is the probability of a given model fitting the data and K is the number of free parameters in the model.AIC scores are often shown as ΔAIC scores, or difference between the best model (smallest AIC) and each model (so the best model has a ΔAIC of zero).Guerin and Stroup [77] suggested that AIC can be used as model selection criteria when control of Type I error rate is most important.AICc(Corrected Akaikes Information Criterion) and Bayesian Information Criterion (BIC) results are also included in the analysis. The corrected Akaike's Information Criterion (AICc) and the Bayesian Information Criterion (BIC) are information-based criteria that assess model fit. Both are based on $-2LogLikelihood$.

AICc is defined as follows:

$$AICc = -2Log(Likelihood) + 2k + 2k(k+1)/(n-k-1) \tag{12}$$

where k is the number of estimated parameters in the model and n is the number of observations used in the model. This value can be used to compare various models for the same data set to determine the best-fitting model. The model having the smallest value, as discussed in [75], is usually the preferred model.

BIC is defined as follows:

$$BIC = -2Log(Likelihood) + kln(n) \tag{13}$$

where k is the number of estimated parameters in the model and n is the number of observations used in the model. When comparing the BIC values for two models, the model with the smaller BIC value is considered better. The following models will be used for comparison of AIC:

- Full Model
- Random Model
- Fixed Model

General Linear mixed models using SAS 9.2 will be used to build the models. Notation for full models include : $Y_{ij} = \mu + L_i + (\beta + B_i) \times t_{ij} + E_{ij}$, $i = 1, 2, ...n$ and $j = 1, 2, ...T_i$
where

- Y_{ij} is the response for lot i at time point j
- μ is the average y-intercept across all lots
- β is the average slope across all lots
- L_i is a random variable that allows the y-intercept to vary from μ for a given lot; L_i has a normal distribution with mean 0 and variance σ_L^2
- B_i is a random variable that allows the slope to vary from β for a given lot; B_i has a normal distribution with mean 0 and variance σ_B^2
- t_{ij} is the time point for measurement j of lot i

- E_{ij} is a random normal error term created by measurement error and model misspecification with mean 0 and variance σ_R^2
- n is the number of lots
- T_i is the number of responses obtained for lot i
- L_i, B_i and E_{ij} are jointly independent

Assay Transfers

Analytical transfer is a major and integrated step in the method life cycle, within the analytical life cycle management.More specifically, inter-laboratory studies conducted usually for transfer purpose enable to address a major component of the robustness item required by industry guidelines on validation (ICH Q2(R1) and FDA Guidance for Industry). The absence of an aligned regulation and guidance from agencies regarding establishing the acceptance criteria, establishing comparability and equivalence for method transfer leads to multiple approaches that differ in validity of results. Acceptance criteria for comparability/equivalence need to be established based upon the intended use of the method to be transferred. Method transfers for less complex/critical assays can be justified if the results are within the specification limits. Advanced statistical methods like statistical equivalence tests, students t-test and Two one-sided t-test(TOST)need to be applied for complex and critical assays. Equivalence tests provide a direct control of type II error (consumers risk).

The goal of assay transfer is to show the equivalence between the two laboratories. Hence, evidence of equivalence is needed to support the claim. One common practice is based on the equivalence of the means between the two laboratories [28]. This approach has a prespecified (L, U) and will declare equivalence if the 90% confidence bound of the mean difference (or percent difference for log-transformed

data) falls within (L, U). By rejecting both null hypotheses, one establishes that the mean difference is between L and U.

As an alternative, Lin et al. [32] proposed an approach using concordance correlation coefficient (CCC), which was originally proposed [31] to assess the reproducibility of an assay compared to the gold standard assay. CCC is a product of two components of accuracy and precision.The precision component is the Pearson correlation coefficient. Bland and Altman [64] pointed out that the Pearson correlation coefficient is not a measurement of agreement because it depends on the range of the true quantity in the sample set and is invariant to change of scales and systematic biases.Although the accuracy component of CCC allows CCC to be sensitive to changes in scale and to systematic biases, CCC still highly depends on the range of the true quantity in the sample set because of the precision component.

The objective of the study was to study of the effect of sample size for establishing equivalency and comparison of methods during assay transfers and to recommend criterion for out of specification risk mitigation.Equivalence test (TOST) is commonly used for analytical method transfers. The power of Equivalence test(TOST) is a probability that properly accept the equivalence at a given true mean difference Δ. It may be alternatively stated in terms of the type-II error β of falsely rejecting equivalence at a given value of Δ.

In order to achieve the desired power, the minimum number of test results needed is calculated. Three parameters are required for this type of calculation:

- **The equivalence limit E**: The equivalence limit is set based on scientific knowledge or historical data.

- **The population standard deviation** σ: The population standard deviation σ represents the precision of the test method. In practice, σ is unknown

at most of time. The long-term method performance data can be used to establish the estimated standard deviation S which is an estimate of the population standard deviation σ.

- **The true difference parameter** Δ: The true difference parameter Δ is a hypothetical value such that if the absolute value of the observed difference is no more than Δ, there is a strong probability of concluding that the two data sets represent equivalent results. The choice of Δ is arbitrary. A practical solution is to assign a value which is about half to two-thirds of the distance E to Δ.

The following inequality is used to get the minimum number of test results n to maintain a desired level of power $1 - \beta$ of the TOST:

$$\phi(\frac{E - \Delta}{\sigma_D} - Z_{1-\alpha}) - \phi(\frac{-E - \Delta}{\sigma_D} + Z_{1-\alpha}) \geq 1 - \beta \tag{14}$$

where

- $\phi(.)$ is the standard normal cumulative distribution function,
- Δ is the true difference parameter $\mu_1 - \mu_2$,
- σ_D is the standard error of the difference D, estimated by $S \times \sqrt{\frac{2}{n}}$
- $Z_{1-\alpha}$ is the $(1 - \alpha)th$ percentile of the standard normal distribution

The minimum number of n which satisfies the inequality above is defined as the minimum number of independent test results in each of two populations. In the other words, we need at least n results for population 1 and n results for population 2 ($2n$ in total) to conduct an equivalence test between two means with the desired power $1 - \beta$.

3.2.4 Proposed Approach for Method Transfers

The most difficult part of the assay transfers is establishing the equivalence margin. Most of the literature suggests use scientific judgement in establishing one and recently there has been some publication about suggested ways to establish one. According to USP <1010> [78] the equivalence margin can be established as described below:

The acceptable difference, (δ), is calculated in the following manner:

$$\text{A= LTI}-\text{LSL for LTI} \geq \text{LSL}$$

$$\text{B= USL}-\text{ UTI for USL} \geq \text{UTI}$$

Equivalence Margin (δ)= minimum (A, B) where

- LTI = Lower Tolerance Interval
- UTI = Upper Tolerance Interval
- USL = Upper Specification limit
- LSL = Lower Specification limit

This approach works if the specifications are large enough to have enough space for subtracting the tolerance intervals to determine the equivalence margin. In case of where the specifications are narrow and the variability of the methods are higher, this might lead to a situation in which there is no difference between tolerance intervals and specifications. The proposed approach is to use the equivalence margin based upon the confidence interval of difference in means:

$$E = (\overline{x_1} - \overline{x_2}) \pm t_{CL}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = \pm t_{cL}\left(\sqrt{\frac{2\sigma^2}{n}}\right) = 2 \times t_{CL} \times \frac{\sqrt{2}}{\sqrt{n}} \times \sigma \quad (15)$$

where

- where $(\overline{x_1} - \overline{x_2})$ is the difference between sample means,

- where t_{CL} is the upper $(1-\alpha/2)$ critical value for the t distribution with k degrees of freedom (with k equal to either the smaller of n_1-1 and n_2-1 or the calculated degrees of freedom)

- If we assume treating the difference in means $(\overline{x_1} - \overline{x_2}) = 0$ and $\sigma_1 = \sigma_2$ and $n_1 = n_2$

In order to include the comprehensive variability of assays, it is recommended to use the upper confidence interval of the standard deviation to capture the uncertainty of the derived standard deviation from a limited dataset. The equation now becomes:

$$E = 2 \times t_{CL} \times \frac{\sqrt{2}}{\sqrt{n}} \times \frac{S\sqrt{N-1}}{\sqrt{\chi^2_{\alpha,N-1}}} \tag{16}$$

where χ^2_α is the critical value from the chi-square distribution with $N1$ degrees of freedom.

3.2.5 Bootstrapped Equivalence Test (TOST)

Bootstrapping is a methodology that relies on random sampling with replacement. The concept of bootstrapping is to perform computations on the data itself to estimate the variation of statistics that are themselves computed from the same data. The concept of bootstrap developed by Effron (1979) is described below:

- x_1, x_2,...x_n is a data sample drawn from a distribution F.

- u is a statistic computed from the sample.

- F is the empirical distribution of the data (the resampling distribution).

- x_1, x_2, . . ., x_n is a resample of the data of the same size as the original sample n

- u is the statistic computed from the resample

Variation of u statistic depends upon the size of the sample. The concept of bootstrapping can be applied to multiple statistics like mean, median, confidence intervals when the original sample size is small and true population estimates can be estimated.

Another method to establish equivalence, is to construct a two sided $100(1-2\alpha)\%$ confidence interval for the difference between two means and compare it with the equivalence limit. This is described in Chapter 2 in section 2.3.3. The reason the confidence interval is $100(1-2\alpha)\%$ and not the usual $100(1-\alpha)\%$ is because this method is tantamount to performing two one-sided tests. For instance, using a two sided 90% confidence interval yields a 0.05 significance level for testing equivalence.

The upper (UCL) and lower (LCL) confidence limits for the $100(1-2\alpha)\%$ two-sided confidence interval on the true difference are computed as follows:

$UCL = D + t_{1-\alpha,f} S_D$

$LCL = D - t_{1-\alpha,f} S_D$

where $t_{1-\alpha,f}$ is the upper $100(1-\alpha)$ % percentile of the Students t distribution with the degrees of freedom ($f = n_1 + n_2 - 2$). The method proposed is an extension of Two one-sided t test. In case of analytical method transfer, a small sample size (n=18,24 etc.,) is used and equivalence is claimed based upon equivalence test hypothesis.

The approach is described below:

- Step 1: Collect data from sending lab ($\overline{x_1}$) and Receiving lab ($\overline{x_2}$)
- Step 2: Simulate 10,000 sets of datasets of same size from sending and receiving lab using bootstrapping method
- Step 3: Compute the mean difference between averages for sending and receiving lab ($\overline{x_1} - \overline{x_2}$) mean difference of a bootstrap sample ($\overline{x_1} - \overline{x_2}$) *1 . .

., $(\overline{x_1} - \overline{x_2})$ *10000 times.

- Step 4: Compute the confidence interval of the mean difference based upon 10000 resamples
- Step 5: Compare the confidence intervals and compare it with the equivalence limit.
- Step 6: If the confidence interval is completely contained within the equivalence limits (-E,E), then accept equivalence. Otherwise, reject equivalence.

Also as part of the assay transfer, Process Capability needs to be calculated for the sending lab before initiation of transfer process. A process where almost all the measurements fall inside the specification limits is a capable process. This can be represented by the plot below:

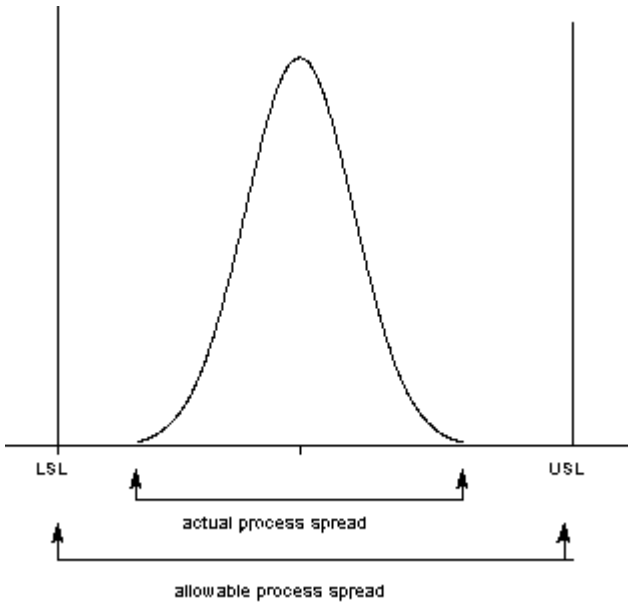


Figure 8: A capable process in which all the measurements fall in the specification limits

C_{pk} statistic assumes that the population of data values is normally distributed. Assuming a two-sided specification, if $\overline{x}$ and σ are the mean and standard

deviation, respectively, of the normal data and USL, LSL, the upper and lower specification limits respectively, then the capability indices are defined as follows:

$$C_{pk} = min\left\{\frac{USL - \overline{x}}{3\sigma}, \frac{\overline{x} - LSL}{3\sigma}\right\} \tag{17}$$

Table 3 below indicates C_{pK} values and associated risk of failure. The greater the C_{pK} value, risk is further reduced. For practical implementation, a minimum value of C_{pK} is required such that future results are $< 0.27\%$ out of specification(OOS) values.

C_{pk}	Sigma level	% out of specification	PPM out of tolerance
0.33	1	31.73	317310.508
0.5	1.5	13.36	133614.403
0.67	2	4.55	45500.264
0.83	2.5	1.24	12419.331
1	3	0.27	2699.796
1.17	3.5	0.05	465.258
1.33	4.0	0.01	63.342
1.50	4.5	0.001	6.795
1.67	5.0	0.0001	0.573
1.83	5.5	0.000004	0.038
2.00	6.0	0.0000002	0.002

Table 3: C_{pk} values and the corresponding associated risk of failure which is determined by the % out of specification.

If the process capability (C_{pK}) of sending lab is less than 1.0, the transfer of assay should not be initiated and assay needs to be remediated to improve capability. The transfer between the labs can be initiated if capability is greater than 1.0 and once the transfer is complete, capability at the sending lab need to be calculated. The following equation is proposed for C_{pK} calculation at the receiving lab.

$$C_{pk} = min\left\{\frac{\overline{\overline{x}} - LSL}{3\sigma}, \frac{USL - \overline{\overline{x}}}{3\sigma}\right\} \tag{18}$$

where

$\overline{\overline{x}} = \overline{x} + (2 \times t_{CL} \times \frac{\sqrt{2}}{\sqrt{n}} \times \frac{S \times \sqrt{N-1}}{\chi^2_{\alpha,N-1}})$

The proposed capability will include any potential bias that might be resulted from the method transfer and will control any out of specification scenarios. Once the testing at the receiving lab is completed, Process capability index needs to be calculated and if the capability is > 1.0 and TOST results indicate equivalence, the transfer can be claimed successful.

The flowchart of the proposed approach for assay transfer is shown in Figure 9

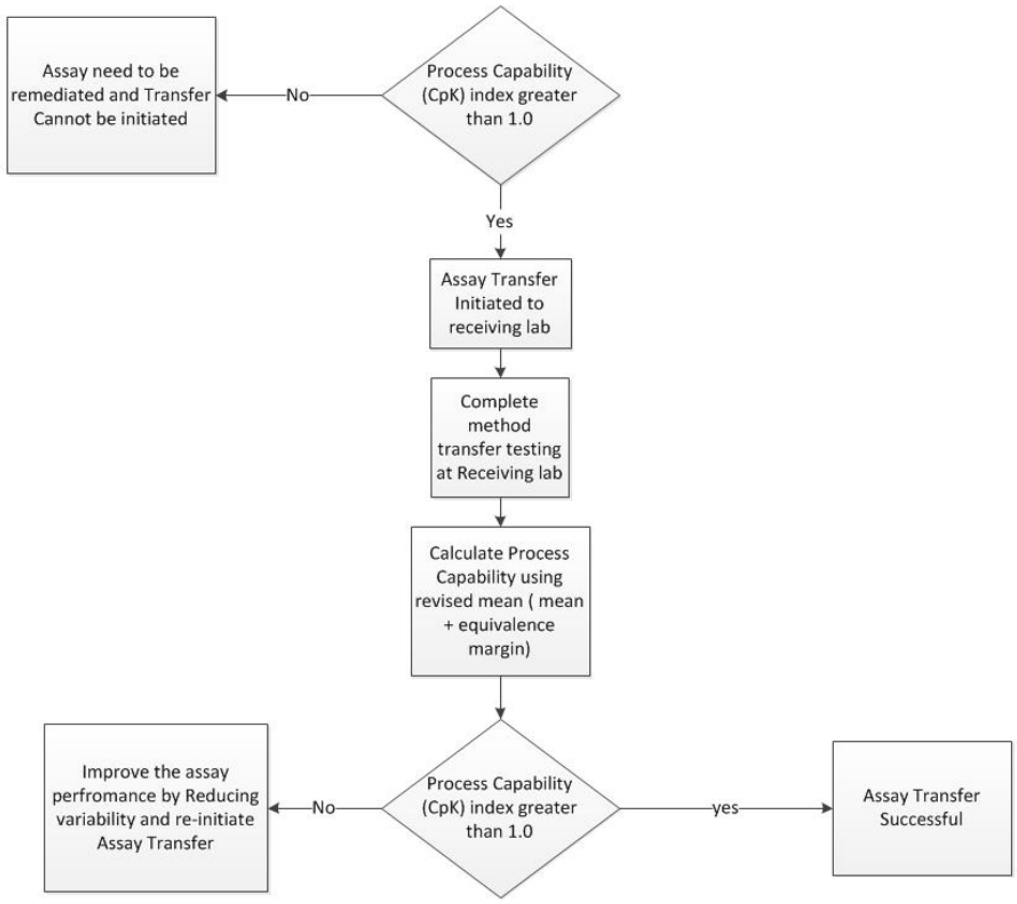


Figure 9: Flowchart of the proposed approach

3.3 Summary

In this chapter, research design has been presented. The details of research methodology, statistical methodologies, data collection are discussed. The research questions and the formulation of hypothesis are also highlighted.

CHAPTER 4

RESULTS AND ANALYSIS

This chapter summarizes the results and analysis of the data. The results are presented and divided in three sections

1. Process Optimization
2. Analytical Transfer
3. Stability Data Analysis

4.1 Process optimization

Simulations are run using the simulation software Pensim V2.0 as described in Chapter 3 under closed-loop control of pH and temperature since those variables play an important role on the quality and quantity of the final product, whereas glucose addition is performed under open-loop. In all runs, a batch culture is followed by a fed-batch operation by the depletion of carbon source. In general, the system switches to the fed batch mode after about 45 h. A constant glucose feed is used during the fed-batch operation. A more detailed description of the process including the state equations and simulation conditions is given by Birol et al.[68]. The variables used in the modeling of Penicillin manufacturing simulation are tabulated in Table 4.

Variables Number	Variable Name
1	Aeration rate(I/h)
2	Agitator Power (W)
3	Substrate feed rate(I/h)
4	Substrate feed temperature(K)
5	Dissolved oxygen concentration (g/l)
6	Culture volume (l)
7	Carbon dioxide concentration (g/l)
8	pH
9	Bioreactor temperature (K)
10	Generated heat (kcal)
11	Cooling water flow rate (l/h)
12	Substrate concentration (g/l)
13	Biomass concentration (g/l)
14	Acid Flow Rate (I/h)
15	Base Flow Rate
16	Penicillin concentration

Table 4: Variables used for Modeling of Penicillin Manufacturing process

The pH was varied between $4.95 - 5.0$ in order to simulate the observed behavior of penicillin production by utilizing an on/off or a proportional-integral-derivative (PID) controller. The pH is regulated by adding highly concentrated (3 M) acid or base solution when necessary. Two PID controllers are used to manipulate the acid and base control values. The PID controllers are tuned for a certain range of initial conditions considered to be the normal operation. The initial conditions are summarized in Table 5.

Substrate Concentration	15g/L
Dissolved Oxygen Concentration	1.16 g/L
Biomass Concentration	0.1g/L
Penicillin Concentration	0 g/L
Culture Volume	100 L
Carbon Dioxide Concentration	0.5g/L
pH	5.0
Fermentor Temperature	298K

Table 5: Initial conditions for simulation

The duration of each batch is 400 h, comprising a preculture stage (about 45 h) and a fed-batch stage (about 355 h). All batches are assumed to be of the same duration. The sampling interval is 0.5 h. Small variations were added to the simulation input data to mimic the variations in the normal operating conditions encountered in the real process. The ranges of the set points and scenarios with setpoint variations are listed in Table 6. The set points were varied within the range and 30 batches with five different combination of ranges were included in the simulation to generate a total of 150 batches. As discussed in Chapter 3, due to high dimensionality of data, few methods such as Principal component analysis(PCA) and Partial least squares (PLS) were chosen to analyze the data for dimensional reduction. PCA and PLS methodologies belong to classification of unsupervised algorithms.

The statistical methodologies used for the current study are:

- Principal component analysis (PCA) (Unsupervised learning)
- Partial least Squares (PLS) (Unsupervised Learning)
- Symbolic Regression (Supervised Learning)
- Support Vector Regression (SVR) (Supervised Learning)

Set points	Range	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Aeration Rate (g/h)	$8-9$	8.6	8.7	8.8	8.9	9.0
Agitator Power (W)	29-31	29.9	29	29.5	30	31
Substrate feed flow rate (l/h)	0.039-0.045	0.0426	0.039	0.041	0.043	0.045
Substrate Feed Tempera-ture(K)	295-296	296	296	295	296	295
Bioreactor Tempera-ture (K)	297-298	298	298	297	298	297
pH	4.95-5.05	5	4.95	5	5.05	5

Table 6: Set point ranges and scenarios for simulation

The results obtained were compared to a 2-stage **Scores based Symbolic regression (SBSR)** method.The results from each model depend on the model parameters. RMSECV minimization was used for optimization in all cases and for all models.

4.1.1 Results

Principle Component Analysis(PCA)

PCA gives an overview of the information in a data table. This summary shows how the observations are related and if there are any deviating observations or groups of observations in the data. Of particular interest in process data analysis is the ability of PCA to uncover both smooth time trends and sudden shifts in the data. In addition, with PCA we also gain an understanding of the relationships among the variables: which variables contribute similar information to the PCA model, and which provide unique information about the observations. PCA describes the correlation structure in X (all the 16 variables described in Table

4). Geometrically, PCA finds lines, planes and hyperplanes in the K-dimensional variable space that approximate the data as well as possible in the least squares sense. Before analyzing the data, some of the variables were transformed due to the skewness of the distribution. All variables are scaled to Unit Variance. The table below shows the formulae for transformation of the two variables.

