Real time release testing for antigen production process

Nusselder, A.C.

Award date:
2021

Department of Industrial Engineering & Innovation Sciences
Operations, Planning, Accounting, & Control

Master Thesis Report

Real Time Release Testing for Antigen Production Process

A.C. (Annick) Nusselder
0955826
a.c.nusselder@student.tue.nl

Supervisors TU/e
dr. T.G. (Tugce) Martagan
t.g.martagan@tue.nl
First Supervisor TU/e

prof.dr.ir. G.J.A.N. (Geert-Jan) van Houtum
g.j.v.houtum@tue.nl
Second Supervisor TU/e

prof.dr.ir. I.J.B.F. (Ivo) Adan
i.adan@tue.nl
Third Supervisor TU/e

Supervisor MSD
B.G. (Bram) van Ravenstein
bram.van.ravenstein@merck.com

Boxmeer
April 29, 2021

Subject headings: real time release testing, quality control strategy, quality prediction, joint optimization modelling, contamination prediction, yield prediction, quality testing, process monitoring, quality consistency, trending analysis

Abstract

In this master thesis the application of real time release testing (RTRT) within the bio-pharmaceutical industry is researched. A framework is formulated that describes a 6-phases methodology towards RTRT application. This framework is used to define RTRT approaches for yield and purity. A joint optimization model is developed to find the optimal acceptance and rejection criteria for critical process parameters for the classification of purity. For yield, a prediction model is constructed that uses growth rate, duration and starting biomass for each bacterial growth phase. The models are validated using an industry case study at MSD. This analysis shows for purity accurate classifications for early contaminations though the purity test remains superior. To increase accuracy and expand validation, it is recommended to introduce additional process parameters and conduct parallel testing on commercial scale. For yield comparative data analysis using the results obtained from the RTRT model and from the traditional yield test found that the RTRT model produces accurate results however it is inconclusive which method, RTRT or quality end testing, is superior. Therefore, it is advised to increase data availability by conducting further parallel testing and expand the measurement system analysis for the traditional yield test. Additionally, a dashboard is developed and implemented for the identification of product candidates for RTRT and for realizing process monitoring benefits resulting from RTRT implementation.

Acknowledgements

This master thesis is the last phase of the master 'Manufacturing System Engineering' at Eindhoven University of Technology. This work represents both the end of a wonderful university experience and the start of working life.

First, I would like to thank Tugce Martagan for mentoring me during this project and for being a role model. The weekly meetings were she shared her knowledge and insights have played an important role in the success of this research. In addition, i want to thank Geert-Jan van Houtum and Ivo Adan, my second and third supervisor, for their time and effort and for sharing their knowledge.

Secondly, i would like to express my gratitude for Bram van Ravenstein for providing me with the required process and product knowledge and helping me overcome the many challenges of this project. Bram has given me many opportunities to develop my professional skills and has supported my personal development. Moreover, I want to thank Rico Kersten, the improvement team, the BTS team and the production team for all their support, time and effort. They have taught me the MSD's company culture, showed me the many activities that are conducted at Boxmeer site and helped me with testing. Also, I want to express my gratitude to Niels de Haas who have me the opportunity to conduct my research at MSD. Also, I'm profoundly grateful for Oscar Repping and Marc Baaijens for their support during my project and for their inspiring drive for innovation.

Last but not least, I would like to thank my friends and family for being there during the difficult times but to celebrate the successes.

Annick Nusselder
Eindhoven, April 2021

Executive Summary

This research is conducted at the bacteriological production department at MSD Animal Heath Boxmeer. The summary will present the problem addressed, the methodology and, the main findings and recommendations for MSD.

Problem Statement

In order to become more cost efficient and meet the increasing demand, MSD Animal Health is interested in introducing Real Time Release Testing (RTRT). Additionally, it is expected that RTRT will show less test variation and as a result provide better quality assurance compared to quality testing as these biological tests include biological variation. According to the European Medicines Agency (EMA, 2012), RTRT is a release strategy that provides assurance of the intended quality of the product by using the information collected during the production process, through product knowledge and on process understanding and control, and can replace traditional quality testing. Importantly, there are two key issues that need to be resolved for MSD to implement RTRT. First of all, the availability of guidelines provided by the European Medicines Agency (EMA) and Food and Drug Administration (FDA) on RTRT is limited. This entails that at the moment it is unclear which conditions need to be met in order to get RTRT approved by EMA and what approach can be used to formulate the RTRT strategy. Secondly, the relation between the process parameters and the quality is not entirely clear. The question still remains which production criteria need to be fulfilled to ensure adequate quality or how process parameters can be used to predict quality. In addition, after more insight is gained on RTRT MSD is interested in wider application of this method and improving process monitoring after RTRT implementation. However, MSD struggles to identify product candidates for RTRT and to process RTRT data is such a way that benefits can be realized. This research is conducted for one specific antigen product on one production line for quality attributes yield and purity.