Variable	**Transform**	**Formula**
Substrate Concentration	Log	$\log_{10}(x)$
flb:Base flow rate	Log	$\log_{10}(x + 5e^{-006})$

Table 7: Variable Transformation and the formulae used for transformation

PCA and PLS work best with normally distributed data. log transformation can decrease the variability of data and make data conform more closely to the normal distribution. The PCA model consists of 16 X-variables and 0 Y-variables. Hence, the PCA can be written as PCA-X.

Table 8 shows the summary of the model.

Principal Component	$R^2(X)$	$R^2(X)$(cum)	Eigenvalue	Q^2	Q^2(cum)
1	0.509	0.509	8.14	0.468	0.468
2	0.188	0.697	3.01	0.258	0.605
3	0.0876	0.785	1.4	-0.0925	0.569
4	0.0833	0.868	1.33	0.243	0.673

Table 8: Summary of the PCA model

R^2X is the percent of the variation of all the X explained by the model and Q^2 is the percent of the variation of all the X that can be predicted by the model. Eigenvalues measure the amount of variation in the total sample, accounted for by each factor.The number of principal components retained in the model are based upon two criteria:

1. Eigenvalue one criterion: Based on this criterion the first components with eigenvalues higher than 1 are chosen.

2. Amount of explained variance: based on this, the chosen factors should explain 70 to 80% of the variance at least.

Based upon the above criterion, first 4 principal components were chosen for the analysis. The PCA model explains 86.8% of variance between X variables.

Figure 10 below displays the cumulative R^2 and Q^2 for the X matrix, after each component.

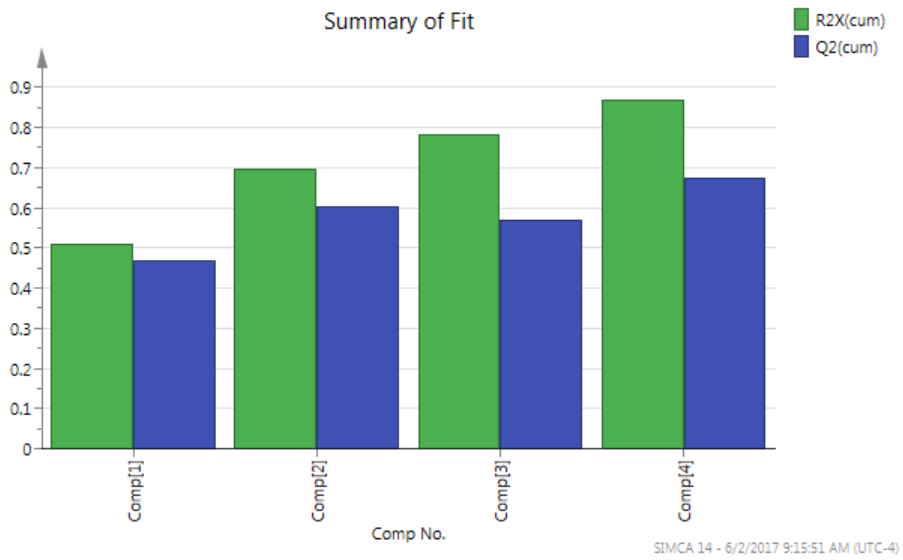


Figure 10: R^2 and Q^2 cum for each componentfor the X-matrix

Figure 11 shows a plot that displays the cumulative $R^2V(X)$ and $Q^2V(X)$ for each X variable.

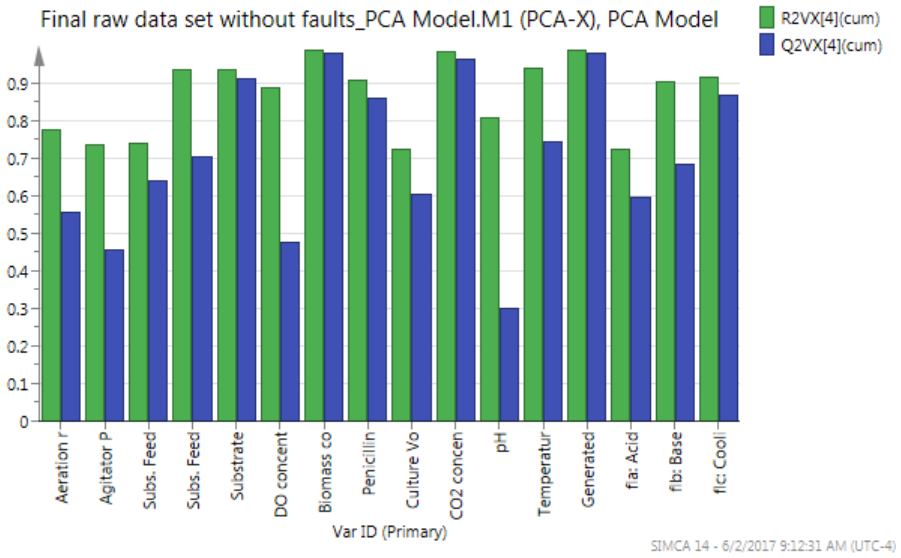


Figure 11: Goodness of fit of the X variables

X-axis denotes the variable and y-axis indicates the amount of variation. $R^2V(X)$ describes the cumulative percent of the variation of the X variable explained by the model after the last component. R^2 is a measure of fit, i.e. how well the model fits the data. $Q^2V(X)$ describes the cumulative percent of the variation of the X variable predicted by the model, after the last component, according to cross validation. Q^2 is a measure of how well the model predicts the variable. The plot indicates that substrate feed, substrate concentration, biomass concentration, CO_2 concentration and generated heat contribute to the maximum amount of variation.

Figure 12 shows a loading scatter plot $P[2]$ vs $P[1]$. The scores are weighted averages of the variables with weights p1(principal component) in the first dimension and p2(principal component) in the second dimension. These weights, the loadings p, are computed from the correlation structure of the X's.Loading scatterplot displays all the variables at the same time. Variables contributing similar information are grouped together and are positively correlated. Variables that are inversely correlated are positioned on opposite sides of the plot origin, in diagonally opposed quadrants. Figure12 indicates that the Penicillin concentration is positively correlated to biomass concentration, culture volume and negatively related to substrate feed and dissolved Oxygen concentration.

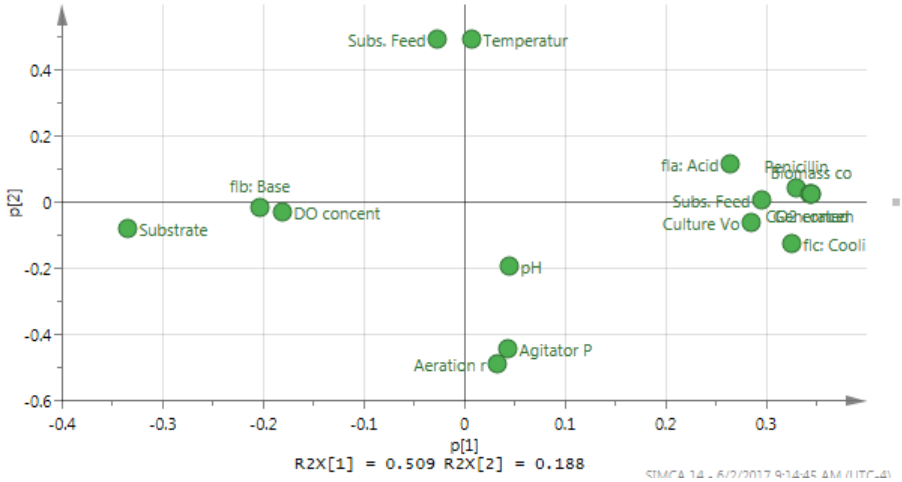


Figure 12: Loading Scatter plot P(2) vs P(1)

Hence P2 vs P1 displays how the X variables correlate with each other and contribute to the model. Points that are far away from the origin have a strong impact on the model, whereas points that are closer to the center have a weaker influence.

Partial Least Squares (PLS)

PLS is a method for relating two data matrices, X and Y, to each other by a linear multivariate model. The PLS model has 16 variables, out of which 15 are X variables and 1 is the Y variable. Table 9 shows all the variables used in the PLS model.

The model is trained on a workset for which both X and Y are available. It may then be applied to a prediction set of new observations. In this prediction, known X data for the prediction set are utilized to predict the unknown Y data. The PLS model may be interpreted to understand how the X variables influence the Y variables. This information is useful when the goal is to modify the X variables to achieve an improved profile among the Y variables. The PLS method derives its usefulness from its ability to analyze data with many, noisy, collinear, and even incomplete variables in both X and Y. In this study, all the variables are

X variables	Y variable
Aeration rate (l/h) Agitator power (W) Substrate feed rate (l/h) Substrate feed temperature (K) Dissolved oxygen concentration (g/l) Culture volume (l) Carbon dioxide concentration (g/l) pH Bioreactor temperature (K) Generated heat (kcal) Cooling/heating water flow rate (l/h) Substrate Concentration Biomass Concentration Acid Flow Rate (l/h) Base Flow Rate (l/h)	Penicillin concentration

Table 9: Variables used in the PLS model

denoted as X (observation variables) except penicillin concentration that will be treated as Y (response variable). Before analyzing the data, some of the variables were transformed due to the skewness of the distribution. PCA and PLS work best with normally distributed data. Log transformation can decrease the variability of data and make data conform more closely to the normal distribution Table 10 shows the transformed variables. All variables are scaled to Unit Variance.

Variable	**Transformation**	**Formula**
Substrate concentration	Log	$log_{10}(x)$
Generated heat	Log	$log_{10}(x + 0.01569)$
Base flow rate	Log	$log_{10}(x + 5e^{-0.006})$

Table 10: Variables and the formulae used for transformation

Table 11 displays the data of cumulative R^2 and Q^2 for the Y matrix after each component.

Principal Component	$R^2(X)$	$R^2(X)(cum)$	Eigenvalue	$R^2(Y)$	$R^2(Y)(cum)$	Q^2	$Q^2(cum)$
1	0.465	0.465	6.97	0.86	0.86	0.86	0.86
2	0.0996	0.564	1.49	0.0827	0.943	0.0591	0.919
3	0.14	0.704	2.1	0.0204	0.963	0.0356	0.954
4	0.101	0.805	1.52	0.0103	0.973	0.00278	0.9574

Table 11: Data of the cumulative R^2 and Q^2

R^2X is the percent of the variation of all the X explained by the model and Q^2 is the percent of the variation of all the X that can be predicted by the model. R^2Y is the percent of the variation of all the Y explained by the model and Q^2 is the percent of the variation of all the Y that can be predicted by the model. The above model explains that 97.3% variation and 95.75% of variation can be predicted by the model.

The plot shown in Figure 13 shows the data in a graphical format.

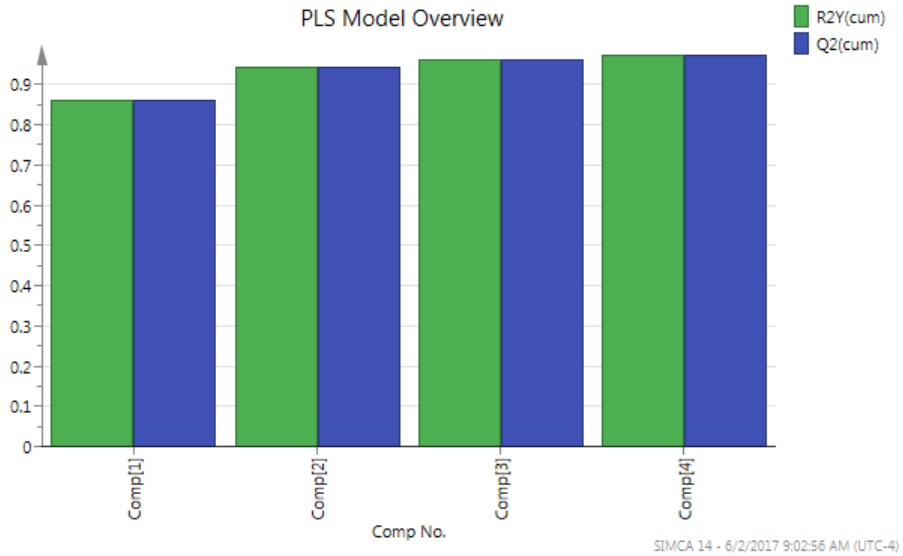


Figure 13: Data and plot of $R^2(cum)$ and $Q^2(cum)$ for the Y-matrix for each component

Here, $R^2V(Y)$ describes the cumulative percent of the variation of the Y

variable explained by the model after the last component. R^2 is a measure of fit, i.e. how well the model fits the data. $Q^2V(Y)$ describes the cumulative percent of the variation of the Y variable predicted by the model, after the last component, according to cross validation. Q^2 is a measure of how well the model predicts new data.

Model Summary

A scatter plot of the X and Y weights (w* and c) is shown below.

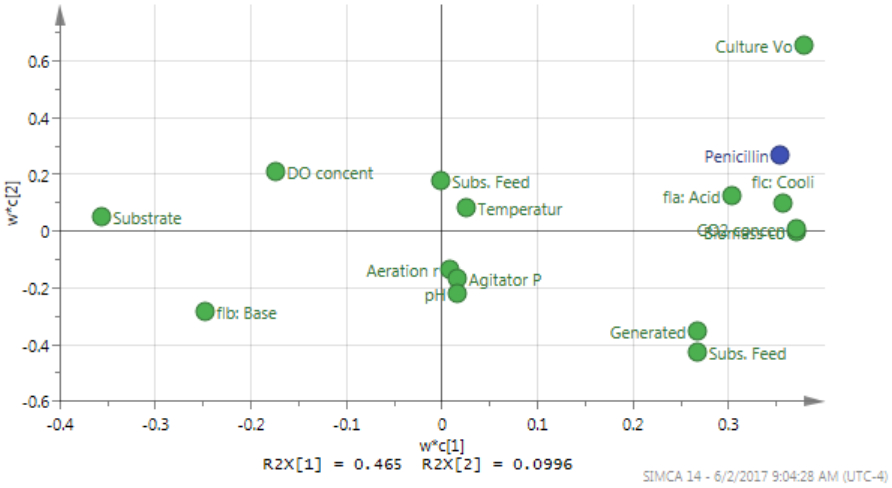


Figure 14: Loadings w*c[1] vs. w*c[2]

The w* are PLS weights. These are the weights that in PLS combine the original variables in X to form the new variables, scores t. "c" represent the weights used to combine the Y's to form the scores u.

Figure 14 shows the relation between the X variables and the Y variables, and the relation within the X's and the Y's. Points that are far away from the origin have a strong influence on the model, whereas points that are closer to the center have a weaker influence. The loadings plot indicates that culture volume, CO_2 concentration, biomass concentration and substrate concentration have stronger influence on the model.

Variable importance plot (VIP) can also be plotted as part of the PLS analysis. The VIP plot reflects the importance of terms in the model both with respect to Y, i.e., correlation to all the responses, and with respect to X (the projection). The VIP values summarize the overall contribution of each X-variable to the PLS model, summed over all components and weighted according to the Y variation accounted for by each component. VIP is normalized, which means the average squared VIP value is 1. Thus, terms in the model with a VIP > 1 are the most important. In Figure 15 shown below, the variables that have a VIP > 1 are considered important while the other variables are insignificant. The plot indicates culture volume, CO_2 concentration, bio mass concentration, Substrate concentration, cooling generated, acid added, substrate feed and generated heat are the most important terms in the model.

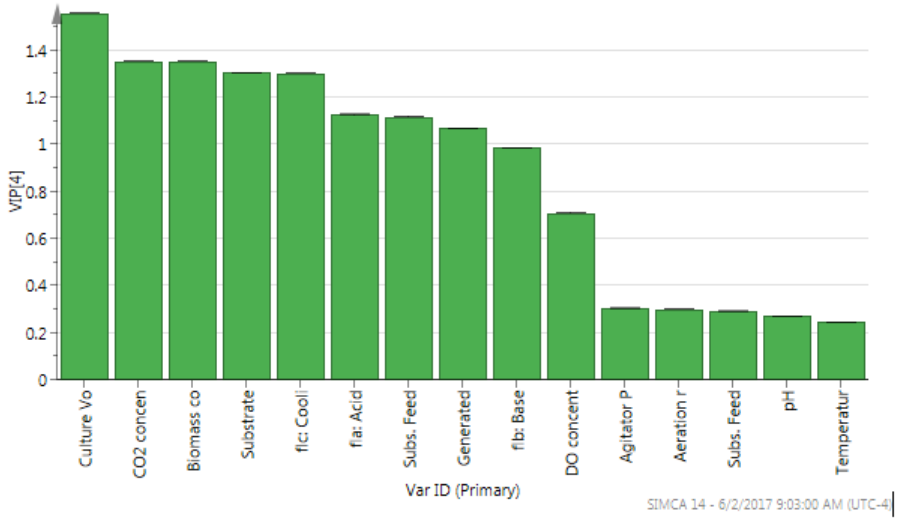


Figure 15: The variables that have a VIP value > 1 are the only significant ones.

Figure 16 shows the Observed vs. Predicted plot that displays the observed values vs. fitted values for the Y-variable (i.e, Penicillin concentration), using PLS model. The RMSEE (Root Mean Square Error of Estimation) indicates the fit error of the observations to the model. RMSECV is a similar measure, but estimated using cross validation.

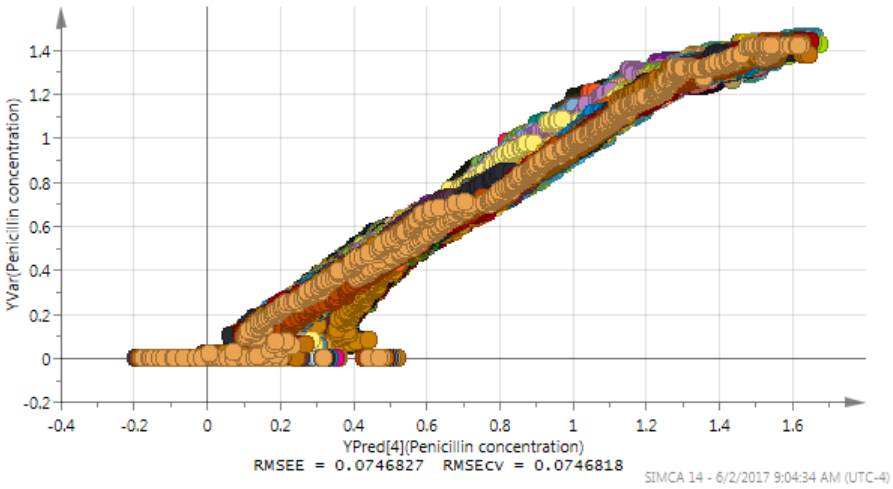


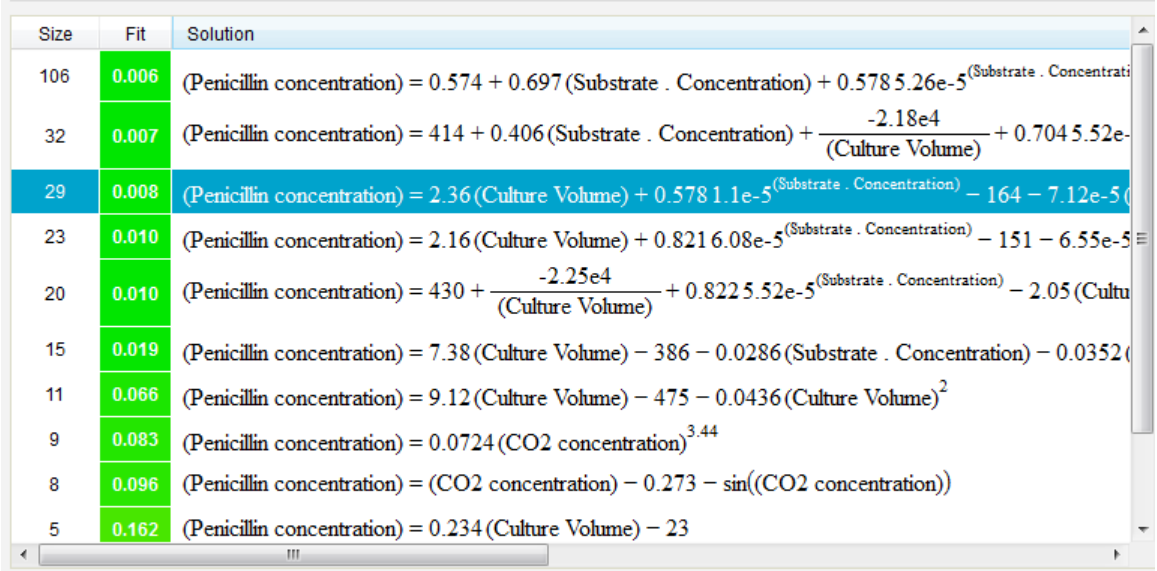
Figure 16: Observed vs. Predicted plot of the Penicillin concentration for all the 150 batches. The different colors in the plot represent different batches.

As seen from the plot, the RMSE is low which indicates better predictability for the Penicillin concentration.

Symbolic Regression(SR)

Symbolic regression analysis has 16 variables, out of which 15 are X variables and 1 is the Y variable (penicillin concentration). The dataset collected from Pensim simulator V2.0 is used for the analysis. The output from Eureqa consists of series of equations that best fit the relation between X variables and Y variable. In the output shown in Figure 17 below, the first column(size) indicates the complexity measure, the second column indicates error measure and the solution column indicates the corresponding candidate equation. Eureqa's complexity metric(or size) is measured both by the number of variables used within the solution as well

as the relative weights of each of the building blocks used in the solution. The best solution is the one that has the optimum balance between error and complexity.



Size	Fit	Solution
106	0.006	(Penicillin concentration) = 0.574 + 0.697 (Substrate . Concentration) + 0.578 5.26e-5$^{(\text{Substrate . Concentrati}}$
32	0.007	(Penicillin concentration) = 414 + 0.406 (Substrate . Concentration) + $\frac{\text{-2.18e4}}{\text{(Culture Volume)}}$ + 0.704 5.52e-
29	0.008	(Penicillin concentration) = 2.36 (Culture Volume) + 0.578 1.1e-5$^{(\text{Substrate . Concentration})}$ − 164 − 7.12e-5 (
23	0.010	(Penicillin concentration) = 2.16 (Culture Volume) + 0.821 6.08e-5$^{(\text{Substrate . Concentration})}$ − 151 − 6.55e-5
20	0.010	(Penicillin concentration) = 430 + $\frac{\text{-2.25e4}}{\text{(Culture Volume)}}$ + 0.822 5.52e-5$^{(\text{Substrate . Concentration})}$ − 2.05 (Cultu
15	0.019	(Penicillin concentration) = 7.38 (Culture Volume) − 386 − 0.0286 (Substrate . Concentration) − 0.0352 (
11	0.066	(Penicillin concentration) = 9.12 (Culture Volume) − 475 − 0.0436 (Culture Volume)^2
9	0.083	(Penicillin concentration) = 0.0724 (CO2 concentration)$^{3.44}$
8	0.096	(Penicillin concentration) = (CO2 concentration) − 0.273 − sin((CO2 concentration))
5	0.162	(Penicillin concentration) = 0.234 (Culture Volume) − 23

Figure 17: The series of equations obtained from Eureqa Formulize.

Eureqa Formulize produces a series of equations and ranks each in terms of goodness of fit. The highlighted equation can be considered as the optimum solution obtained from the series of equations. After this point, increasing the complexity resulted in the amount of errors that can be ignored due to the discrepancies.

Figure 18 shows the mathematical solutions accuracy vs. its complexity. The solution with a mean absolute error 0.03 and complexity of 29 was considered as optimum solution. After this point, increasing the complexity resulted in the amount of errors that were ignored due to the discrepancies. The stability and percent converged of final solutions after 1.79×10^3 generations were 1.977% and 97.3%, respectively.

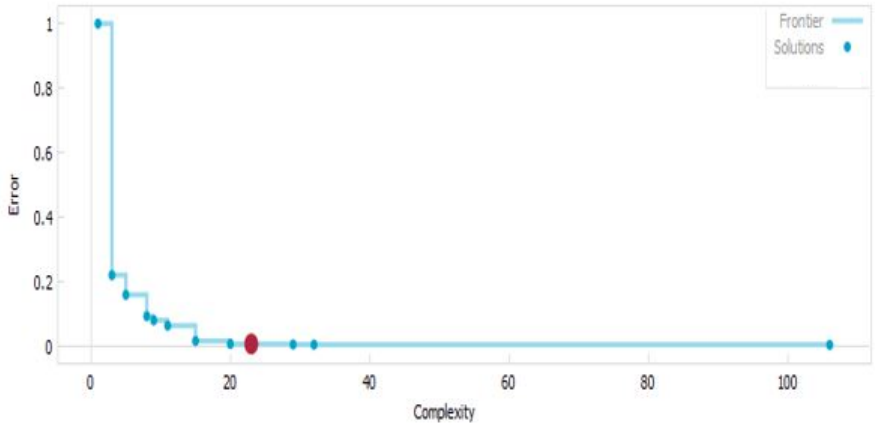


Figure 18: Error vs complexity obtained from the mathematical solutions

Ultimately based on the optimum point, the proposed mathematical solution can be presented as:

$(Penicillinconcentration) = 2.36 * (CultureVolume) + 0.5781.1e - 5^{(}Substrate.Concentration) - 164 - 7.12e - 5 * (CultureVolume)^2 - 8.60944211627824e - 6 * (CultureVolume) * (Substrate.Concentration)^2$

The parameters of the solution are summarized in the Table 12 below:

R^2 goodness of fit	0.99034313
Correlation Coefficient	0.99542721
Maximum Error	0.48358019
Mean Squared Error	0.002019551
Mean Absolute Error	0.03344815
Coefficients	5
Complexity	29

Table 12: Summary of the solution

Each of the terms in the table is described as:

- R^2 Goodness of Fit: (1- SS_{res}/SS_{tot}) where SS_{tot} is proportional to the total variance, and SS_{res} is the residual sum of squares.
- Maximum Error: Minimizes the single highest error of the residuals. It is used to minimize the worst-case error or to force algorithm to model a small residual feature.
- Mean Absolute Error: Minimizes the mean of the absolute value of residual errors, mean(abs(error)). Assumes noise follows a double exponential distribution.
- Mean Squared Error: Minimizes the mean of the squared residual errors. Assumes noise follows a normal distribution.

The solutions thus obtained are plotted against the data, that includes both training and validation data. The training data is a subset of the current data that is used by Eureqa to search for solutions. The validation data is a second subset that is used only to measure accuracy.

Figure 19 displays the correlation between actual observed values which appear in the training and validation data against the values predicted by the selected model.

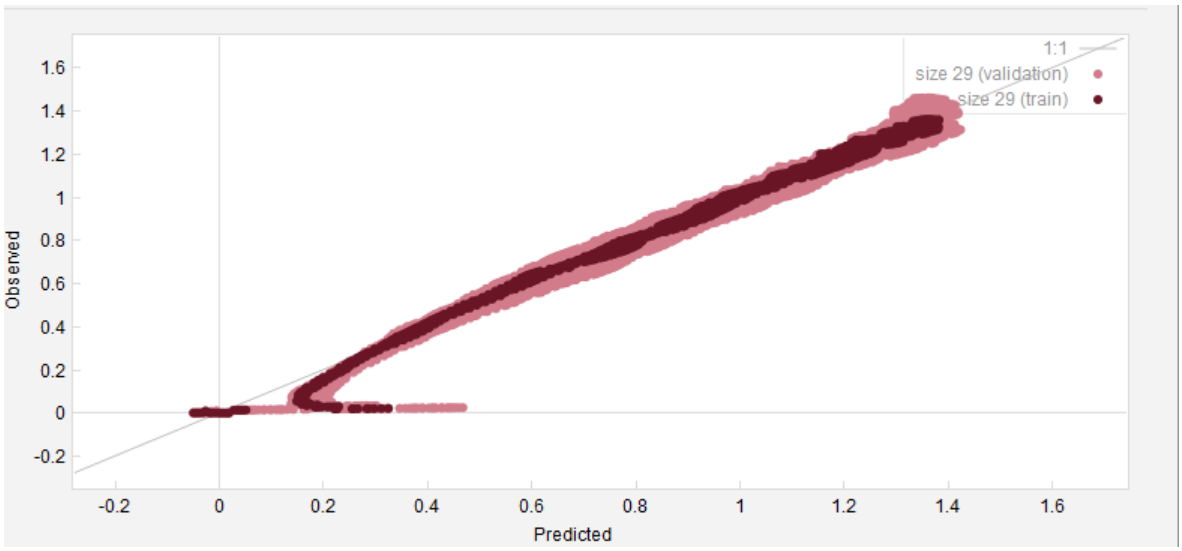


Figure 19: Observed vs predicted data. The plot shows both training and validation data.

The plot indicates high correlation between observed and predicted values. Also the solution summary indicates a high R^2 value that explains more than 99% of the variation by the model.

The plot shown in Figure 20 displays the size of the error for each point in the training and validation data. As seen in the plot, the residual error is close to 0 for training and validation data.

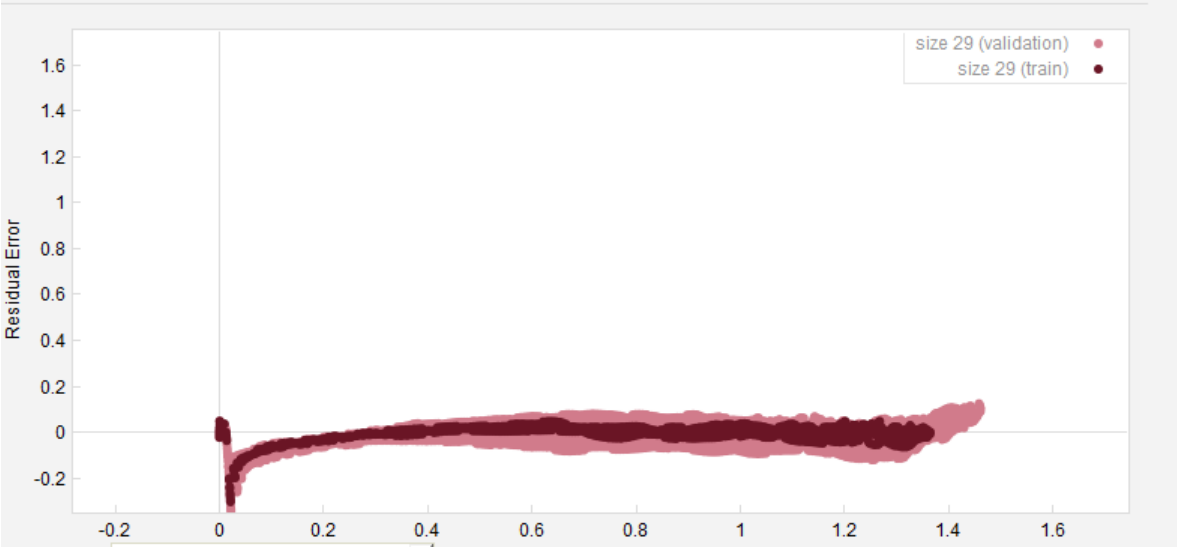


Figure 20: Residual Error plot for both training and validation data.

Figure 21 displays the receiver operating characteristic(ROC) Curve for the data. ROC curves are used for classification models which predict target variables

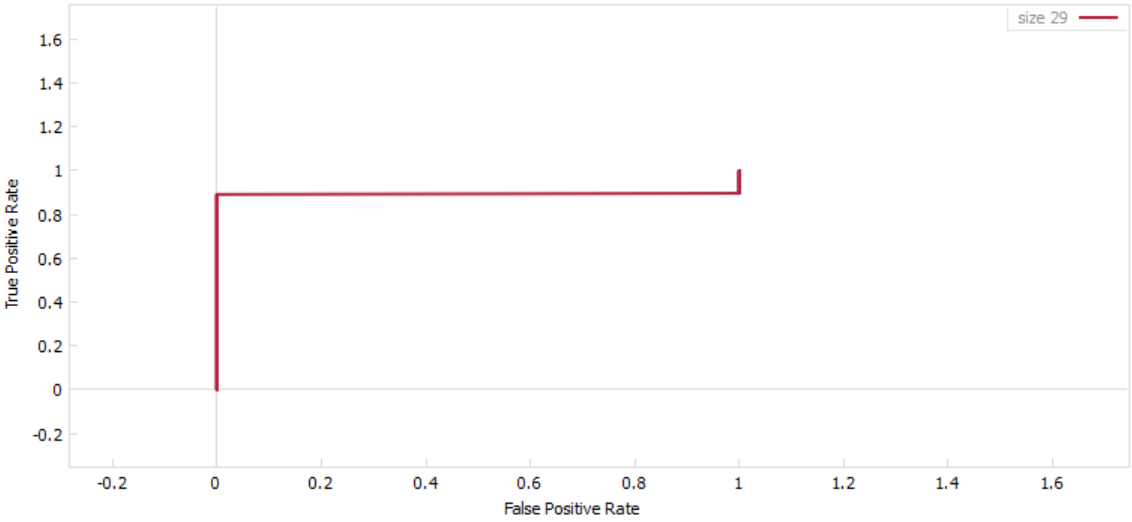


Figure 21: ROC curve.

that always have a value of 0 or 1. This procedure is a useful way to evaluate the performance of classification schemes in which there is one variable with two categories by which subjects are classified. The ROC curve is plotted using True Positive rate on y-axis and False Positive rate on x-axis. Area under the ROC

curve is a good measure for the model. A larger area under the curve indicates a better model.

Table 13 presents a sensitivity analysis over optimum solution. Sensitivity means the relative impact that a predictor has on the target variable (Penicillin Concentration).%Positive is the likelihood that increasing this variable will increase the target variable; Positive Magnitude is when an increase in this variable leads to an increase in the target variable, this is generally how big the positive impact is. %Negative is the likelihood that increasing this variable will decrease the target variable, and finally Negative Magnitude is when an increase in this variable leads to a decrease in the target variable and this is generally how big the negative impact is. The model is most sensitive to Substrate Concentration (sensitivity = 12.82) even more than to culture volume (sensitivity = 0.597). Results show that at 100% of time, Substrate Concentration will have negative impact with magnitude of 12.82, and at 96% of times, culture volume will have positive impact on forecasting penicillin concentration with a magnitude of 0.11763.

Variable	**Sensitivity**	**% Positive**	**Positive Magnitude**	**% Negative**	**Negative Magnitude**
Substrate concentration	12.82	0%	0	100%	12.82
Culture Volume	0.5967	96%	0.61566	4%	0.11763

Table 13: Sensitivity report

Support vector regression (SVR)

SVR is a support vector machine (SVM) algorithm that is used to solve regression problems. This leads to an optimized generalization performance by maximizing a space that exists between two layers. Instances that are the most adjacent to a hyperplane that has the maximum space or instances located at the shortest distance from a plane are called support vectors. The SVR equation described by Kim.et, al (2016) is shown as:

$$f(x) = \sum_{i=1}^{l} \alpha_i t_i K(x_i.x) + b \tag{19}$$

All the results calculated by using a kernel function K and a test sample x for every x_i having l support vector(s) are added together. α is a Lagrange multiplier, t is an integer that represents the category, and b is a constant that represents the location on the hyperplane.

The analysis has 16 variables, out of which 15 are X variables and 1 is the Y variable (penicillin concentration). The dataset collected from Pensim simulator V2.0 is used for the analysis.

SVR analysis was performed using WEKA and R software. Waikato Environment for Knowledge Analysis(WEKA) is a machine learning program developed by the University of Waikato in New Zealand enabling the user to analyze data and to perform prediction modeling by using various machine learning algorithms. SMOreg (weka. classifiers. functions. SMOreg) was used in the study for developing regression model. Support vector machine for regression can be implemented using SMOreg module of WEKA with arbitrary kernel functions. Nominal attributes can be transformed in to binary form and SMO algorithm can replace all missing values. This algorithm has number of features that includes fast learning and better scaling properties. The fitness and the statistical significance of the models developed in this study was assessed using the statistical parameters such

as R, R^2, MAE and RMSE.

Summary of results from linear-epsilon SVR Below is the summary of results obtained using linear SVR.

Statistic	Value
Corelation coefficient	0.9977
Mean absolute error	0.0448
Root mean squared error	0.0498
Relative absolute error	0.123684
Root relative squared error	0.115497

Table 14: Summary of results from SVR

The following classifier options are used:

- C: The complexity constant C. (default 1) item N: 0=normalize
- I: weka. classifiers. functions. supportVector.RegSMOImproved
- K: The Kernel to use: weka. classifiers. functions.supportVector.PolyKernel
- T: The tolerance parameter for checking the stopping criterion. (0.001)
- V: Use variant 1 of the algorithm when true, otherwise use variant 2. (true)
- P: The epsilon for round-off error. ($1.0e^{-12}$)
- L: The epsilon parameter in epsilon-insensitive loss function. ($1.0e^{-3}$)
- W: The random number seed. (1)
- D: Enables debugging output (if available) to be printed. (off)
- no-checks: Turns off all checks - use with caution! (checks on)
- C: The size of the cache (a prime number), 0 for full cache and -1 to turn it off. (250007)

- E: The Exponent to use. (1.0)
- L: Use lower-order terms. (no)

Complexity parameter controls the process flexibility for drawing the line to fit the data. Another important aspect in SVR is the type of kernel to use. The linear kernel is simplest that classifies data with a straight line or hyperplane. Polynomial Kernel is used to fit non-linear data. Radial basis function kernel is also used to learn complex shapes to fit training data.

In-order not to over fit the data and to detect the best cost parameter, SVR algorithm was run multiple times by changing the cost from 0.1 to 1.0 in increments of 0.1. Figure 22 shows how R^2 changes based upon the change in cost parameter.

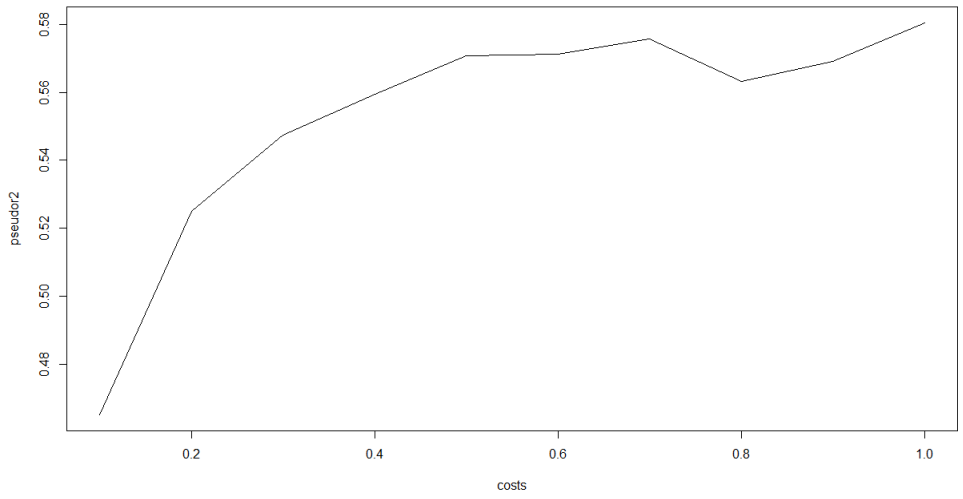


Figure 22: Cost Paramter vs pseudo R^2

As seen in the plot, the Pseudo R^2 increases with the cost parameter upto a certain value. Beyond a cost parameter of approximately 0.5, the R^2 remains fairly constant. Since the cost parameter controls the trade-off between achieving a low error on the training data and minimizing the weights, it need not be increased beyond a certain value since a further increase in cost parameter only results in

the increase of the complexity of the hypothesis class.

Suggested Methodology: Scores Based Symbolic Regression

PCA Scores

In the case of Penicillin manufacturing process, suppose n variables have been measured at p time intervals. The information can be formulated in a matrix form as:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

where $x1 = [x_{11}x_{12}...x_{1p}]$ is the row vector containing the penicillin concentrations measured at p time intervals for the first lot, x_2 is the row vector containing the penicillin concentrations for the second sample and so on. PCA creates new orthogonal variables (latent variables) that are linear combinations of the original x-variables. Singular value decomposition of the matrix can be used for X:

$$X_{nxp} = U_{nxp}\Lambda_{pxp}P'_{pxp} = T_{nxp}P'_{pxp} \tag{20}$$

U is the unweighted/normalized score matrix and T is the weighted /unnormalized score matrix. The new matrices contain the new variables. P is the loading matrix and the column vectors are called eigen vectors. Elements of matrix P are the loadings of original variables on each eigenvector. Diagonal matrix is denoted by L. The decrease in principal components explain decreasing amount of variability in X.

Model Summary The analysis has 16 variables, out of which 15 are X variables and 1 is the Y variable (penicillin concentration). The dataset collected from Pensim simulator V2.0 is used for the analysis. In the case of Penicillin manufacturing only the first 3 principal components were retained. The scores were calculated on

the 3 principal components of X variables and calculated on the Y variables. The data set for scores consists of 12000 rows and 4 columns. Due to large size of the dataset used, only a subset of data is displayed below:

Obs ID (time stamp (h))	Principal Component.t[1]	Principal Component.t[2]	Principal Component.t[3]	Penicillin Concentration.t[1]
0	-7.60569	0.654546	-4.70945	-1.82627
0.5	-7.60064	0.66779	-4.59685	-1.82627
1	-7.52415	0.691612	-4.43994	-1.82627
1.5	-7.52901	0.688805	-4.56128	-1.82624
2	-7.55764	0.68729	-4.60966	-1.82624
2.5	-7.44015	0.70481	-4.43102	-1.82624
3	-7.93164	0.684611	-3.24621	-1.82624
3.5	-7.84162	0.69266	-3.2239	-1.82624
4	-7.91394	0.685491	-3.1825	-1.82622
4.5	-7.88512	0.685142	-3.22685	-1.82622
5	-7.83358	0.691493	-3.19015	-1.82622
5.5	-7.7484	0.706234	-3.03812	-1.82622
6	-7.82135	0.696165	-3.16513	-1.82622
6.5	-7.72326	0.713198	-3.05555	-1.8262
7	-7.70703	0.714202	-3.13309	-1.8262
7.5	-7.66604	0.72109	-3.01039	-1.8262
8	-7.67406	0.72308	-2.97925	-1.82618
8.5	-7.70064	0.726664	-2.98583	-1.82618
9	-7.57703	0.750119	-2.79395	-1.82618
9.5	-7.57645	0.753466	-2.80763	-1.82616
10	-7.60586	0.753112	-2.78363	-1.82616
10.5	-7.52205	0.755196	-2.75026	-1.82616
11	-7.56105	0.744883	-2.82163	-1.82614
11.5	-7.4021	0.775713	-2.49554	-1.82614
12	-7.4062	0.7672	-2.60709	-1.82611
12.5	-7.41492	0.759467	-2.62606	-1.82611
13	-7.46607	0.755358	-2.58294	-1.82609
13.5	-7.31172	0.779479	-2.39546	-1.82609
14	-7.284	0.782144	-2.44577	-1.82607
14.5	-7.35634	0.777324	-2.5043	-1.82607
15	-7.28826	0.793944	-2.4283	-1.82605
15.5	-7.46109	0.789531	-1.89294	-1.82605
16	-7.36748	0.790131	-1.79102	-1.82603
16.5	-7.4422	0.788301	-1.75151	-1.826
17	-7.36847	0.812163	-1.64458	-1.826
17.5	-7.28623	0.814422	-1.66352	-1.82598
18	-7.25694	0.822467	-1.52328	-1.82596
18.5	-7.28969	0.823138	-1.56675	-1.82594
19	-7.11886	0.854665	-1.43437	-1.82592
19.5	-7.19283	0.85795	-1.37301	-1.8259
20	-7.10153	0.866682	-1.37098	-1.82587
20.5	-7.26597	0.875357	-0.931723	-1.82585
21	-7.1342	0.909291	-0.808382	-1.82583
21.5	-7.05516	0.907247	-0.75764	-1.82581
22	-6.94928	0.903226	-0.676207	-1.82579
22.5	-7.0092	0.913747	-0.529517	-1.82574
23	-7.01986	0.932531	-0.539768	-1.82572
23.5	-6.91586	0.941327	-0.153081	-1.82568
24	-6.86116	0.938897	-0.143005	-1.82565

Figure 23: Subset of the data used for the suggested methodology

Symbolic Regression was applied to the scores using Eureqa Formulize software. The output from Eureqa consists of series of equations that best fit the relation between X variables and Y variable. This is shown in Figure 24, where

the first column (i.e,size) indicates the complexity measure, second column indicates error measure and the solution column indicates the corresponding candidate equation. Eureqa's complexity metric (or size) is measured both by the number of variables used within the solution as well as the relative weights of each of the building blocks used.