Methodology and Results

A general methodology for RTRT application is formulated using a 6-phased framework. This framework is used for the RTRT application for yield and purity. First, there is focused on risk assessment where the critical process parameters for yield and purity are identified. From case studies and expert knowledge it is concluded that especially real-time measured process parameters are suitable for predicting quality. The critical process parameters can be derived from the curve of real-time measured process parameters

by growth curve modelling. Two growth curve models are developed that expand the existing research in this field and enable the modelling of indirectly controlled process parameters and parameters that include a death phase.

The RTRT approach for purity is based on control limits for the critical process parameters that are used as acceptance and rejection criteria. Defining these control limits is a trade-off between waste and quality risk; when the control limits are too wide the probability of having a quality issue increases whereas when the control limits are too narrow the probability of having unnecessary waste increases. An optimization model is developed that determines the optimal control limits for the critical process parameter based on the risk of having a false negative or a false positive misclassification. This model is validated with production and experimental data including pure and contaminated batches. It is found that the model detects contamination accurately for batches contaminated early in the process however contaminations late in the process are hard to detect using the model. In comparison to the traditional purity test it can be concluded that the purity test is superior to the RTRT approach eventhough the purity test also experiences difficulties with detecting late contaminations. Comparing the performance of the optimization model to existing methods for defining control limits shows significant improvements resulting from the optimization model. Moreover, the analysis shows that the RTRT approach improves in the number of critical process parameters especially for critical process parameters with a strong positive relation to purity. In addition, the improvement of the optimization model compared to existing approaches increases when the weight for false negatives compared false positives increases.

For yield, RTRT is introduced by formulating a prediction method that estimates yield using process parameters measured during the production process. Authorities put emphasis on building a RTRT model based on a firm's understanding of the production process and of the relationships between the process parameters and yield. In addition, MSD expressed its desire to expand the model to other products therefore the model should be generalisable. A prediction model is developed that models the formation of or reduction in biomass for each bacterial growth rate using the growth rate, duration and starting amount of biomass for a phase. This model is validated using two products that are similar with regards to their production process but differ in terms of data availability. A comparative analysis with the predicted and measured yield shows that the prediction model is able to generate accurate predictions for yield especially for the product with high data availability. The comparison of the quality testing method and RTRT method is inconclusive which is most likely a result of low data availability especially for the measurement system analysis for the yield testing method.

Looking at RTRT strategy in a broader perspective considering the future of RTRT, two questions arise "which products are suitable for RTRT?" and "how can we benefit for RTRT implementation?". A dashboard is developed and implemented that functions both as a decision-support tool and a process monitoring tool. Even without actual RTRT implementation MSD can benefit from this process monitoring tool because of the significant reduction of 33 to 50% in information leadtime realized by the implementation of this dashboard. The decision support tool provides insight on which products are suitable candidates for RTRT by evaluating the quality consistency of products using trending and out-of-spec analysis. The

process monitoring tool helps to process quality data such as RTRT data and gain insight in process behavior and performance. This tool supports departments by identifying out-of-spec observations which could require adaptions to production schedules and detecting observed process issues and potential future process issues. The implementation insights show that this dashboard can improve return on investment and decrease inventory and lost sales.

Recommendations

Based on the conclusions of this research five main recommendations are formulated.

Implement growth curve modelling for all process parameters real-time during the production process It is recommended to implement growth curve modelling for all process parameters measured real-time to obtain the required input for the RTRT models and to conduct quantitative analysis that can support process improvement and problem solving. This implementation could best be conducted in two-phases: (1) starting with growth curve modelling implementation for the product and critical process parameters used in this research for the purpose of parallel testing for RTRT and (2) implementing automatic growth curve modelling for all products and all process parameters on all production lines after the fermentors allow for automatic data processing.

Conduct parallel testing for purity and yield using the developed RTRT strategies The RTRT approach for purity has been validated using contamination data on experimental scale instead of commercial scale. For further model validation, the recommendation is to conduct parallel testing on commercial scale where misclassifications should be used to update and improve the model. Moreover, it is advised to continue parallel testing for yield using the prediction model to generate substantial comparative data to ensure the performance of the RTRT strategy.