Eureqa Formulize produces a series of equation and ranks each in terms of goodness of fit and are displayed in Figure 24.

Size	Fit	Solution
64	0.219	(Pencillin Concentration.M2.t[1]) = 0.39 (M4.t[1]) + 0.373 (M1.t[5]) + 0.13 (M4.t[1]) sin((M4.t[1]) + si
52	0.225	(Pencillin Concentration.M2.t[1]) = 0.414 (M4.t[1]) + 0.31 (M1.t[5]) + 0.143 (M4.t[1]) sin((M4.t[1]) + :
28	0.234	(Pencillin Concentration.M2.t[1]) = 0.0649 + 0.413 (M1.t[5]) + 0.337 (M4.t[1]) + 0.0582 (M4.t[1]) sin(
17	0.236	(Pencillin Concentration.M2.t[1]) = 0.00769 + 0.396 (M1.t[5]) + 0.361 (M4.t[1]) − 0.0613 (M1.t[4]) −
13	0.245	(Pencillin Concentration.M2.t[1]) = 0.412 (M1.t[5]) + 0.364 (M4.t[1]) − 0.00359 − 0.158 (M4.t[3])
12	0.274	(Pencillin Concentration.M2.t[1]) = 1.98 cos(4.77 + 0.309 (M4.t[1])) − 0.256
10	0.275	(Pencillin Concentration.M2.t[1]) = 2.04 sin(0.289 (M4.t[1])) − 0.196
9	0.308	(Pencillin Concentration.M2.t[1]) = 0.032 + 0.405 (M1.t[5]) + 0.332 (M4.t[1])
7	0.308	(Pencillin Concentration.M2.t[1]) = 0.41 (M1.t[5]) + 0.338 (M4.t[1])
5	0.475	(Pencillin Concentration.M2.t[1]) = 0.0388 + 0.384 (M4.t[1])
3	0.475	(Pencillin Concentration.M2.t[1]) = 0.394 (M4.t[1])

Figure 24: Results obtained from Eureqa Formulize

The mathematical solutions accuracy vs its complexity is shown in Figure 25. The solution with a mean absolute error 0.15 and complexity of 64 as the optimum point on the frontier was considered as optimum solution. After this point, increasing the complexity amount of errors had ignorable discrepancies.

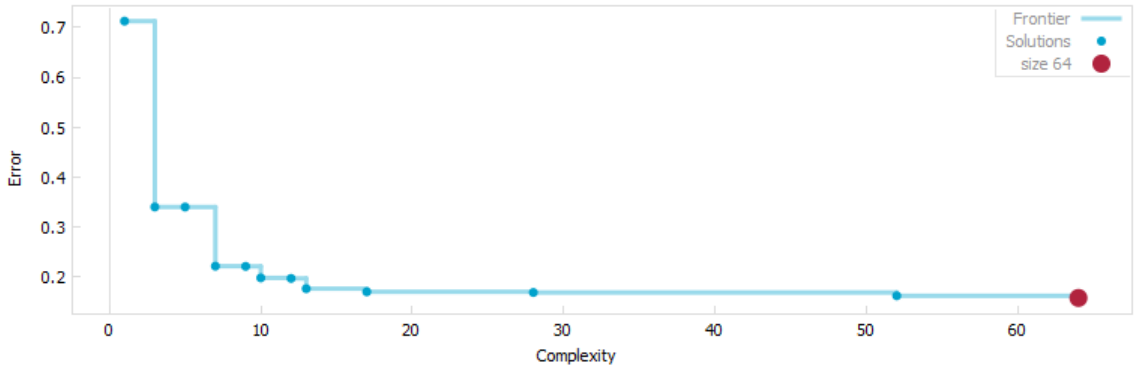


Figure 25: Error as a function of complexity

The stability and percent converged of final solutions after 9.36×10^5 generations were 5.05% and 100%, respectively. Ultimately, based on the optimum point, the proposed mathematical solution can be presented as:

$(PencillinConcentration.M2.t[1]) = 0.389736496840901 * (M4.t[1]) + 0.373132243013773 * (M1.t[5]) + 0.130101662986574 * (M4.t[1]) * sin((M4.t[1]) + sin(0.442743055690672 * (M4.t[1]))) - 0.0890641626334202 - 0.0625878223198603 * (M1.t[5])^2$

This solution is summarized in Table 15

R^2 goodness of fit	0.9603151
Correlation Coefficient	0.98039149
Maximum Error	0.75668024
Mean Squared Error	0.039637283
Mean Absolute Error	0.15593035
Coefficients	6
Complexity	64

Table 15: Summary Parameters of the solution

These solutions are plotted against the data in which both training and validation data are shown. The training data is a subset of the data that is used by Eureqa to search for solutions. The validation data is a second subset that is used only to measure accuracy.

Figure 26 displays the correlation between actual observed values which appear in the training and validation data against the values predicted by the selected model. The plot indicates high correlation between observed and predicted values. Also the solution summary indicates a high R^2 that explains more than 96% of the variation by the model.

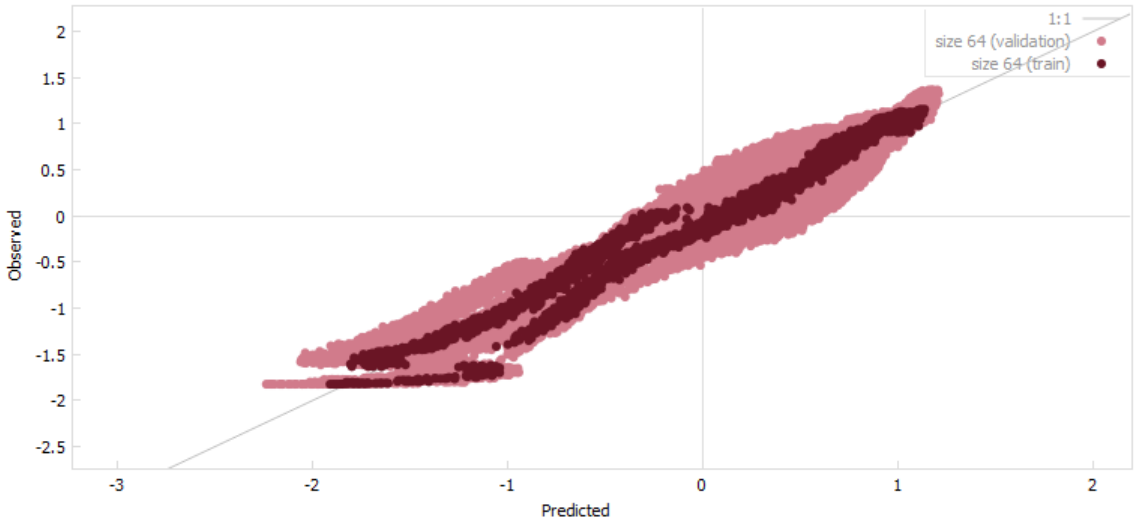


Figure 26: Observed vs predicted data. Both the training and validation data are shown

The size of the error for each point in the training and validation data is plotted as seen in Figure 27. As seen from the plot, the residual error is close to 0 for training and validation data, that indicates that the model used is a more accurate one.

The receiver operating characteristic(ROC) Curve for the data is plotted and shown in Figure 28 and is used for classification models that predict target variables that always have a value of 0 or 1. Area under the ROC curve is a good measure for the model. A larger area under the curve indicates a better model. Unlike the ROC curve from symbolic regression analysis seen in Figure 21, the ROC curve obtained using the proposed model has a larger area under the curve which indicates that this is a better model compared to the model that uses just symbolic regression

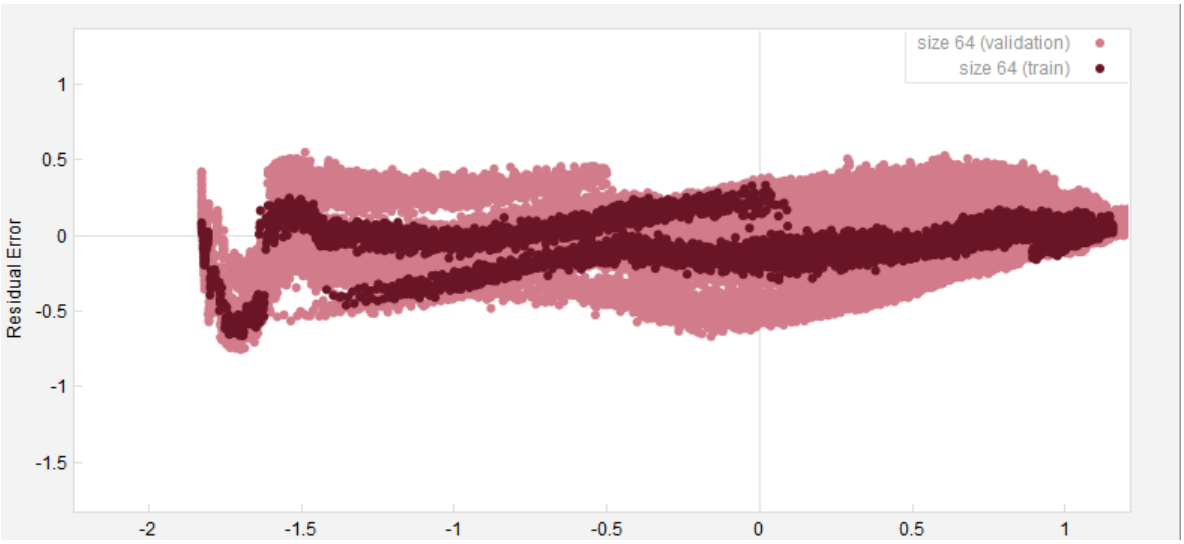


Figure 27: Residual error plot for both training ans validation data

analysis.

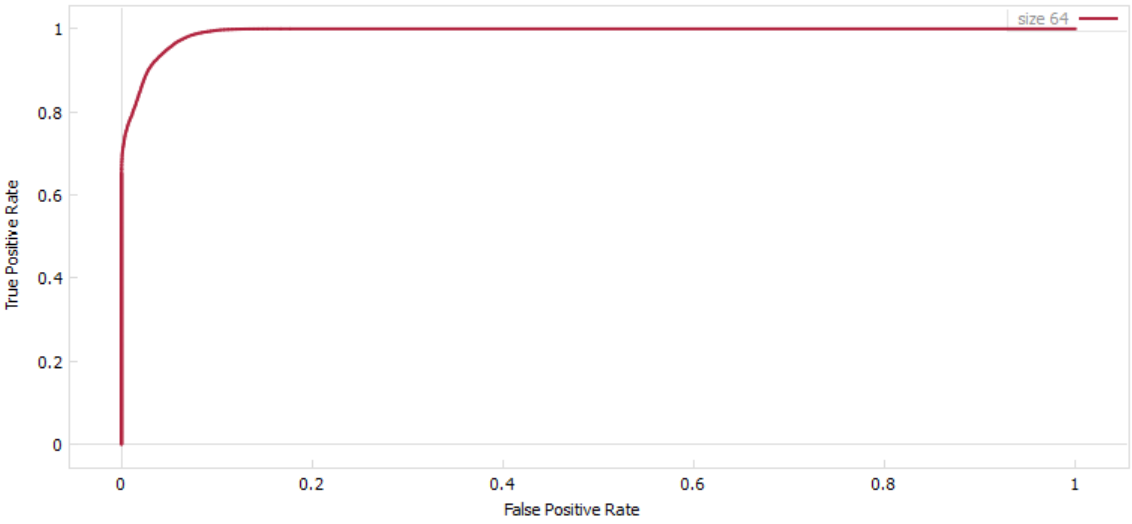


Figure 28: ROC curve for the proposed model.

Table 16 presents a sensitivity analysis over optimum solution. Sensitivity means the relative impact that a predictor has on the target variable (Penicillin Concentration).

From the sensitivity report, it can be deduced that this model is more sensitive to Principal component 1 compared to Principal Component 5. Also, 15% of the

variable	sensitivity	%positive	Positive Magnitude	% Negative	Negative magnitude
(t[1])	0.72092	85%	0.83508	15%	0.095281
(t[5])	0.29591	100 %	0.29591	0%	0

Table 16: Sensitivity Report obtained using the proposed model

time, the principal component 1 will have a negative impact with magnitude of 0.72092 and at 100% of times, Principal component 5 will not have a positive impact on forecasting penicillin concentration at all.

4.2 Summary

Multiple statistical approaches were applied to penicillin manufacturing process and a new methodology was proposed that will improve the accuracy and predictability of the process. A summary of the results obtained for all models is shown in Table 17.

PCA methodology relies on linear assumptions and normality assumptions. A statistic has scale invariance if changing the scale by a certain amount does not change the system, function, or statistics shape or properties. PCA is not scale invariant.

PLS has some limitations that include greater difficulty of interpreting the loadings of the independent latent variables and significance of parameter estimates cannot be estimated, since the distributional properties of estimates are not known.

Since the search space for symbolic regression can be large, SR algorithms might take much longer time to find a suitable model and parametrization compared to traditional regression methods. SR also need huge computational power.

SVR has high algorithmic complexity, extensive memory requirements and requires choosing appropriately hyper parameters that will allow for sufficient generalization performance. Also, the design for multi-class SVR classifiers is complex.

Table 17 indicates that other approaches like SVR and Symbolic regression might be better than SBSR. For instance, the mean absolute errors are smaller for both SR and SVR compared to SBSR. However these methods have several limitations as discussed earlier. The proposed approach has multiple advantages, which offset the limitations. Some of the several advantages of SBSR are discussed below:

Symbolic Regression does not dependent on linearity assumptions. The algorithms search for best model to fit the data including linear and nonlinear models. Compared to PCA and PLS, results obtained from symbolic regression are easy to interpret.

The other significant advantage of Scores based symbolic regression is that the new variables are orthogonal (variables used are scores compared to actual variables) and collinearity is no longer a concern, which is typically a problem in the case of continuous manufacturing processes. Also, noise remains in the residuals, since few principal components represent most of the variance. The regression coefficients are more stable, since the eigen vectors are orthogonal. Since the scores are already limited to first few principal components, applying symbolic regression is computationally faster and will provide quicker results.Inversion of the matrix is simple, similar to the situation where original variables are correlated.

Models	R^2	**Mean absolute error**	**RMSE**
Principal Component Analysis(PCA)	0.868	N/A	N/A
Partial Least Squares(PLS)	0.805	N/A	0.075
Symbolic Regression(SR)	0.99	0.033	0.045
Support Vector Regression(SVR)	0.99	0.045	0.05
Scores Based Symbolic Regression(SBSR)	0.96	0.156	0.19

Table 17: Summary of different models

4.3 Analytical Method Transfer

Analytical method transfer is a process that involves transferring analytical method from a sending lab to receiving lab to prove that a method is executed similar at both labs. Analytical method transfer is a key component of technology transfer between process development laboratory to commercial manufacturing. Regulatory agencies require that transfer between labs are documented and equivalence is established. The objective of the method transfer is to ensure that results obtained by receiving lab are reliable and comparable to sending lab, to meet the specifications of a product and address any bias between the labs. Two one sided t-test(TOST) is used to establish equivalence of the results obtained by sending and receiving labs. The objective of the study was to study of the effect of sample size for establishing equivalency and comparison of methods during assay transfers and to recommend criterion for out of specification risk mitigation.

4.3.1 Power and Sample size

An important step of TOST is determining the desired power and the corresponding sample size needed for conducting the test. The power of TOST is defined as the probability of correctly accepting equivalence at a given true mean difference between two populations. This may be alternatively stated in terms of the type-II error β of falsely rejecting equivalence at a given true mean difference between two populations.

Larger sample sizes give more power to the test. In order to achieve the desired power, the minimum number of test results required is calculated. Four parameters need to be specified in this calculation:

1. The desired power $1 - \beta$

2. The equivalence limit E

3. population standard deviation σ

4. the true difference parameter Δ

The desired power is usually set to be 90%. Due to the cost of replicates, the desired power can be decreased but should not be lower than 80%. The equivalence limit E is set based on scientific knowledge. The population standard deviation represents the precision of the test method. In practice, σ is unknown most of the time. The estimated standard deviation S, calculated by the long-term method performance data, can be used to estimate the population standard deviation σ. The true difference parameter Δ is a hypothetical value such that if the absolute value of the observed difference is no more than Δ, there is a strong probability of concluding that the two data sets represent equivalent results. The choice of Δ is arbitrary.

The following inequality is used to determine the minimum number of test results n to maintain a desired level of power $1 - \beta$ of the TOST:

$$\phi(\frac{E - \Delta}{\sigma_D} - Z_{1-\alpha}) - \phi(\frac{-E - \Delta}{\sigma_D} + Z_{1-\alpha}) \geq 1 - \beta \tag{21}$$

where

- $\phi(.)$ is the standard normal cumulative distribution function,
- Δ is the true difference parameter $\mu_1 - \mu_2$,
- σ_D is the standard error of the difference D, estimated by $S \times \sqrt{\frac{2}{n}}$
- $Z_{1-\alpha}$ is the $(1 - \alpha)th$ percentile of the standard normal distribution

The minimum number of n which satisfies the inequality above is defined as the minimum number of independent replicates in each of the two populations. In other words, at least n results for population 1 and n results for population 2 ($2n$

	Transfer accepted	Transfer not accepted
Acceptable	Right Decision	Rejecting an effective transfer (Producer's risk)-Type I Error
Not Acceptable	Approving ineffective transfer (Consumer's Risk)-Type II Error	Right Decision

Table 18: Hypothesis testing for Analytical Transfer

in total) are required to conduct an equivalence test between two means with the desired power $1 - \beta$. It is not required but highly recommended that the sample sizes from two populations are equivalent.

Table 18 describes consumer and producers risk for analytical method transfers.

Power and Sample size calculation Results

The variation of sample size as a function of power was studied using equivalence limit, standard deviation and the true difference for the analytical method. The data collected for method transfer is described in Chapter 3. Based upon historical data, the parameters for Analytical Method A for sample size calculation are:

- s = 4%, the estimated standard deviation based on the long-term method performance data
- $\alpha = 0.05$, controlled type-I error which provides a 95% level of confidence
- E = 8%, the equivalence limit
- $\Delta = 4\%$, true mean difference
- $\beta = 0.10$, controlled Type-II error which provides the power which is probability $1 - \beta$ of properly accepting equivalence at a given value of Δ.

The results obtained for sample size for Analytical method A with $\alpha = 0.05$ and power=0.9 are shown in Figure 29.

Power and Sample Size

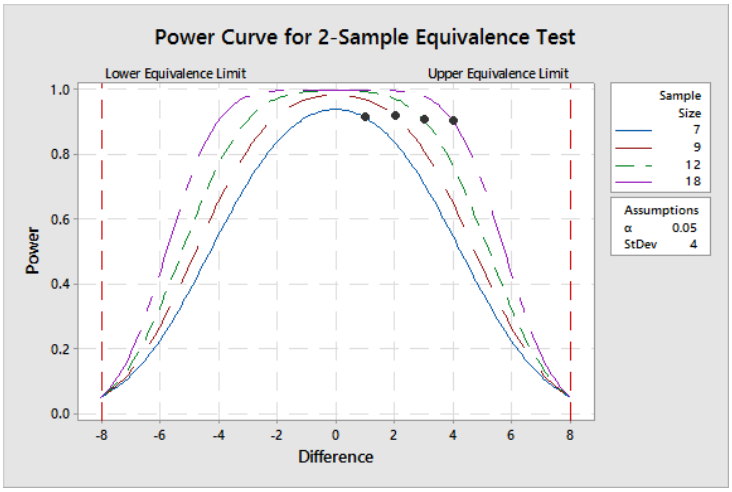
2-Sample Equivalence Test

Power for difference: Test mean - reference mean
Null hypothesis: Difference ≤ -8 or Difference ≥ 8
Alternative hypothesis: -8 < Difference < 8
α level: 0.05
Assumed standard deviation: 4

Difference	Sample Size	Target Power	Actual Power
1	7	0.9	0.914203
2	9	0.9	0.918678
3	12	0.9	0.906710
4	18	0.9	0.902272

The sample size is for each group.

(a) Power and sample size for $\alpha = 0.05$ and power 0.9



(b) Power Curve for $\alpha = 0.05$ and power 0.9

Figure 29: Summary of the sample size results

The analysis shows that, if the difference is 1, 7 observations are required in each group to achieve a power of 0.9. A sample size of 7 gives a power of approximately 0.91. If the difference is closer to lower equivalence limit or the upper equivalence limit (-8 or 8), more observations are needed to achieve the same power. For example, if the difference is 4, at least 18 observations are needed in each group to achieve a power of 0.9. For any sample size, the power of the test decreases and approaches a constant value for all the sample sizes, as the difference approaches the lower equivalence limit or the upper equivalence limit.

Additional analysis was performed for varying the difference to detect and the power to detect differences is calculated at $\alpha = 0.05$ and summarized below in Figure 30.

Power and Sample Size

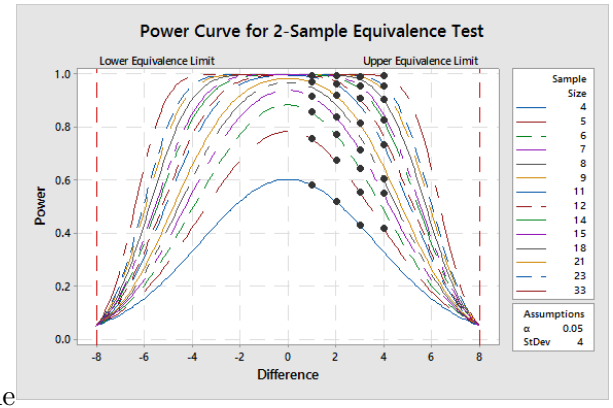
2-Sample Equivalence Test

```
Power for difference:       Test mean - reference mean
Null hypothesis:            Difference ≤ -8 or Difference ≥ 8
Alternative hypothesis:     -8 < Difference < 8
α level:                    0.05
Assumed standard deviation: 4
```

Difference	Sample Size	Target Power	Actual Power
1	4	0.40	0.580924
1	4	0.50	0.580924
1	5	0.60	0.754232
1	5	0.70	0.754232
1	6	0.80	0.855094
1	7	0.90	0.914203
1	9	0.95	0.969773
1	12	0.99	0.993765
2	4	0.40	0.519053
2	4	0.50	0.519053
2	5	0.60	0.672703
2	6	0.70	0.770967
2	7	0.80	0.837877
2	9	0.90	0.918678
2	11	0.95	0.959941
2	15	0.99	0.990888
3	4	0.40	0.429700
3	5	0.50	0.553408
3	6	0.60	0.641719
3	7	0.70	0.711158
3	9	0.80	0.813790
3	12	0.90	0.906710
3	15	0.95	0.954863
3	21	0.99	0.990249
4	5	0.40	0.417487
4	7	0.50	0.547303
4	8	0.60	0.601464
4	11	0.70	0.732327
4	14	0.80	0.824086
4	18	0.90	0.902272
4	23	0.95	0.954817
4	33	0.99	0.991195

The sample size is for each group.

(a) Power and sample size results using the 2 sample equivalence test



(b) Power curve

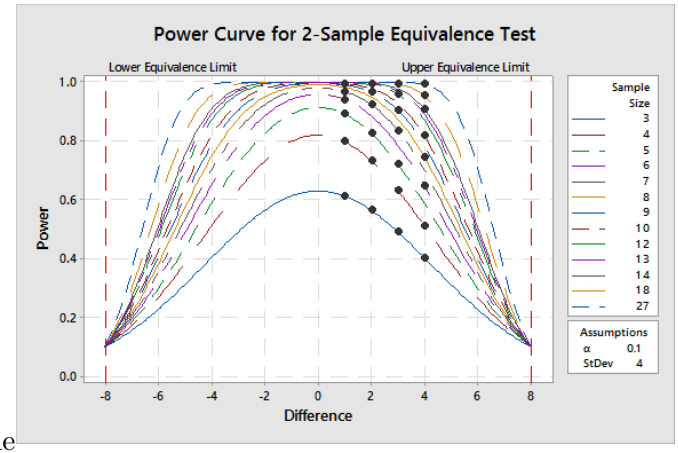
Figure 30: Summary of the analysis for $\alpha = 0.05$

Additional analysis was performed for varying the difference to detect and the power to detect differences is calculated at $\alpha = 0.1$ and summarized below in Figure 31

Power and Sample Size

```
2-Sample Equivalence Test
Power for difference:       Test mean - reference mean
Null hypothesis:            Difference ≤ -8 or Difference ≥ 8
Alternative hypothesis:     -8 < Difference < 8
α level:                    0.1
Assumed standard deviation: 4
            Sample   Target
Difference    Size   Power     Actual Power
         1       3    0.40      0.611901
         1       3    0.50      0.611901
         1       3    0.60      0.611901
         1       4    0.70      0.796378
         1       5    0.80      0.890294
         1       6    0.90      0.939966
         1       7    0.95      0.966833
         1      10    0.99      0.994294
         2       3    0.40      0.562874
         2       3    0.50      0.562874
         2       4    0.60      0.730575
         2       4    0.70      0.730575
         2       5    0.80      0.825015
         2       7    0.90      0.922260
         2       9    0.95      0.965251
         2      13    0.99      0.993344
         3       3    0.40      0.489313
         3       4    0.50      0.630377
         3       4    0.60      0.630377
         3       5    0.70      0.719925
         3       7    0.80      0.834954
         3       9    0.90      0.903212
         3      12    0.95      0.957441
         3      18    0.99      0.992287
         4       3    0.40      0.401493
         4       4    0.50      0.509095
         4       6    0.60      0.647029
         4       8    0.70      0.744881
         4      10    0.80      0.816594
         4      14    0.90      0.906862
         4      18    0.95      0.953681
         4      27    0.99      0.990943

The sample size is for each group.
```

(a) Power and sample size results using the 2 sample equivalence test



(b) Power curve

Figure 31: Summary of the analysis for $\alpha = 0.1$

Additional analysis to solve for sample size was performed for the following scenarios at $\alpha = 0.15$:

- Power: 0.8 , 0.85, 0.9 , 0.95 , 0.99
- Alpha: 0.15
- Group Allocation: Equal (N1 = N2)
- (Upper Equivalence Limit)|EU|: 4 , 8
- (Lower Equivalence Limit)|EL| : -Upper Limit
- D (True Difference): 0, 2, 4, 8
- SD (Standard Deviation): 2, 4

Figure 32 below shows the results for sample size at different power and Equivalence limits and standard deviation levels.

The columns shown in the figure are defined as follows:

- Target Power is the desired power value (or values) entered in the procedure. Power is the probability of rejecting a false null hypothesis.
- Actual Power is the power obtained in this scenario. Because N1 and N2 are discrete, this value is often (slightly) larger than the target power.
- N1 and N2 are the number of items sampled from each population.
- N is the total sample size, N1 + N2.
- D is the true difference between the means.
- SD is the within-group standard deviation for the two groups.
- α = 0.05/0.1/0.15 controlled type-I error which provides a 95%/90%/85% level of confidence

Target Power	Actual Power	N1	N2	N= N1+N2	D(True Difference)	SD	Lower Equiv.limit	Upper Equiv.Limit	Alpha
0.8	0.884057	6	6	12	0	4	-8	8	0.05
0.85	0.884057	6	6	12	0	4	-8	8	0.05
0.9	0.939163	7	7	14	0	4	-8	8	0.05
0.95	0.968601	8	8	16	0	4	-8	8	0.05
0.99	0.99198	10	10	20	0	4	-8	8	0.05
0.8	0.837877	7	7	14	2	4	-8	8	0.05
0.85	0.885016	8	8	16	2	4	-8	8	0.05
0.9	0.918678	9	9	18	2	4	-8	8	0.05
0.95	0.959941	11	11	22	2	4	-8	8	0.05
0.99	0.990888	15	15	30	2	4	-8	8	0.05
0.8	0.824086	14	14	28	4	4	-8	8	0.05
0.85	0.868403	16	16	32	4	4	-8	8	0.05
0.9	0.902272	18	18	36	4	4	-8	8	0.05
0.8	0.819273	4	4	8	0	4	-8	8	0.1
0.85	0.912243	5	5	10	0	4	-8	8	0.1
0.9	0.912243	5	5	10	0	4	-8	8	0.1
0.95	0.957797	6	6	12	0	4	-8	8	0.1
0.99	0.990538	8	8	16	0	4	-8	8	0.1
0.8	0.825015	5	5	10	2	4	-8	8	0.1
0.85	0.8838	6	6	12	2	4	-8	8	0.1
0.9	0.92226	7	7	14	2	4	-8	8	0.1
0.95	0.965251	9	9	18	2	4	-8	8	0.1
0.99	0.993344	13	13	26	2	4	-8	8	0.1
0.8	0.816594	10	10	20	4	4	-8	8	0.1
0.85	0.868931	12	12	24	4	4	-8	8	0.1
0.9	0.906862	14	14	28	4	4	-8	8	0.1
0.95	0.953681	18	18	36	4	4	-8	8	0.1
0.99	0.990943	27	27	54	4	4	-8	8	0.1
0.8	0.90206	4	4	8	0	4	-8	8	0.15
0.85	0.90206	4	4	8	0	4	-8	8	0.15
0.9	0.90206	4	4	8	0	4	-8	8	0.15
0.95	0.9558	5	5	10	0	4	-8	8	0.15
0.99	0.99112	7	7	14	0	4	-8	8	0.15
0.8	0.82742	4	4	8	2	4	-8	8	0.15
0.85	0.89184	5	5	10	2	4	-8	8	0.15
0.9	0.93094	6	6	12	2	4	-8	8	0.15
0.95	0.95564	7	7	14	2	4	-8	8	0.15
0.99	0.9925	11	11	22	2	4	-8	8	0.15
0.8	0.82231	8	8	16	4	4	-8	8	0.15
0.85	0.85278	9	9	18	4	4	-8	8	0.15
0.9	0.9166	12	12	24	4	4	-8	8	0.15

Figure 32: Results of sample sizes for varying power, equivalence limits and standard deviation levels.

The lower and upper equivalence limits are the maximum allowable differences that still result in equivalence.

Figure 33 shows the variation of N1 with power for varying difference in means. As seen from the figure, the sample size requirement increases as the power increases. A power of 0.9 is an optimum balance for mitigating consumer's risk.

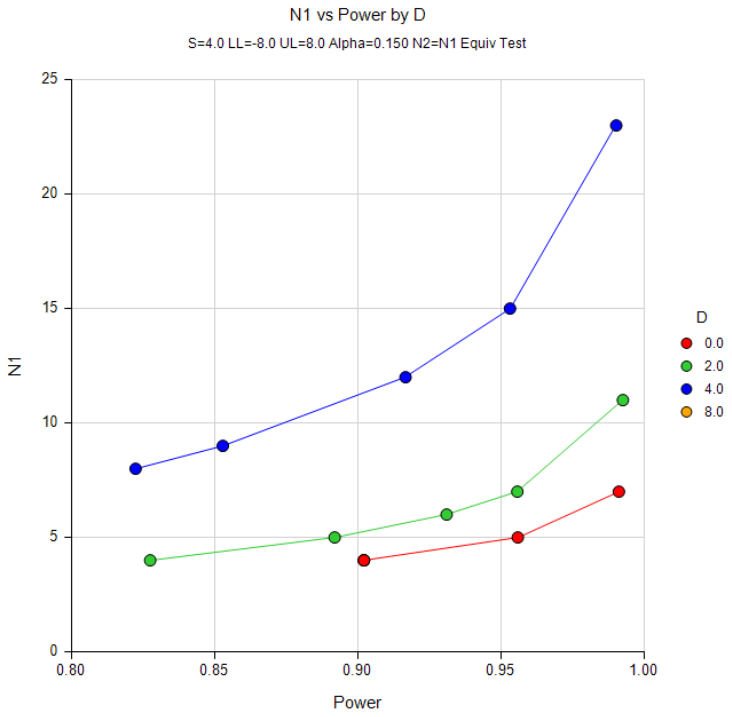


Figure 33: N1 vs power for varying means

A contour plot is shown below for target power vs sample size. The darker regions identify higher alpha risk. The contour levels reveal a peak centered in the vicinity of sample size of 30 and power of 0.95. alpha risk in this peak region is greater than 0.15. Similarly, for alpha risk at 0.05, to achieve a power of 0.9, a minimum sample size of 36 is required.

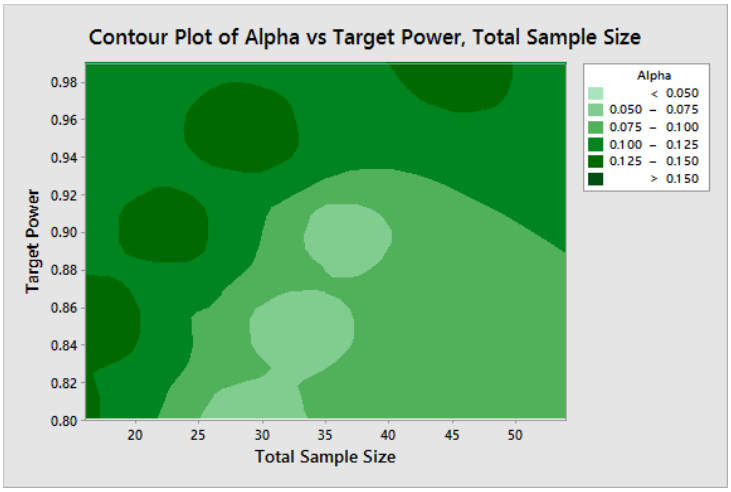


Figure 34: The contour plot of power as a function of sample size (N1 + N2) for varying alpha.

An equivalence test of means using two one-sided tests on data from a parallel-group design with sample sizes of 18 in the sending lab and 18 in the receiving lab achieves 90% power at a 5% significance level when the true difference between the means is 4.0, the standard deviation is 4.0, and the equivalence limits are -8.0 and 8.0.

Based upon the analysis above, it can be deduced that increasing sample size, widening equivalence margin, or reducing variability result in higher power, as expected. The power of the test does NOT depend on where the center of the distribution is, but is a function of the true difference between the group means. The highest power was obtained when $\Delta = 0$. When $\Delta = 0$ and the goal post was set to one standard deviation, i.e., $E = s$, the power of the test is simplified to a function that is only dependent on sample size. When $\Delta = 0$, E and s cancel each other from the formula. It can therefore be concluded from the table in Figure 32, that when $2s$ is used as the equivalence limit and there is a true difference in means of $1s$ roughly 36 total samples are needed (18 each group if equal size) for $> 90\%$ power. When $\Delta \neq 0$, E and s do not cancel each other in the formula, and

hence, the difference powers for the same sample size depend on the values of E and s.

For a fixed sample size, the power of the test is maximized at $\Delta = 0$, but may not reach 100%. Type I error is maximized at $\Delta = E$, and eventually goes to zero as $\Delta \to \infty$.

4.3.2 Analytical Method Transfer Results

18 experiments at each lab were conducted to test for Equivalence. The data from the experiments are tabulated below:

Replicate	**Sending Lab**	**Receiving Lab**
1	94.01	93.99
2	95.61	100.1
3	96.43	106.2
4	98.62	99.6
5	102.45	103.46
6	99.84	106.36
7	104.76	105.21
8	100.46	99.01
9	97.49	106.23
10	100.06	100.77
11	92.84	95.78
12	94.56	98.16
13	103.33	101.05
14	98.5	103.52
15	109.96	110.58
16	105.46	108.45
17	106.12	112.79
18	107.92	104.25

Table 19: Data for all the 18 experiments performed using Analytical Method Transfer

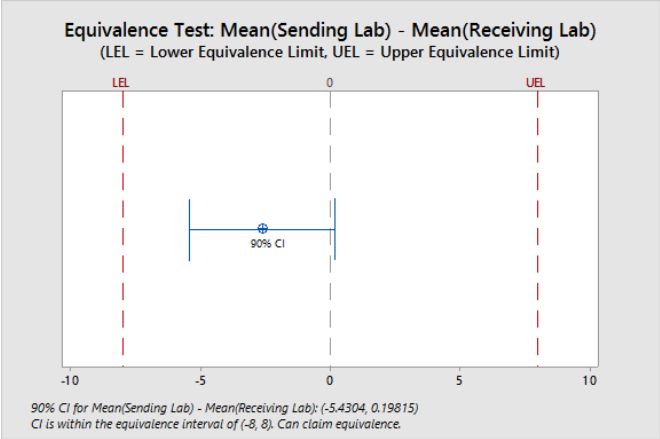
The results of the test are shown in Figure 35. As seen in Figure 35b, the 90% confidence interval of mean difference (-5.4304, 0.19815) is well within the equivalence margin (-8 ,8). Also, the greater of the two p-values is 0.001, which is

less than the chosen $\alpha(0.05)$. These two results claim equivalence.

Results from TOST:

```
Method

Test mean = mean of Sending Lab
Reference mean = mean of Receiving Lab
Equal variances were assumed for the analysis.


Descriptive Statistics

Variable        N    Mean   StDev SE Mean
Sending Lab    18  100.47  4.9968   1.1778
Receiving Lab  18  103.08  4.9892   1.1760

Pooled StDev = 4.99300


Difference: Mean(Sending Lab) - Mean(Receiving Lab)

Difference      SE        90% CI        Equivalence Interval
   -2.6161  1.6643  (-5.4304, 0.19815)         (-8, 8)

CI is within the equivalence interval. Can claim equivalence.


Test

Null hypothesis:         Difference ≤ -8 or Difference ≥ 8
Alternative hypothesis:  -8 < Difference < 8
α level:                 0.05

Null Hypothesis  DF  T-Value  P-Value
Difference ≤ -8  34   3.2349    0.001
Difference ≥ 8   34  -6.3786    0.000

The greater of the two P-Values is 0.001. Can claim equivalence.
```

(a) Statistical parameters obtained from TOST



(b) 90% confidence interval of both the labs.

Figure 35: Result from Analytical Method Transfer

4.4 Analytical method transfer using Lin's CCC

The analysis below captures the power for several sample sizes of a study designed to compare sending lab results to receiving lab based upon CCC. The power was captured for several sample sizes of a study designed to compare sending lab results to receiving lab. The minimum value of ρ that can be tolerated is 0.8. The power was computed at ρ_1 = 0.91,0.99, 0.999. The analysis is summarized and shown in Figure 36 below:

The columns in the tables are defined as follows:

- X is the Sending lab results
- Y is the Receiving lab results
- Power is the probability of rejecting a false null hypothesis.
- Sample Size (n) is the number of measurement pairs (Sending - Receiving) in the study.
- CCC0 is the value of Lin's concordance correlation coefficient assuming H0. CCC0 serves as a lower bound on acceptable values of CCC.
- CCC1 is the value of Lin's concordance correlation coefficient assuming H1. It is the value at which the power is calculated.
- ρ0 is correlation between the new measurement (Y) and the gold-standard measurement (X) assuming H0. It serves as a lower bound on acceptable values of ρ
- ρ1 is correlation between Y and X assuming H1. It is the value at which the power is calculated

Power	Sample Size (n)	Lin's CCC |H0 (CCC0)	Lin's CCC |H1 (CCC1)	ρ(Y,X) |H0 (ρ0)	ρ(Y,X) |H1 (ρ1)
0.811	22	0.8	0.91	0.8	0.91
0.8565	26	0.8	0.91	0.8	0.91
0.9054	32	0.8	0.91	0.8	0.91
0.9532	42	0.8	0.91	0.8	0.91
0.9904	64	0.8	0.91	0.8	0.91
0.8755	4	0.8	0.99	0.8	0.99
0.8755	4	0.8	0.99	0.8	0.99
0.95	5	0.8	0.99	0.8	0.99
0.9803	6	0.8	0.99	0.8	0.99
0.9923	7	0.8	0.99	0.8	0.99
0.9521	3	0.8	0.999	0.8	0.999
0.9521	3	0.8	0.999	0.8	0.999
0.9521	3	0.8	0.999	0.8	0.999
0.9521	3	0.8	0.999	0.8	0.999
0.9973	4	0.8	0.999	0.8	0.999
0.8002	1156	0.9	0.91	0.9	0.91
0.85	1407	0.9	0.91	0.9	0.91
0.9	1759	0.9	0.91	0.9	0.91
0.95	2353	0.9	0.91	0.9	0.91
0.99	3700	0.9	0.91	0.9	0.91
0.8408	5	0.9	0.99	0.9	0.99
0.9053	6	0.9	0.99	0.9	0.99
0.9053	6	0.9	0.99	0.9	0.99
0.9671	8	0.9	0.99	0.9	0.99
0.9936	11	0.9	0.99	0.9	0.99
0.9017	3	0.9	0.999	0.9	0.999
0.9017	3	0.9	0.999	0.9	0.999
0.9017	3	0.9	0.999	0.9	0.999
0.988	4	0.9	0.999	0.9	0.999
0.9986	5	0.9	0.999	0.9	0.999

Figure 36: CCC Analysis

The results of the analysis is shown in Figure 37. As seen from the figure, the sample size increases when H0 ,H1 and $\rho 0$ and $\rho 1$ are closer to each other. Sample size increases as more power is required at different levels of CCC0 and CCC1.

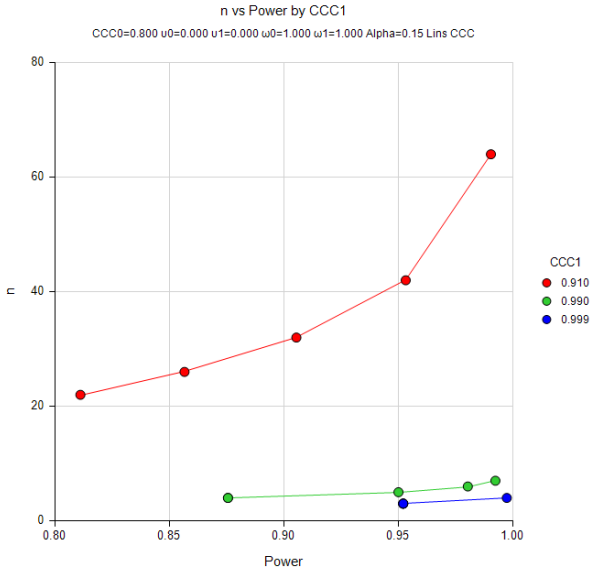


Figure 37: Plot of n vs power for CCC0 = 0.8

The above plots show that using Lin's concordance correlation coefficient to compare the sending lab to receiving lab results, a minimum sample of 32 subjects results in 0.9110 power to determine whether the new method is similar to the sending lab method. The statistical test uses a one-sided z test with a 0.15 significance level.

4.4.1 Analytical Method Transfer Results based on CCC

The concordance correlation coefficient ρ_c [31, 33] evaluates the degree to which pairs of observations fall on the 45° line through the origin. It contains a measurement of precision ρ and accuracy C_b:

$\rho_c = \rho C_b$

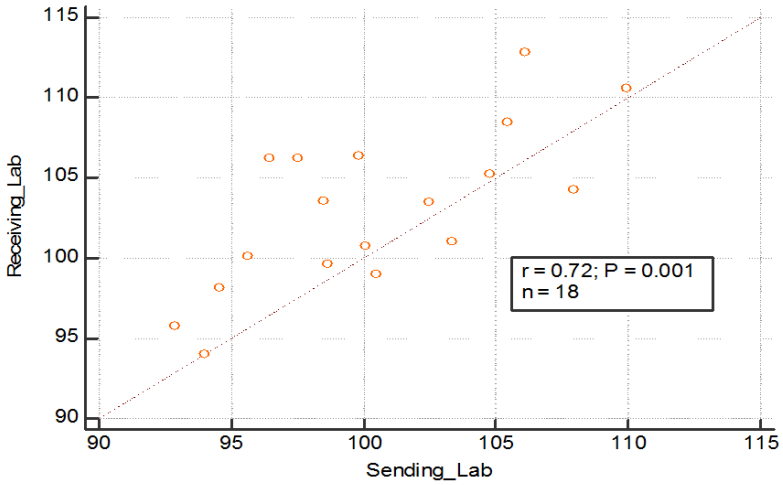
where

- ρ is the Pearson correlation coefficient, which measures how far each obser-

vation deviates from the best-fit line, and is a measure of precision, and

- C_b is a bias correction factor that measures how far the best-fit line deviates from the 45° line through the origin, and is a measure of accuracy

Figure 38 shows a plot of receiving lab vs sending lab for 18 samples, and the statistical parameters for these samples. (ρ) is the Pearson correlation coefficient, which measures how far each observation deviates from the best-fit line, and is a measure of precision.