Introduce additional process parameters for purity and test other contaminating bacteria The accuracy of the purity model could be improved by finding additional process parameters that are indicative for the presence of contaminations. Therefore, a two-fold recommendation is proposed: (1) expand the RTRT approach to products that have more contamination data and (2) conduct research on potential additional process parameters that can be used to predict purity. On commercial scale process parameters are measured that have not been included in this research due to experimental limitation but may be critical to purity. Products with more contamination data on commercial scale can be used to analyze these additional process parameters. In addition, to classify purity more contaminating bacteria should be researched as the difference in growth characteristics of these bacteria can influence their detectability.

Expand RTRT approaches to other products The advice is to use the yield dashboard to identify candidates for RTRT application for yield. It is recommended to use this dashboard as it provides a data-driven decision on a product's suitability for RTRT which supports the approval process for RTRT and can improve return on investment.

Develop action plans for trending analysis The recommendation is to formulate action plans for the trends included in the yield dashboard. This action plan should describe checks that should be done

to determine if there is a problem and, if a problem is observed, formulate analyses that can be conducted to identify the cause and determine the solution direction.

Contents

Abbreviations

CPP: Critical Process Parameters

CQA: Critical Quality Attribute

DSP: Downstream Process

EMA: European Medicines Agency

FDA: U.S. Food and Drug Administration

GMP: Good Manufacturing Practices

IPC: In Process Control

MSD: Merck Sharp & Dohme

OD: Optical Density

QP: Qualified Person

RTRT: Real Time Release Testing

STRs: Stirred Tank Reactors

USP: Upstream Process

Chapter 1

Introduction

The bio-pharmaceutical industry for animal health focuses on producing vaccines and antibiotics that help maintain the health and welfare of animals. Animals are our companions and play an important role in the food supply therefore animal health is not only important for animals but also for us. This research is conducted at the company Merck Sharp & Dohme (MSD) Animal Health, a leader in the bio-pharmaceutical industry. In Section 1.1 background information on the company MSD is provided and Section 1.2 describes the production process of MSD relevant to this research. Subsequently, Section 1.3 discusses the problem that will be addressed and Section 1.4 presents the goal of this research and the research questions that will be answered. Than, Section 1.5 describes the scope of this research. Lastly, the outline for this report is presented in Section 1.6.

1.1 MSD Animal Health

Merck Sharp & Dohme (MSD) Animal Health is a world leader in the development, production and sale of Veterinary medicines. Their wide range of veterinary pharmaceuticals, vaccines and even health management solutions and services is offered to veterinarians, farmers, pet owners and governments in more than 150 markets. The portfolio of species for which MSD provides its products includes companion animals, equine, poultry, swine, aquaculture and ruminants. MSD Animal Health is dedicated to preserving and improving animal health and thereby support safe and affordable food supplies and help pets live longer, healthier lives.

MSD Animal Health was founded in the 1940s when MSD decided to expand from only being active in human pharmaceuticals to also offering Veterinary medicines. Soon the first vaccine, a poultry coccidiostat, called sulfaquinoxaline, was discovered. Throughout the years, many vaccines and antibiotics are developed to treat and prevent illness in animals, from the first recombinant DNA vaccine in the 1980s to treat diarrhea in piglets, to the development of Canine Influenza Vaccine H3N2, a contagious respiratory disease in dogs, in 2015.

MSD Animal health is present in more than 50 countries. Boxmeer and De Bilt are the production and research locations in The Netherlands. Currently, the establishment at Boxmeer is world's largest

production location within pharmaceutical industry when it comes to Veterinary vaccines. Here, all activities needed for vaccine development take place: (1) research and development, (2) production and quality control and (3) logistics, marketing and sales.

The vision of MSD Animal Health is to provide innovative solutions to improve animal health. MSD wants to support the bond between people and their pets, and help protect international public health, ensure food safety and increase protein supplies. Research operations are focused on the development of vaccines and pharmaceuticals for today's most challenging diseases and the improvement of existing production processes in order to meet changing market needs.

This research was conducted at the bacteriological processing department MSD Animal Heath at production site Boxmeer.

1.2 Overview of Biomanufacturing Operation

The production process with the bacteriological process department is composed of the upstream process (USP) and the downstream process (DSP). This process is illustrated in Figure 1.1. USP is used for cultivating the bacteria; this process is also called the fermentation process. After the fermentation process, the function of the DSP is to concentrate and purify the product obtained from USP.

In the USP, the first step is to produce the pre-culture (the inoculum process); the working seed is fed with the suitable medium. The cells of the working seed will start to divide and grow. From there the fermentation process is conducted, the pre-culture is transferred into fermentors (stirred tank reactors (STRs)) in which the main fermentation takes places. Within these STRs, the environment is highly controlled facilitating the cell growth. After a defined period, the fermentation process will be stopped by adding inactivation chemicals or lowering the temperature. Resulting from the fermentation is a mixture of the biologics of interest, referred to as antigen, and unwanted impurities.