Variable Y	Receiving_Lab
Variable X	Sending_Lab

Sample size	18
Concordance correlation coefficient	0.6306
95% Confidence interval	0.3013 to 0.8255
Pearson ρ (precision)	0.7223
Bias correction factor C_b (accuracy)	0.8731

Figure 38: Results of CCC Analysis

There is no literature yet, giving a descriptive scale for the degree of agreement for Lins Concordance Correlation coefficient, but McBride [79] suggests the following descriptive scale for values of the concordance correlation coefficient:

Based upon the above recommendations, the level of agreement of 0.6306 indicates

Value of ρ_c	Strength of agreement
< 0.90	Poor
0.90- 0.95	Moderate
0.95 - 0.99	Substantial
0.95 - 0.99	Almost Perfect

poor agreement between the sending and receiving labs.

4.4.2 Application of Bootstrapped Equivalence Test to Analytical Method Transfer

18 experiments at each lab were conducted to test for Equivalence. The data collected was tabulated in Table 19. Suggested methodology was applied to the data collected and summary of results are included below. Specifications of the parameter are 85-125.

Step 1: C_{pk} calculation of sending lab

C_{pk} has been calculated using Equation 18. This is illustrated in Figure 39

A C_{pk} calculation indicates a value of 1.48 based upon the specifications limit range of 85-125. Hence, the analytical method transfer is initiated.

Step 2: Calculation of Equivalence Margin

The equivalence margin has been calculated using Equation 15. The values of $n = 18$, $\alpha = 0.05$ and $\sigma = 5$ were used in the equation that results an Equivalence margin value of **5.8**.

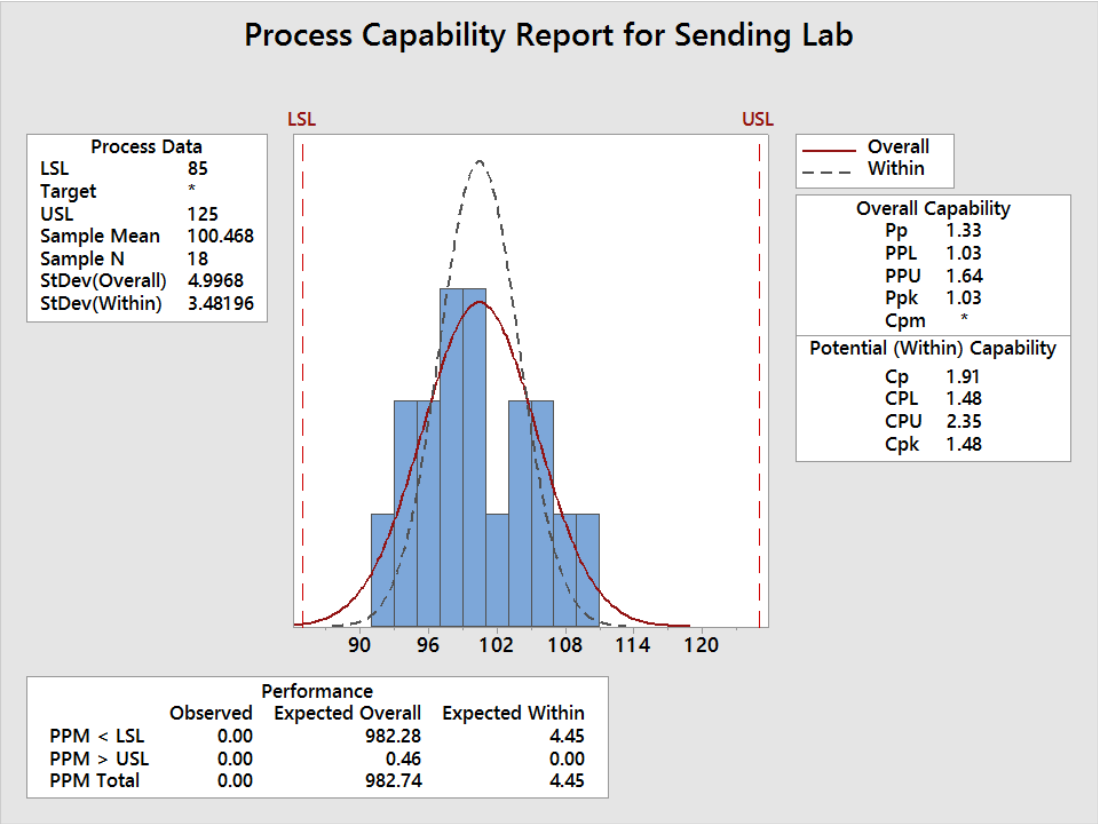


Figure 39: Process capability of sending lab

Step 3: Perform Bootstrapped Equivalence test based upon proposed equivalence margin

(a) The proposed equivalence margin is (-5.8, 5.8)

(b) Simulate 10,000 sets of datasets of same size from sending and receiving lab using bootstrapping (random selection with replacement)

(c) Confidence interval of the mean difference based upon 10000 resamples is calculated. Confidence interval = (-5.1973056, 0.1033611)

(d) Since the confidence interval is completely contained within the equivalence limits (-5.8,5.8), Equivalence can be claimed.

Figure 40 below displays the histogram of mean differences for 10,000 resamples based upon bootstrapping.

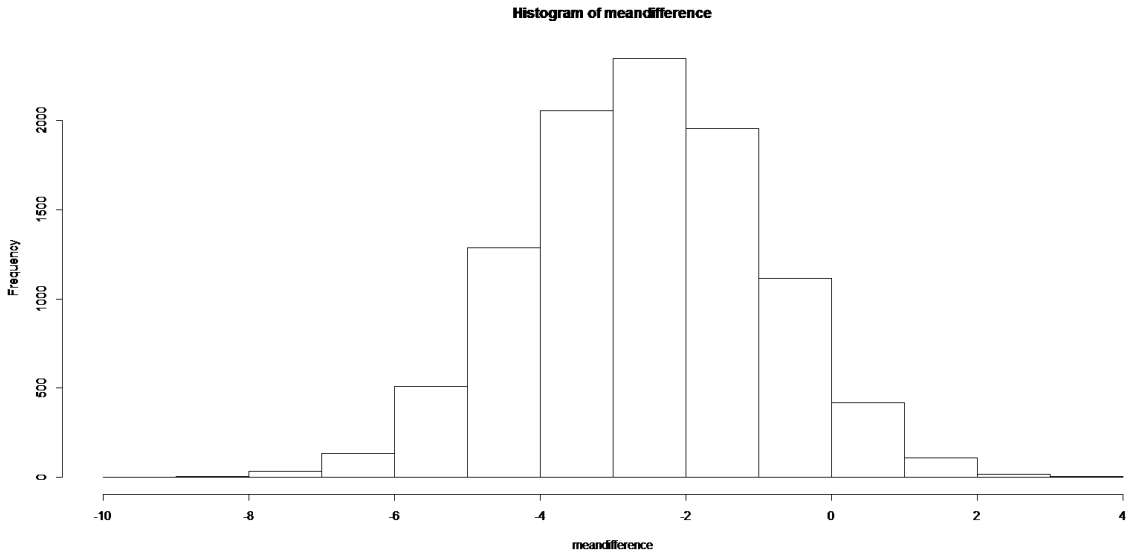


Figure 40: Histogram of mean differences

Step 4: Calculate Process capability using revised mean to control risk of Out of Specification(OOS)

The process mean is revised to include the potential for bias by adding the equivalence margin to process mean and process capability is calculated based upon the revised mean.

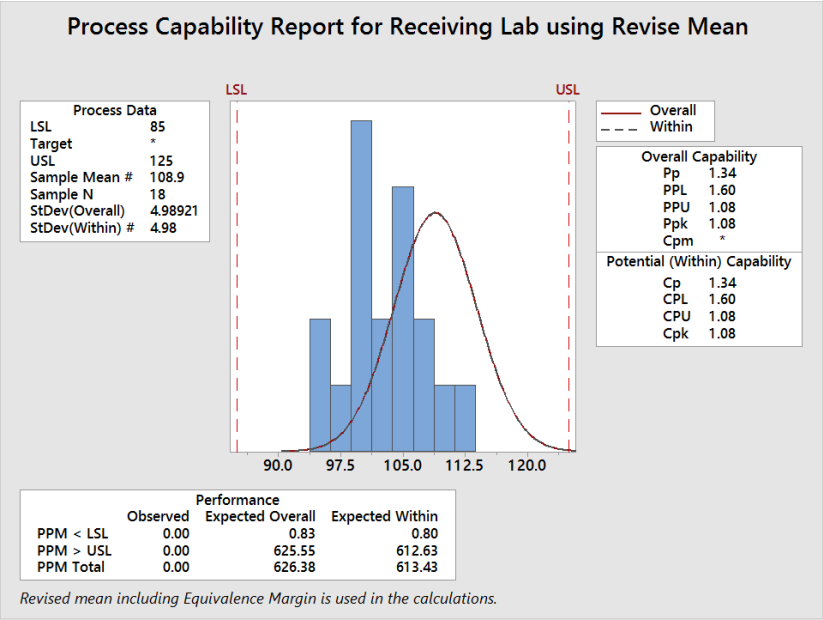


Figure 41: Process Capability of Receiving lab using the revised mean

The C_{pk} calculation indicates a value of 1.08 based upon specification limit 85-125. Hence the analytical method transfer can be claimed successful.

4.4.3 Summary of the current study using Analytical Method Transfer

The study analyzed the relation between sample size, power, and difference in means to detect at multiple levels of Type I error and recommended optimum samples size for the study with atleast 90% power to detect the differences. Two different statistical methodologies that are used for analytical method transfers were compared and an extension of Equivalence test using bootstrapping was proposed to optimize producer risk and consumer risk while controlling the risk for out of specification results. In a traditional two one-sided t test, confidence interval of the mean difference is based upon single point estimate. Applying bootstrap-

ping techniques provides better understanding of population mean differences and provide a robust estimate for confidence interval. Equivalence between sending and receiving labs can be claimed with more confidence. Also using the equivalence margin for process capability computations provide better estimate of out of specification risk and can be mitigated.

Below is a summary of the approaches used in the current study from a producer/consumer risk perspective:

Statistical Approach	Producer Risk (Type I Error)	Consumer's Risk (Type II Error)	Risk of OOS
Equivalence	Sample Size dependent	Sample Size dependent	Not Controlled
Concordance Correlation Coefficient	Not Controlled	Not Controlled	Not Controlled
Proposed approach	Sample Size dependent	Sample Size dependent	Controlled

Figure 42: Summary from a producer/consumer risk persepctive

As seen from Figure 42, the proposed approach is the only one that addresses both sample size and the risk of Out of specification

4.5 Stability Data Analysis

Stability studies are an essential component of biopharma manufacturing, allowing evaluation of active pharmaceutical ingredient (API) stability or drug product stability under the influence of a variety of environmental factors such as temperature, humidity, and light. Data analysis from these evaluations will provide guidance for recommended storage conditions, retest intervals and shelf lives to be established. Ruberg and Stegeman [49] and Ruberg and Hsu [50] discuss methods based on multiple comparison for poolability of slopes and batch degradation. The

stability data is taken from Ruberg and Stegeman [49] and is included in Figure 7. There are six batches in this data and concentration for the batches is measured over years.

It is desirable to pool the data from different batches to set a single shelf life for all the batches. The batches can only be pooled if there are no statistical significant differences between slopes and intercepts. Current Food and Drug Administration (FDA) guideline recommends using $\alpha = 0.25$ for testing the poolability of batches. The scenarios below are representative of a dataset where slopes can be pooled.

- **Scenario 1 : Considering batch as fixed factor and α=0.25 (Poolable slopes)**

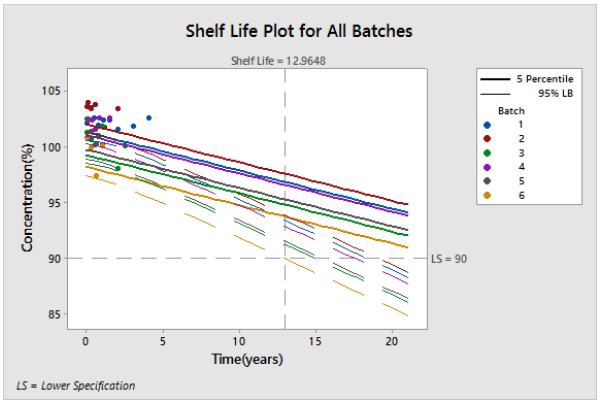
  Figure 43 shows the results obtained considering batch as a fixed factor.

  The p-value based upon the analysis above for model selection with $\alpha = 0.025$ for the Time by Batch interaction (Time*Batch) is 0.359. Because the p-value is greater than the α-level of 0.25, the model can be reduced. Both Time and Batch are significant. Thus, the regression equations for each batch have different intercepts and common slopes. Batch 6 has the smallest intercept, 99.723, which indicates that this batch had the lowest concentration at time zero. The shortest shelf life estimate is 12.965 years, so the overall shelf life for the product is estimated as 12.965 years.

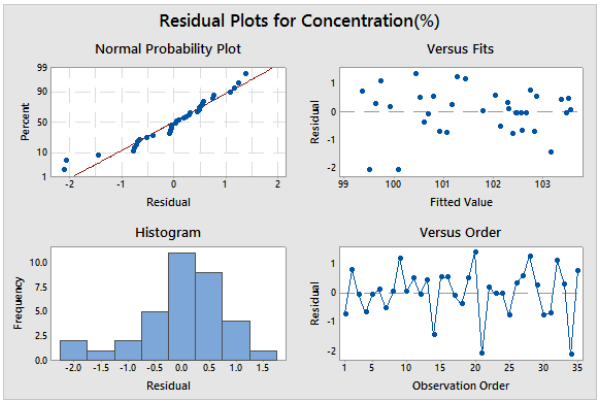
```
Factor Information

Factor   Type       Number of Levels   Levels
Batch    Fixed                 6       1, 2, 3, 4, 5, 6

Model Selection with α = 0.25
Source             DF   Seq SS   Seq MS  F-Value  P-Value
Time(years)         1   0.8304   0.8304     1.04    0.317
Batch               5  54.5519  10.9104    13.73    0.000
Time(years)*Batch   5   4.6039   0.9208     1.16    0.359
Error              23  18.2777   0.7947
Total              34  78.2640

Source       DF   Seq SS   Seq MS  F-Value  P-Value
Time(years)   1   0.8304   0.8304     1.02    0.322
Batch         5  54.5519  10.9104    13.35    0.000
Error        28  22.8817   0.8172
Total        34  78.2640
Terms in selected model: Time(years), Batch

Analysis of Variance

Source       DF  Adj SS   Adj MS  F-Value  P-Value
Time(years)   1   2.979   2.9785     3.64    0.067
Batch         5  54.552  10.9104    13.35    0.000
Error        28  22.882   0.8172
Total        34  78.264

Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.903992  70.76%     64.50%      51.38%

Regression Equation
Batch
1      Concentration(%) = 102.831 - 0.344 Time(years)

2      Concentration(%) = 103.555 - 0.344 Time(years)

3      Concentration(%) = 100.785 - 0.344 Time(years)

4      Concentration(%) = 102.566 - 0.344 Time(years)

5      Concentration(%) = 101.273 - 0.344 Time(years)

6      Concentration(%) = 99.723 - 0.344 Time(years)

Shelf Life Estimation

Lower spec limit = 90
Shelf life = time period in which you can be 95% confident that at
least 95% of response is above lower spec limit

Batch    Shelf Life
1            18.360
2            19.114
3            14.931
4            17.532
5            15.409
6            12.965
Overall      12.965
```

Figure 43: Results of analysis obtained using batch as fixed factor

Figure 44 below shows both the residuals and the shelf life plots of the product based upon ICHQ1E guideline. This model considers batch as a fixed factor and assumes there is no future production.



(a) Shelf life plots



(b) Residual plots

Figure 44: Shelf life and residual plots treating batch as the fixed factor for $\alpha = 0.25$

The residuals appear to be reasonably normal and randomly scattered about zero. Thisindicates there are no outliers or unusual observations. Batch can also be treated as random factor where all future production from comparable processes can be treated similar. This is discussed in Scenario 2.

- **Scenario 2: Considering Batch as Random factor and $\alpha = 0.25$ (Poolable Slopes)**

  Figure 45 shows the summary of the model when batch is treated as random factor. When the batch factor is random, 95th percentile provides a better estimate of shelf life compared to 50th percentile. Random factor takes future batches in to consideration. The shelf life for the present data is approximately 14.5434 years.

  The coefficients table in Figure 45 shows the estimated coefficients for the fixed effects in the model. These coefficients are the intercept (Constant) and the slope (Month) for the marginal fitted equation 22, which predicts the fitted value for any random batch.

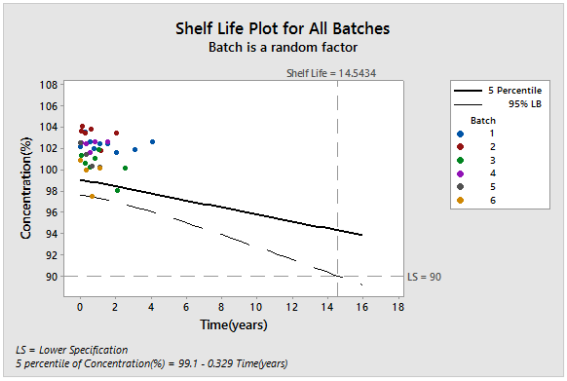
$$Concentration(\%) = 101.79 - (0.329 \times month) \tag{22}$$

```
Factor Information

Factor  Type    Number of Levels  Levels
Batch   Random                 6  1, 2, 3, 4, 5, 6

Model Selection with α = 0.25
Model                                  -2 LogLikelihood  Difference  P-Value
Time(years) Batch Time(years)*Batch           106.394
Time(years) Batch                             106.959      0.5644    0.603
Time(years)                                   128.759     21.8006    0.000

Terms in selected model: Time(years), Batch

Variance Components

Source       Var  % of Total    SE Var   Z-Value  P-Value
Batch   1.901810      69.93%  1.306239  1.455943    0.073
Error   0.817613      30.07%  0.218621  3.739869    0.000
Total   2.719423
Model Summary

       S    R-sq  R-sq(adj)
0.904220  70.35%     69.45%

Coefficients

Term               Coef   SE Coef     DF     T-Value  P-Value
Constant     101.791443  0.603173   5.59  168.760084    0.000
Time(years)   -0.329069  0.180047  28.62   -1.827679    0.078


Random Effect Predictions

Term        BLUP     StDev    DF    T-Value  P-Value
Batch
  1     0.971888  0.641329  6.86   1.515428    0.174
  2     1.636401  0.652400  7.25   2.508280    0.039
  3    -0.962380  0.644369  6.96  -1.493523    0.179
  4     0.704024  0.664465  7.67   1.059536    0.322
  5    -0.475179  0.682828  8.29  -0.695898    0.506
  6    -1.874755  0.682828  8.29  -2.745573    0.024



Shelf Life Estimation

Lower spec limit = 90
Shelf life = time period in which you can be 95% confident that
at least 95% of response is above lower spec limit

Shelf life for all batches = 14.5434
```

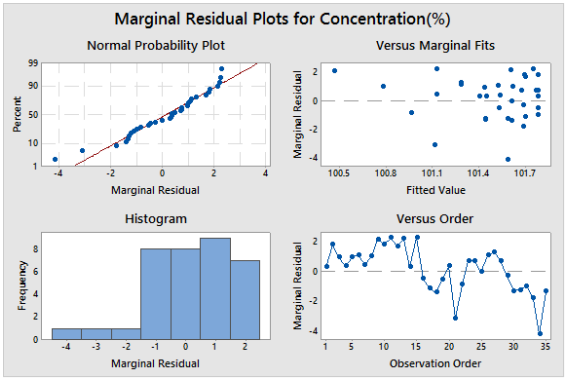
Figure 45: Results of analysis obtained using batch as random factor and α=0.25

The table of random effect predictions in Figure 45 shows the predicted values for the random terms. With these predicted values, intercept and the slope for the conditional fitted equations can be determined, which predict the fitted values for the specific batches.

The plots shown in Figure 46 below show the shelf life and residuals plot of the product based upon ICHQ1E guideline.



(a) Shelf life plots



(b) Residual plots

Figure 46: Shelf life and residual plots treating batch as a random factor for $\alpha = 0.25$

The marginal residual plots for concentration% in Figure 46b show that the marginal residuals may not be normally distributed with constant variance. The histogram is flatter than the normal distribution, and the points in the normal probability plot do not follow the line well. One reason for the

non-normal behavior of the marginal residuals is that, the variance of the marginal residuals depends on the time variable and may not be constant.

- **Analysis for non-poolable batches at** $\alpha = 0.25$

  Another scenario that uses data with non-poolable batches is also discussed. In this case, a different data set has been used and is shown in Figure 47.

Batch	Time(years)	Concentration (%)
1	0.014	100.4
1	0.28	100.3
1	0.514	99.7
1	0.769	99.2
1	1.074	98.9
1	1.533	98.2
1	2.03	97.3
1	3.071	95.7
1	4.049	94.5
2	0.022	100.7
2	0.118	100.6
2	0.272	100.3
2	0.566	99.9
2	1.165	98.6
2	2.022	97.6
2	3.077	96.4
3	0.025	100.2
3	0.275	99.7
3	0.547	99.2
3	0.797	99
3	1.041	98.8
3	2.058	96.4
3	2.519	96.2
4	0.066	100.4
4	0.343	100
4	0.533	99.5
4	0.802	99.3
4	1.033	99.3
4	1.538	98.2
5	0.011	100.5
5	0.31	99.8
5	0.624	99.1
5	1.063	98.4
6	0.011	100.1
6	0.31	99.5
6	0.624	98.5
6	1.063	98.4

Figure 47: Data used for analysis. In this case, non-poolable batches have been used.

4.6 Proposed approach

The following notation will be used for the Full Model:

$$Y_{ij} = \mu + L_i + \beta + B_i + t_{ij} + E_{ij}, \forall i = 1, 2, ..n; and j = 1, 2, ...T_j \qquad (23)$$

Here,

- Y_{ij} is the response for lot i at time point j
- μ is the average y-intercept across all lots
- β is the average slope across all lots
- L_i is a random variable that allows the y-intercept to vary from μ for a given lot; L_i has a normal distribution with mean 0 and variance $\frac{2}{L}$
- B_i is a random variable that allows the slope to vary from β for a given lot; B_i has a normal distribution with mean 0 and variance σ_B^2
- t_{ij} is the time point for measurement j of lot i
- E_{ij} is a random normal error term created by measurement error with mean 0 and variance σ_E^2
- n is the number of lots
- T_i is the number of responses obtained for lot i
- L_i, B_i , E_{ij} are jointly independent

The dataset used for the analysis is included in Figure 47. Results from the analysis are included in Figure 48.

Class Level Information		
Class	Levels	Values
Batch	6	1 2 3 4 5 6

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	0.04276

Fit Statistics	
-2 Res Log Likelihood	8.5
AIC (smaller is better)	10.5
AICC (smaller is better)	10.7
BIC (smaller is better)	11.7

(a) Fixed model results

Class Level Information		
Class	Levels	Values
Batch	6	1 2 3 4 5 6

Covariance Parameter Estimates	
Cov Parm	Estimate
Batch	0.09334
Residual	0.04737

Fit Statistics	
-2 Res Log Likelihood	12.3
AIC (smaller is better)	16.3
AICC (smaller is better)	16.6
BIC (smaller is better)	15.8

(b) Random Model Results

Class Level Information		
Class	Levels	Values
Batch	6	1 2 3 4 5 6

Covariance Parameter Estimates	
Cov Parm	Estimate
Batch	0.06312
Time_years_*Batch	0.01579
Residual	0.04185

Fit Statistics	
-2 Res Log Likelihood	10.8
AIC (smaller is better)	16.8
AICC (smaller is better)	17.6
BIC (smaller is better)	16.2

(c) Full Model Results

Figure 48: Results from different models

All the models were analyzed using SAS PROC Mixed procedure. In this case, batch is identified as a classification variable and has 6 levels. Fit statistics has the estimates for Akaikes Information Criterion (AIC), AICC and BIC.

AIC criteria is summarized in Table 20:

Model	AIC
Fixed	10.5
Random	16.3
Full	16.8

Table 20: Summary of the AIC criteria for the three models used

Based upon the results, as seen from the table, the AIC criteria is the smallest for the fixed model. Hence, fixed model is recommended for shelf life determination of the product. According to the ICHQ1E guideline, if there is a significant difference in slopes among batches, it is not appropriate to combine the data from all batches. The shelf life for individual batches in the stability study can be estimated by using individual intercepts and individual slopes. If the lots have similar slopes and intercepts, overall process lot mean can be used to determine the shelf life. If the lots have different slopes or intercepts, worst case lot(WCL) is recommended to determine the shelf life.

Three linear degradation models are fit, and the three models are the following:

- Model 1 : Different slopes and different intercepts for the batches
- Model 2: Common slope and different intercepts for the batches.
- Model 3: Common slope and common intercept for the batches.

The recommended model is determined by the following procedure:

1. Model 1 is used to fit with the time effect coming first in the model, followed by the batch effect, and then the interaction. Using Type I (Sequential) sums-of-squares, a test for equal slopes was performed (Source C in the output).
    - If the p-value is less than 0.25, the slopes are assumed to be different across batches. The procedure stops and Model 1 is used to estimate the expiration date.
    - If the p-value is greater than or equal to 0.25, the slopes are assumed to be common across batches. The procedure continues to Step 2.

2. If the conclusion from Step 1 is common slopes, then a test for equal intercepts is performed using Type I (Sequential) sums-of-squares from Model 1 (Source B in the output).

    - If the p-value is less than 0.25, the intercepts are assumed to be different across batches, and Model 2 is used to estimate the expiration date.
    - If the p-value is greater than or equal to 0.25, the intercepts are assumed to be common across batches, and Model 3 is used to estimate the expiration date.

When Model 1 (different slopes and different intercepts) is used for estimating the expiration date, the MSE (mean squared error) is not pooled across batches. Prediction intervals are computed for each batch using individual mean squared errors, and the interval that crosses the specification limit first is used to estimate the expiration date. Figure 49 gives the summary of the output.

The test for equal slopes has a p-value of 0.1861. Because this is smaller than a significance level of 0.25, the test was rejected, and it can be concluded that the degradation slopes are not equal between batches.

The test for equal intercepts and slopes has a p-value of < 0.0001. Because this is smaller than a significance level of 0.25(ICHQ1E recommends a p-value$<$0.25 for significance), the test was rejected, and it can be concluded that the intercepts are different between batches. Using the p-value criteria $<$0.25 will cause the expiry to shorten and favor the consumer/patient with shortened shelf life.

Intercept	Slope	Earliest Crossing Time
Different	Different	3.509979
Different	Common	6.172841
Common	Common	6.568067

Source	DF	SS	Mean Square	F Statistic	Prob>F
A	10	3.180083	0.318008	7.436234	<.0001*
B	5	2.829324	0.565865	13.23205	<.0001*
C	5	0.35076	0.070152	1.640417	0.1861
D	25	1.069118	0.042765		
E	12	361886	30157.17		

Figure 49: Summary of the output obtained based on different slopes and intercepts

Because the test for equal slopes and the test for equal intercepts was rejected, the chosen model is the one with different intercepts and different slopes. The best model accepted at the significance level of 0.25 has different intercepts and different slopes. Figure 50 shows a graph of concentration % vs time when different slopes and intercepts were used. Based upon the plot, the model suggests the earliest crossing time is 3.509979 years with 95 percent confidence. ICH Guidelines indicate an expiration time of 3.509979 years. The plot indicates where the 95% confidence interval intersects the specification and the intersection point is the shelf life of product based upon ICHQ1E.

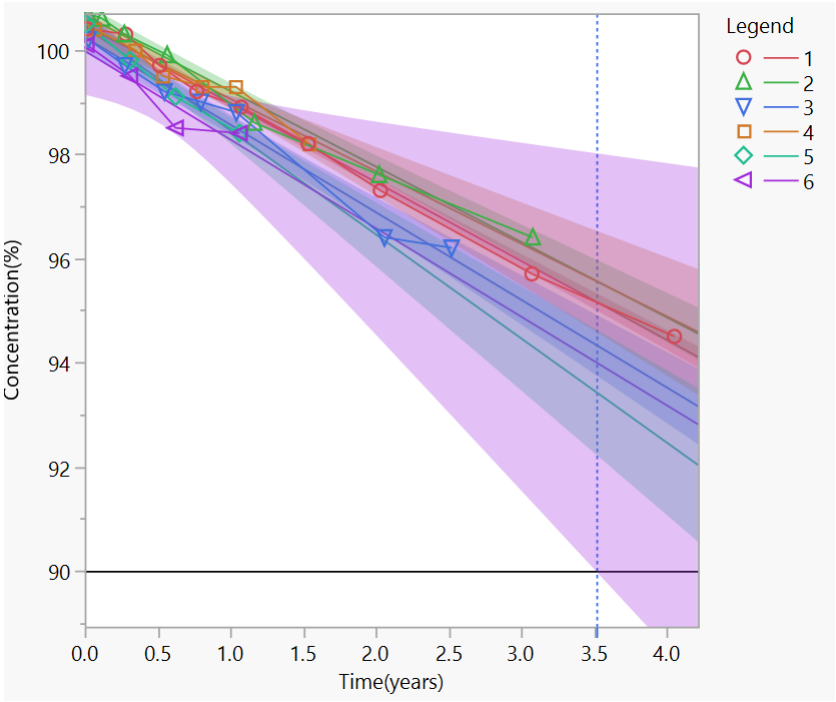


Figure 50: Different slopes and different intercepts(two-sided interval)

Figure 51 shows a graph of concentration % vs time when common slopes and intercepts were used.

If the common slope and Common intercept model was used, the 95% confidence interval of the mean slope intersects the specification and hence the shelf life in this case is determined as 6.57 years.

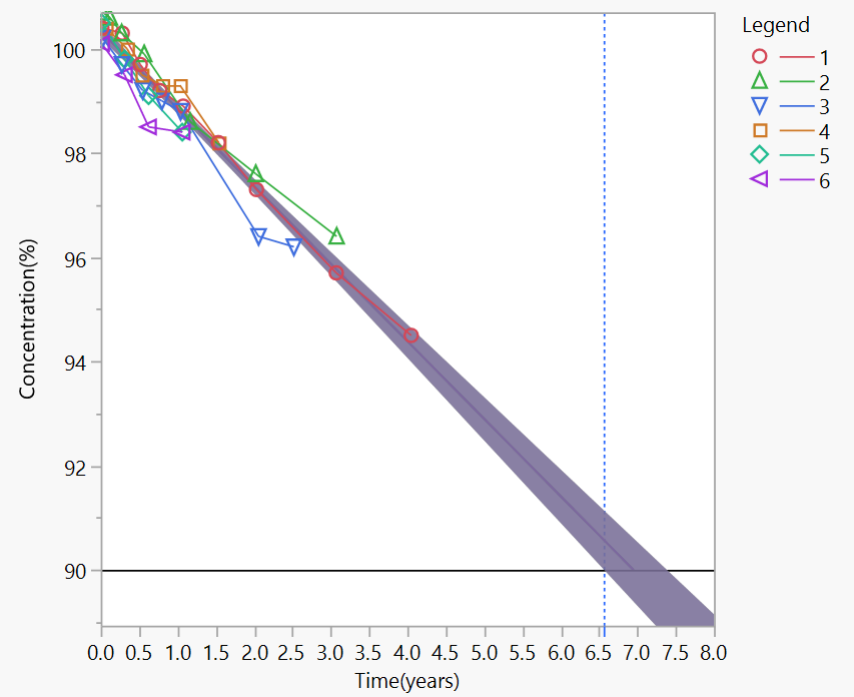


Figure 51: Common slopes and intercepts(two-sided interval)

The above methodology fits a 2-sided interval for a parameter with one-sided specification and that will result in a shorter expiration date/shelf life.

The proposed approach is to fit a one-sided confidence interval for a parameter with one-sided specification. Also, more appropriate method for establishing shelf life is to fit a 95% one-sided confidence interval for the worst case lot (WCL), that represents the true shelf life. Results from the analysis are shown in the Figure 52 below:

```
Factor Information

Factor  Type   Number of Levels  Levels
Batch   Fixed                6  1, 2, 3, 4, 5, 6


Model Selection with α = 0.25

Source             DF   Seq SS   Seq MS  F-Value  P-Value
Time(years)         1  76.9340  76.9340  1799.01    0.000
Batch               5   2.8293   0.5659    13.23    0.000
Time(years)*Batch   5   0.3508   0.0702     1.64    0.186
Error              25   1.0691   0.0428
Total              36  81.1832

Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.206796  98.68%     98.10%      95.57%


Regression Equation

Batch
1      Concentration(%) = 100.489 - 1.5153 Time(years)

2      Concentration(%) = 100.656 - 1.4490 Time(years)

3      Concentration(%) = 100.245 - 1.6823 Time(years)

4      Concentration(%) = 100.451 - 1.393 Time(years)

5      Concentration(%) = 100.454 - 1.999 Time(years)

6      Concentration(%) = 99.979 - 1.701 Time(years)
```

Figure 52: Results from the analysis using 95% one-sided confidence interval for the WCL (worst case lot)

Based upon the above analysis, Batch 5 has the steepest slope and is considered the worstcase lot. The shelf life will be based upon the earliest time at which the 95 percent confidence limit for the mean intersects the specification limit (lower specification limit=90) in the dataset.

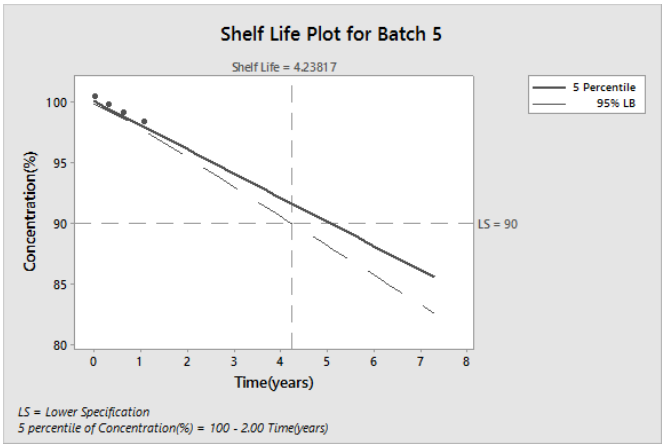


Figure 53: Worst case lot with one-sided lower confidence interval and estimation of shelf life

As seen from Figure 53, the analysis suggests a shelf life of 4.3 years compared to 3.5 years for the same study based upon applying appropriate confidence interval.

Shelf life determination is a key component in the overall strategy for biopharmaceutical products. Current guidelines were reviewed and new methods were proposed to improve the determination of shelf life. The proposed methods include consideration of sidedness of intervals based upon one-sided specification or two-sided specification, selecting the model based upon AIC criteria instead of a given significance level and based upon poolability of slopes using the worst case lot to determine the shelf life. As demonstrated in the analysis, shelf life can be extended or shortened based upon the statistical methodology used.

CHAPTER 5

Conclusion

This chapter concludes the current study and future work.

5.1 Conclusion

Bio-technology manufacturing processes are complex and consist of multiple processing operations with numerous parameters including inputs and their correlation structures impact the final output or productivity of the process. Modeling the relation between input parameters, process conditions to the output attributes require advanced analytical tools and complex models. Shelf life of a biopharmaceutical product is typically based upon the stability data. The shelf life of a pharmaceutical product is the maximum time at which the response of a stability limiting characteristic for all parameters (or other unit) in the batch does not exceed the specification limit. ICH Q1E provides general guidance for establishing the shelf life based upon average of the results, but does not provide any guidance on individual results. Analytical methods are used to measure the quality of the product.

Typically, like a manufacturing process, they are developed in a pre-commercial setting and need to be transferred to commercial manufacturing. Main objective of methods transfer is to avoid release of product that does not meet specifications (safety impact) often considered as Type I error as well as avoid rejection of good product (business impact- Type II error). Because of these potentially high-impact implications, an assay cannot be used by the new lab until transfer is successfully completed. The assay transfer process involves robust studies from the receiving and sending labs to establish equivalence. There are no official regulatory guideline exists for assay transfers. FDA guidance for validation of bioanalytical

methods, assay transfer between labs is considered as partial validation.

One of the objective of this work is to compare multiple statistical approaches to model fermentation process for penicillin manufacturing process and propose a new methodology that is more appropriate for auto correlated and continuous data. A comprehensive review of available statistical models was conducted in Chapter 2. All the methods were associated with certain advantages and disadvantages and complex to implement. A new method, which is a combination of Principal components analysis and Symbolic regression was proposed in Chapter 4. Advantages of proposed method is that new variables are orthogonal and address the concern for collinearity and variance is explained by the principal components and noise remain in the residuals.

Another objective of the study is to analyze current practices for analytical method transfers and propose a more robust methodology to avoid out of specification results. Statistical approaches two-one sided t test (TOST) and Concordance Correlation Coefficient(CCC) were compared to establish equivalence. An important step of TOST and CCC analysis is determining the desired power and the corresponding sample size needed for conducting the test. The power is defined as the probability of correctly accepting equivalence at a given true mean difference between two populations.

A detailed analysis with sample sizes and power was summarized and optimal sample size based upon the cost and resources balancing power is recommended. A new methodology was proposed for establishing equivalence margin and incorporated Process capability in to analytical method transfer to avoid risk for potential out of specification results and summarized in Chapter 4. Another objective of the study is to study the current regulations and statistical methodologies for establishing shelf life of biopharmaceutical products. Different factors were included

in the study including poolability of data, worst case lot and including fixed and random effects for establishing and extension of shelf life. Different statistical models were built to establish the best suitable model for the data. A new method was proposed for selecting the best model that fits the data based upon Akaikes Information Criterion. The proposed method also included usage of appropriate sided-ness of confidence intervals based upon the specifications. i,e; using a two-sided confidence interval for a two-sided specification and one-sided upper or lower confidence intervals for one-sided specification. Comparison of shelf life using existing guidance and proposed methodology was summarized in Chapter 4.

5.2 Future work

A practical optimization methodology for Penicillin manufacturing process was developed in this work. The methodology was compared to and tested with various other techniques. A natural extension of this work will be to implement this for an actual biopharmaceutical manufacturing process and a comparison of the results with those obtained from the simulations used in this work. The data was used to build the models and predict the performance of the process after manufacturing is completed. This methodology can be used for real-time monitoring and predicting the process performance by comparing the current batch with models built with historical data. Multiple statistical methods were compared to establish the equivalence during method transfers and a new method was proposed to control the risk of out of specification results. This can be further expanded to incorporate the effects of accuracy and precision of the measurement systems to determine the equivalence margin before initiating the transfer. Measurement system Analysis like Gage R&R studies need to be performed to quantify accuracy and precision and variance components to understand the variability of the analytical method. The proposed methodology of establishing shelf life is based upon

the data collected in the literature and studied under normal operating conditions. The study can be further expanded for accelerated storage conditions like exposing the drug substance to heat, humidity to study the impact of environmental conditions on shelf life.

Also, when random effects are considered, best linear unbiased predictors (BLUPs) can be calculated and used for shelf life determination. BLUPs provide more accurate predictions for the observed values that are outputs from random processes. The methodology can be further expanded to combine drug substance and drug product stability studies to come up with a overall shelf life /specifications for the product.

LIST OF REFERENCES

[1] "Biologic medicines in development," Tech. Rep., 2013. [Online]. Available: http://www.phrma.org/sites/default/files/pdf/biologics2013.pdf

[2] J.-M. Lee, C. K. Yoo, and I.-B. Lee, "Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis," *Journal of Biotechnology*, vol. 110, no. 2, pp. 119–136, 2004.

[3] S. Singh and M. Bakshi, "Stress test to determine inherent stability of drugs," *Pharm Technol*, vol. 4, pp. 1–14, 2000.

[4] J. T. Carstensen, *Advanced pharmaceutical solids*. CRC Press, 2000.

[5] I. H. T. Guideline, "Evaluation for stability data q1e," in *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*, 2003.

[6] R. Capen, D. Christopher, P. Forenzo, C. Ireland, O. Liu, S. Lyapustina, J. ONeill, N. Patterson, M. Quinlan, D. Sandell, *et al.*, "On the shelf life of pharmaceutical products," *AAPS PharmSciTech*, vol. 13, no. 3, pp. 911–918, 2012.

[7] E. Rozet, W. Dewé, E. Ziemons, A. Bouklouze, B. Boulanger, and P. Hubert, "Methodologies for the transfer of analytical methods: a review," *Journal of Chromatography B*, vol. 877, no. 23, pp. 2214–2223, 2009.

[8] O. Marjanovic, B. Lennox, D. Sandoz, K. Smith, and M. Crofts, "Real-time monitoring of an industrial batch process," *Computers and chemical engineering*, vol. 30, no. 10, pp. 1476–1481, 2006.

[9] V. Aryadoust, "Application of evolutionary algorithm-based symbolic regression to language assessment: Toward nonlinear modeling," *Psychological Test and Assessment Modeling*, vol. 57, no. 3, p. 301, 2015.

[10] D. B. Fogel, "An overview of evolutionary programming," in *Evolutionary Algorithms*. Springer, 1999, pp. 89–109.

[11] M. D. Schmidt and H. Lipson, "Data-mining dynamical systems: Automated symbolic system identification for exploratory analysis," in *ASME 2008 9th Biennial Conference on Engineering Systems Design and Analysis*. American Society of Mechanical Engineers, 2008, pp. 643–649.

[12] T. D. Gwiazda, *Crossover for single-objective numerical optimization problems*. Tomasz Gwiazda, 2006, vol. 1.

[13] M. Schmidt and H. Lipson, "Symbolic regression of implicit equations," in *Genetic Programming Theory and Practice VII*. Springer, 2010, pp. 73–85.

[14] J. R. Koza, "Human-competitive results produced by genetic programming," *Genetic Programming and Evolvable Machines*, vol. 11, no. 3-4, pp. 251–284, 2010.

[15] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[16] R. S. Michalski, "Learnable evolution model: Evolutionary processes guided by machine learning," *Machine learning*, vol. 38, no. 1, pp. 9–40, 2000.

[17] V. Aryadoust and S. Liu, "Predicting efl writing ability from levels of mental representation measured by coh-metrix: A structural equation modeling study," *Assessing Writing*, vol. 24, pp. 35–58, 2015.

[18] M. D. Schmidt and H. Lipson, "Age-fitness pareto optimization," in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. ACM, 2010, pp. 543–544.

[19] V. Vapnik, *The nature of statistical learning theory*. Springer science and business media, 2013.

[20] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling*. Springer, 1998, pp. 55–85.

[21] X. Peng and Y. Wang, "A normal least squares support vector machine (nls-svm) and its learning algorithm," *Neurocomputing*, vol. 72, no. 16, pp. 3734–3741, 2009.

[22] "Mathworks - makers of matlab and simulink." [Online]. Available: http://www.mathworks.com/

[23] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[24] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[25] "International society for pharmaceutical engineering — ispe — gmp." [Online]. Available: http://www.ispe.org/

[26] L. Kaminski, U. Schepers, and H. Wätzig, "Analytical method transfer using equivalence tests with reasonable acceptance criteria and appropriate effort: extension of the ispe concept," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 53, no. 5, pp. 1124–1129, 2010.

[27] W. Home, "Who guidelines on transfer of technology in pharmaceutical manufacturing language: English," 2011.

[28] D. J. Schuirmann, "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 15, no. 6, pp. 657–680, 1987.

[29] "Standard practice for conducting equivalence testing in laboratory applications." [Online]. Available: http://www.astm.org/

[30] G. B. Limentani, M. C. Ringo, F. Ye, M. L. Bergquist, and E. O. MCSorley, "Beyond the t-test: statistical equivalence testing," 2005.

[31] I. Lawrence and K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255–268, 1989.

[32] I. Lawrence and K. Lin, "Assay validation using the concordance correlation coefficient," *Biometrics*, pp. 599–604, 1992.

[33] L. I. Lin *et al.*, "Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence," *Statistics in medicine*, vol. 19, no. 2, pp. 255–270, 2000.

[34] P. Nomikos and J. F. MacGregor, "Monitoring batch processes using multiway principal component analysis," *AIChE Journal*, vol. 40, no. 8, pp. 1361–1375, 1994.

[35] P. Nomikos and J. F. MacGregor, "Multivariate spc charts for monitoring batch processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, 1995.

[36] T. Kourti, P. Nomikos, and J. F. MacGregor, "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway pls," *Journal of process control*, vol. 5, no. 4, pp. 277–284, 1995.

[37] D. Dong and T. J. McAvoy, "Nonlinear principal component analysisbased on principal curves and neural networks," *Computers & Chemical Engineering*, vol. 20, no. 1, pp. 65–78, 1996.

[38] E. Martin and A. Morris, "An overview of multivariate statistical process control in continuous and batch process performance monitoring," *Transactions of the Institute of Measurement and Control*, vol. 18, no. 1, pp. 51–60, 1996.

[39] S. Rännar, J. F. MacGregor, and S. Wold, "Adaptive batch monitoring using hierarchical pca," *Chemometrics and intelligent laboratory systems*, vol. 41, no. 1, pp. 73–81, 1998.

[40] S. Wold, N. Kettaneh, H. Fridén, and A. Holmberg, "Modelling and diagnostics of batch processes and analogous kinetic experiments," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1, pp. 331–340, 1998.

[41] C.-Y. Chou, H.-R. Liu, X. Huang, and C.-H. Chen, "Economic-statistical design of multivariate control charts using quality loss function," *The International Journal of Advanced Manufacturing Technology*, vol. 20, no. 12, pp. 916–924, 2002.

[42] H. Zhang, B. Lennox, P. R. Goulding, and Y. Wang, "Adaptive information sharing factors in federated kalman filtering," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 79–84, 2002.

[43] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," *science*, vol. 324, no. 5923, pp. 81–85, 2009.

[44] V. Centner, J. Verdú-Andrés, B. Walczak, D. Jouan-Rimbaud, F. Despagne, L. Pasti, D.-L. Massart, and O. E. De Noord, "Comparison of multivariate calibration techniques applied to experimental nir data sets," *Applied spectroscopy*, vol. 54, no. 4, pp. 608–623, 2000.

[45] U. Thissen, M. Pepers, B. Üstün, W. Melssen, and L. Buydens, "Comparing support vector machines to pls for spectral regression applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 73, no. 2, pp. 169–179, 2004.

[46] L. Yi and W. Haiqing, "Modelling of the penicillin fermentation process via ls-svm based on pensim simulator," *Chemical Reaction Engineering and Technology*, vol. 22, no. 3, p. 252, 2006.

[47] X. Wang, J. Chen, C. Liu, and F. Pan, "Hybrid modeling of penicillin fermentation process based on least square support vector machine," *Chemical Engineering Research and Design*, vol. 88, no. 4, pp. 415–420, 2010.

[48] D. LeBlond, D. Griffith, and K. Aubuchon, "Linear regression 102: stability shelf life estimation using analysis of covariance," *Journal of Validation Technology*, vol. 17, no. 3, p. 47, 2011.

[49] S. J. Ruberg and J. W. Stegeman, "Pooling data for stability studies: testing the equality of batch degradation slopes," *Biometrics*, pp. 1059–1069, 1991.

[50] S. J. Ruberg and J. C. Hsu, "Multiple comparison procedures for pooling batches in stability studies," *Technometrics*, vol. 34, no. 4, pp. 465–472, 1992.

[51] W. Liu and P. Somerville, "Stepwise multiple tests for successive comparisons of treatment effects," *Computational statistics & data analysis*, vol. 46, no. 1, pp. 189–199, 2004.

[52] W. Liu, S. Lin, and W. W. Piegorsch, "Construction of exact simultaneous confidence bands for a simple linear regression model," *International Statistical Review*, vol. 76, no. 1, pp. 39–57, 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1751-5823.2007.00027.x

[53] S. K. Sumie Yoshioka, Yukio Aso, "Assessment of shelf-life equivalence of pharmaceutical products," *Chemical & Pharmaceutical Bulletin*, vol. 45, no. 9, pp. 1482–1484, 1997.

[54] Y. Tsong, W.-J. Chen, T.-Y. D. Lin, and C. W. Chen, "Shelf life determination based on equivalence assessment," *Journal of biopharmaceutical statistics*, vol. 13, no. 3, pp. 431–449, 2003.

[55] G. D. Djira, L. A. Hothorn, and Y. Tsong, "Equivalence tests for shelf life and average drug content in stability studies," *Journal of biopharmaceutical statistics*, vol. 18, no. 5, pp. 985–995, 2008.

[56] I. Knezevic, "Stability evaluation of vaccines: Who approach," *Biologicals*, vol. 37, no. 6, pp. 357–359, 2009.

[57] W. H. Organizaion, "Guidelines on stability evaluation of vaccines," *Biologicals*, vol. 37, no. 6, pp. 424–434, 2009.

[58] T. L. Schofield, "Vaccine stability study design and analysis to support product licensure," *Biologicals*, vol. 37, no. 6, pp. 387–396, 2009.

[59] K. Komka and S. Kemeny, "A modified error model for the assessment of stability of pharmaceutical products," *Chemometrics and intelligent laboratory systems*, vol. 72, no. 2, pp. 161–165, 2004.

[60] T. E. Norwood, "Statistical analysis of pharmaceutical stability data," *Drug Development and Industrial Pharmacy*, vol. 12, no. 4, pp. 553–560, 1986.

[61] P. D. Association *et al.*, "Pda technical report 57," *Analytical Method Development and Qualification for Biotechnology Products. Bethesda, MD: PDA*, vol. 67, pp. 1–69, 2012.

[62] D. Chambers, G. Kelly, G. Limentani, A. Lister, K. R. Lung, and E. Warner, "Analytical method equivalency," *Pharmaceutical Technology*, 2005.

[63] C. Agut, A. Caron, C. Giordano, D. Hoffman, and A. Ségalini, "Transfer of analytical procedures: a panel of strategies selected for risk management, with emphasis on an integrated equivalence-based comparative testing approach," *Journal of pharmaceutical and biomedical analysis*, vol. 56, no. 2, pp. 293–303, 2011.

[64] J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[65] B. Zhong and J. Shao, "Evaluating the agreement of two quantitative assays with repeated measurements," *Journal of Biopharmaceutical Statistics*, vol. 13, no. 1, pp. 75–86, 2003.

[66] J. Shao and B. Zhong, "Assessing the agreement between two quantitative assays with repeated measurements," *Journal of Biopharmaceutical Statistics*, vol. 14, no. 1, pp. 201–212, 2004.

[67] J. Zhong, K. Lee, and Y. Tsong, "Statistical assessment of analytical method transfer," *Journal of Biopharmaceutical Statistics*, vol. 18, no. 5, pp. 1005–1012, 2008.

[68] G. Birol, C. Ündey, and A. Cinar, "A modular simulation package for fed-batch fermentation: penicillin production," *Computers & Chemical Engineering*, vol. 26, no. 11, pp. 1553–1565, 2002.

[69] R. Bajpai and M. Reuss, "A mechanistic model for penicillin production," *Journal of Chemical Technology and Biotechnology*, vol. 30, no. 1, pp. 332–344, 1980.

[70] C. Ündey, S. Ertunç, T. Mistretta, and B. Looze, "Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control," *Journal of Process Control*, vol. 20, no. 9, pp. 1009–1018, 2010.

[71] P. Van den Kerkhof, G. Gins, J. Vanlaer, and J. F. Van Impe, "Dynamic model-based fault diagnosis for (bio) chemical batch processes," *Computers & Chemical Engineering*, vol. 40, pp. 12–21, 2012.

[72] K. R. Lung, M. A. Gorko, J. Llewelyn, and N. Wiggins, "Statistical method for the determination of equivalence of automated test procedures," *Journal of Analytical Methods in Chemistry*, vol. 25, no. 6, pp. 123–127, 1900.

[73] W. Dewé, B. Govaerts, B. Boulanger, E. Rozet, P. Chiap, and P. Hubert, "Using total error as decision criterion in analytical method transfer," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 2, pp. 262–268, 2007.

[74] [Online]. Available: http://www.vub.ac.be

[75] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[76] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[77] L. Guerin and W. W. Stroup, "A simulation study to evaluate proc mixed analysis of repeated measures data," 2000.

[78] L. Torbeck, "Usp¡ 1010¿ analytical data-interpretation and treatment appendix e: Comparison of methods," *Journal of GXP Compliance*, vol. 16, no. 4, p. 30, 2012.

[79] G. McBride, "A proposal for strength-of-agreement criteria for lins concordance correlation coefficient," *NIWA Client Report: HAM2005-062*, 2005.

BIBLIOGRAPHY

[Online]. Available: http://www.vub.ac.be

"International society for pharmaceutical engineering — ispe — gmp." [Online]. Available: http://www.ispe.org/

"Mathworks - makers of matlab and simulink." [Online]. Available: http://www.mathworks.com/

"Standard practice for conducting equivalence testing in laboratory applications." [Online]. Available: http://www.astm.org/

"Biologic medicines in development," Tech. Rep., 2013. [Online]. Available: http://www.phrma.org/sites/default/files/pdf/biologics2013.pdf

Agut, C., Caron, A., Giordano, C., Hoffman, D., and Ségalini, A., "Transfer of analytical procedures: a panel of strategies selected for risk management, with emphasis on an integrated equivalence-based comparative testing approach," *Journal of pharmaceutical and biomedical analysis*, vol. 56, no. 2, pp. 293–303, 2011.

Akaike, H., "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

Aryadoust, V., "Application of evolutionary algorithm-based symbolic regression to language assessment: Toward nonlinear modeling," *Psychological Test and Assessment Modeling*, vol. 57, no. 3, p. 301, 2015.

Aryadoust, V. and Liu, S., "Predicting efl writing ability from levels of mental representation measured by coh-metrix: A structural equation modeling study," *Assessing Writing*, vol. 24, pp. 35–58, 2015.

Association, P. D. *et al.*, "Pda technical report 57," *Analytical Method Development and Qualification for Biotechnology Products. Bethesda, MD: PDA*, vol. 67, pp. 1–69, 2012.

Bajpai, R. and Reuss, M., "A mechanistic model for penicillin production," *Journal of Chemical Technology and Biotechnology*, vol. 30, no. 1, pp. 332–344, 1980.

Birol, G., Ündey, C., and Cinar, A., "A modular simulation package for fed-batch fermentation: penicillin production," *Computers & Chemical Engineering*, vol. 26, no. 11, pp. 1553–1565, 2002.

Bland, J. M. and Altman, D., "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

Burnham, K. P. and Anderson, D. R., *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

Capen, R., Christopher, D., Forenzo, P., Ireland, C., Liu, O., Lyapustina, S., ONeill, J., Patterson, N., Quinlan, M., Sandell, D., *et al.*, "On the shelf life of pharmaceutical products," *AAPS PharmSciTech*, vol. 13, no. 3, pp. 911–918, 2012.

Carstensen, J. T., *Advanced pharmaceutical solids*. CRC Press, 2000.

Centner, V., Verdú-Andrés, J., Walczak, B., Jouan-Rimbaud, D., Despagne, F., Pasti, L., Massart, D.-L., and De Noord, O. E., "Comparison of multivariate calibration techniques applied to experimental nir data sets," *Applied spectroscopy*, vol. 54, no. 4, pp. 608–623, 2000.

Chambers, D., Kelly, G., Limentani, G., Lister, A., Lung, K. R., and Warner, E., "Analytical method equivalency," *Pharmaceutical Technology*, 2005.

Chou, C.-Y., Liu, H.-R., Huang, X., and Chen, C.-H., "Economic-statistical design of multivariate control charts using quality loss function," *The International Journal of Advanced Manufacturing Technology*, vol. 20, no. 12, pp. 916–924, 2002.

Cortes, C. and Vapnik, V., "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

Dewé, W., Govaerts, B., Boulanger, B., Rozet, E., Chiap, P., and Hubert, P., "Using total error as decision criterion in analytical method transfer," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 2, pp. 262–268, 2007.

Djira, G. D., Hothorn, L. A., and Tsong, Y., "Equivalence tests for shelf life and average drug content in stability studies," *Journal of biopharmaceutical statistics*, vol. 18, no. 5, pp. 985–995, 2008.

Dong, D. and McAvoy, T. J., "Nonlinear principal component analysisbased on principal curves and neural networks," *Computers & Chemical Engineering*, vol. 20, no. 1, pp. 65–78, 1996.

Fogel, D. B., "An overview of evolutionary programming," in *Evolutionary Algorithms*. Springer, 1999, pp. 89–109.

Guerin, L. and Stroup, W. W., "A simulation study to evaluate proc mixed analysis of repeated measures data," 2000.

Guideline, I. H. T., "Evaluation for stability data q1e," in *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*, 2003.

Gwiazda, T. D., *Crossover for single-objective numerical optimization problems*. Tomasz Gwiazda, 2006, vol. 1.

Holland, J. H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

Home, W., "Who guidelines on transfer of technology in pharmaceutical manufacturing language: English," 2011.

Kaminski, L., Schepers, U., and Wätzig, H., "Analytical method transfer using equivalence tests with reasonable acceptance criteria and appropriate effort: extension of the ispe concept," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 53, no. 5, pp. 1124–1129, 2010.

Knezevic, I., "Stability evaluation of vaccines: Who approach," *Biologicals*, vol. 37, no. 6, pp. 357–359, 2009.

Komka, K. and Kemeny, S., "A modified error model for the assessment of stability of pharmaceutical products," *Chemometrics and intelligent laboratory systems*, vol. 72, no. 2, pp. 161–165, 2004.

Kourti, T., Nomikos, P., and MacGregor, J. F., "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway pls," *Journal of process control*, vol. 5, no. 4, pp. 277–284, 1995.

Koza, J. R., "Human-competitive results produced by genetic programming," *Genetic Programming and Evolvable Machines*, vol. 11, no. 3-4, pp. 251–284, 2010.

Lawrence, I. and Lin, K., "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255–268, 1989.

Lawrence, I. and Lin, K., "Assay validation using the concordance correlation coefficient," *Biometrics*, pp. 599–604, 1992.

LeBlond, D., Griffith, D., and Aubuchon, K., "Linear regression 102: stability shelf life estimation using analysis of covariance," *Journal of Validation Technology*, vol. 17, no. 3, p. 47, 2011.

Lee, J.-M., Yoo, C. K., and Lee, I.-B., "Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis," *Journal of Biotechnology*, vol. 110, no. 2, pp. 119–136, 2004.

Limentani, G. B., Ringo, M. C., Ye, F., Bergquist, M. L., and MCSorley, E. O., "Beyond the t-test: statistical equivalence testing," 2005.

Lin, L. I. *et al.*, "Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence," *Statistics in medicine*, vol. 19, no. 2, pp. 255–270, 2000.

Liu, W. and Somerville, P., "Stepwise multiple tests for successive comparisons of treatment effects," *Computational statistics & data analysis*, vol. 46, no. 1, pp. 189–199, 2004.

Liu, W., Lin, S., and Piegorsch, W. W., "Construction of exact simultaneous confidence bands for a simple linear regression model," *International Statistical Review*, vol. 76, no. 1, pp. 39–57, 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1751-5823.2007.00027.x

Lung, K. R., Gorko, M. A., Llewelyn, J., and Wiggins, N., "Statistical method for the determination of equivalence of automated test procedures," *Journal of Analytical Methods in Chemistry*, vol. 25, no. 6, pp. 123–127, 1900.

Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., and Crofts, M., "Real-time monitoring of an industrial batch process," *Computers and chemical engineering*, vol. 30, no. 10, pp. 1476–1481, 2006.

Martin, E. and Morris, A., "An overview of multivariate statistical process control in continuous and batch process performance monitoring," *Transactions of the Institute of Measurement and Control*, vol. 18, no. 1, pp. 51–60, 1996.

McBride, G., "A proposal for strength-of-agreement criteria for lins concordance correlation coefficient," *NIWA Client Report: HAM2005-062*, 2005.

Michalski, R. S., "Learnable evolution model: Evolutionary processes guided by machine learning," *Machine learning*, vol. 38, no. 1, pp. 9–40, 2000.

Nomikos, P. and MacGregor, J. F., "Monitoring batch processes using multiway principal component analysis," *AIChE Journal*, vol. 40, no. 8, pp. 1361–1375, 1994.

Nomikos, P. and MacGregor, J. F., "Multivariate spc charts for monitoring batch processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, 1995.

Norwood, T. E., "Statistical analysis of pharmaceutical stability data," *Drug Development and Industrial Pharmacy*, vol. 12, no. 4, pp. 553–560, 1986.

Organizaion, W. H., "Guidelines on stability evaluation of vaccines," *Biologicals*, vol. 37, no. 6, pp. 424–434, 2009.

Peng, X. and Wang, Y., "A normal least squares support vector machine (nls-svm) and its learning algorithm," *Neurocomputing*, vol. 72, no. 16, pp. 3734–3741, 2009.

Rännar, S., MacGregor, J. F., and Wold, S., "Adaptive batch monitoring using hierarchical pca," *Chemometrics and intelligent laboratory systems*, vol. 41, no. 1, pp. 73–81, 1998.

Rozet, E., Dewé, W., Ziemons, E., Bouklouze, A., Boulanger, B., and Hubert, P., "Methodologies for the transfer of analytical methods: a review," *Journal of Chromatography B*, vol. 877, no. 23, pp. 2214–2223, 2009.

Ruberg, S. J. and Hsu, J. C., "Multiple comparison procedures for pooling batches in stability studies," *Technometrics*, vol. 34, no. 4, pp. 465–472, 1992.

Ruberg, S. J. and Stegeman, J. W., "Pooling data for stability studies: testing the equality of batch degradation slopes," *Biometrics*, pp. 1059–1069, 1991.

Schmidt, M. and Lipson, H., "Distilling free-form natural laws from experimental data," *science*, vol. 324, no. 5923, pp. 81–85, 2009.

Schmidt, M. and Lipson, H., "Symbolic regression of implicit equations," in *Genetic Programming Theory and Practice VII*. Springer, 2010, pp. 73–85.

Schmidt, M. D. and Lipson, H., "Data-mining dynamical systems: Automated symbolic system identification for exploratory analysis," in *ASME 2008 9th Biennial Conference on Engineering Systems Design and Analysis*. American Society of Mechanical Engineers, 2008, pp. 643–649.

Schmidt, M. D. and Lipson, H., "Age-fitness pareto optimization," in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. ACM, 2010, pp. 543–544.

Schofield, T. L., "Vaccine stability study design and analysis to support product licensure," *Biologicals*, vol. 37, no. 6, pp. 387–396, 2009.

Schuirmann, D. J., "A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 15, no. 6, pp. 657–680, 1987.

Shao, J. and Zhong, B., "Assessing the agreement between two quantitative assays with repeated measurements," *Journal of Biopharmaceutical Statistics*, vol. 14, no. 1, pp. 201–212, 2004.

Singh, S. and Bakshi, M., "Stress test to determine inherent stability of drugs," *Pharm Technol*, vol. 4, pp. 1–14, 2000.

Sumie Yoshioka, Yukio Aso, S. K., "Assessment of shelf-life equivalence of pharmaceutical products," *Chemical & Pharmaceutical Bulletin*, vol. 45, no. 9, pp. 1482–1484, 1997.

Suykens, J. A. and Vandewalle, J., "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

Thissen, U., Pepers, M., Üstün, B., Melssen, W., and Buydens, L., "Comparing support vector machines to pls for spectral regression applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 73, no. 2, pp. 169–179, 2004.

Torbeck, L., "Usp¡ 1010¿ analytical data-interpretation and treatment appendix e: Comparison of methods," *Journal of GXP Compliance*, vol. 16, no. 4, p. 30, 2012.

Tsong, Y., Chen, W.-J., Lin, T.-Y. D., and Chen, C. W., "Shelf life determination based on equivalence assessment," *Journal of biopharmaceutical statistics*, vol. 13, no. 3, pp. 431–449, 2003.

Ündey, C., Ertunç, S., Mistretta, T., and Looze, B., "Applied advanced process analytics in biopharmaceutical manufacturing: Challenges and prospects in real-time monitoring and control," *Journal of Process Control*, vol. 20, no. 9, pp. 1009–1018, 2010.

Van den Kerkhof, P., Gins, G., Vanlaer, J., and Van Impe, J. F., "Dynamic model-based fault diagnosis for (bio) chemical batch processes," *Computers & Chemical Engineering*, vol. 40, pp. 12–21, 2012.

Vapnik, V., "The support vector method of function estimation," in *Nonlinear Modeling*. Springer, 1998, pp. 55–85.

Vapnik, V., *The nature of statistical learning theory*. Springer science and business media, 2013.

Wang, X., Chen, J., Liu, C., and Pan, F., "Hybrid modeling of penicillin fermentation process based on least square support vector machine," *Chemical Engineering Research and Design*, vol. 88, no. 4, pp. 415–420, 2010.

Wold, S., Kettaneh, N., Fridén, H., and Holmberg, A., "Modelling and diagnostics of batch processes and analogous kinetic experiments," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1, pp. 331–340, 1998.

Yi, L. and Haiqing, W., "Modelling of the penicillin fermentation process via ls-svm based on pensim simulator," *Chemical Reaction Engineering and Technology*, vol. 22, no. 3, p. 252, 2006.

Zhang, H., Lennox, B., Goulding, P. R., and Wang, Y., "Adaptive information sharing factors in federated kalman filtering," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 79–84, 2002.

Zhong, B. and Shao, J., "Evaluating the agreement of two quantitative assays with repeated measurements," *Journal of Biopharmaceutical Statistics*, vol. 13, no. 1, pp. 75–86, 2003.

Zhong, J., Lee, K., and Tsong, Y., "Statistical assessment of analytical method transfer," *Journal of Biopharmaceutical Statistics*, vol. 18, no. 5, pp. 1005–1012, 2008.