The culture resulting from the USP is transferred to the DSP by pipes, also called closed system. The DSP is used to separate the unwanted impurities from the antigen, for example by centrifugation. The order and combination of DSP activities is dependent on the characteristics of the desired end-product.

After downstream processing, the end product is tested for its quality attributes by quality control. This is a separate department that evaluates whether or not the end product meets the quality criteria for release. Typical quality attributes that are tested are inactivation, sterility and potency.

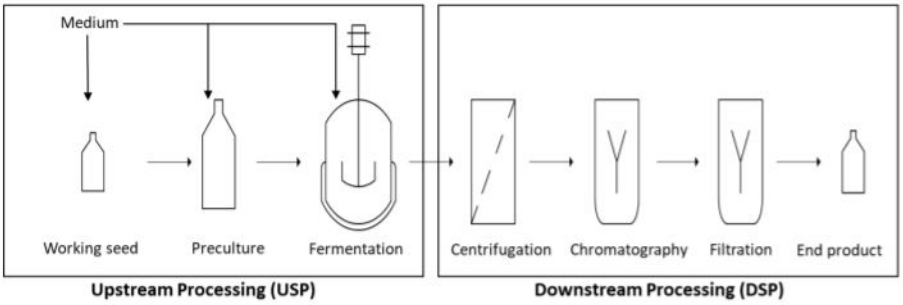


Figure 1.1: Antigen production process (Martagan et al., 2020)

For a typical fermentation process, six phases of bacterial growth can be identified: lag phase, acceleration phase, exponential growth phase, deceleration phase, stationary phase and death phase (see Figure 1.2) The growth process starts with the lag phase; a period of adaption of the cells to their new environment. In the acceleration phase, cells have adjusted to their environment and cell growth begins. The cell growth continues through the exponential growth phase where an exponential increase in the number of cells is observed. Depletion of one or more essential growth nutrients and accumulation of toxic growth associated by-products causes the cell growth to slow down; this is called the deceleration phase. When the number of cells dividing and dying is at equilibrium, the cell growth enters the stationary phase. The death phase starts when the rate of cells dying is greater than the rate of cells diving and the number of viable cells starts decreasing.

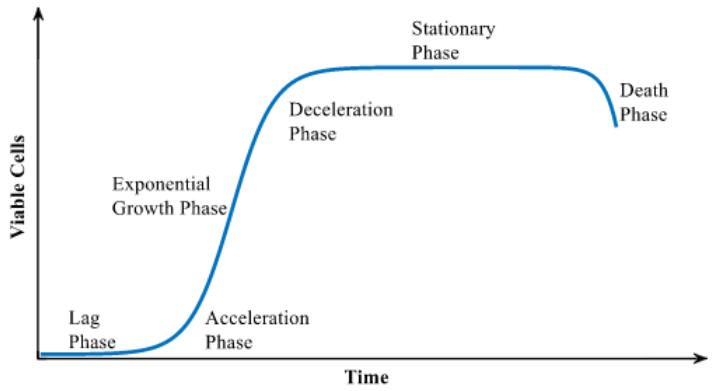


Figure 1.2: The six phases of cell growth during fermentation (Martagan et al., 2020)

1.3 Problem Statement

In terms of regulatory requirements producing human drugs and animal drugs are similar. However, these drugs differ in the allocation of the cost. Human drugs are often subsidized and paid for by insurance companies whereas animal health products are paid for by the animal owners. Therefore, offering an acceptable price that is justifiable compared to the value of a healthy animal products is of great importance. This emphasizes the need for animal drugs manufacturers to be as cost efficient as possible.

Furthermore, over the years, the demand for animal health products has increased substantially and is expected to increase even more. In the period of 2020-2027, market analysts anticipate that the global animal health market size will exhibit a compound annual growth rate of 5.8% (Grand View Research, 2020). The main driver of this growth is the rise in the demand for protein food. The growing human population has resulted in an increase in the human protein consumption. Since animal-based food products count for 33% of this protein consumption, meeting these dietary needs asks for greater rearing of livestock and a focus on veterinary healthcare. Both contribute to the growth in the demand for animal drugs.

In order to become more cost efficient and meet the increasing demand, MSD Animal Health and other bio-pharmaceutical companies are interested in introducing Real Time Release Testing (RTRT). Additionally,

it is expected that RTRT will show less test variation and as a result provide better quality assurance compared to quality testing that includes biological variation. According to the European Medicines Agency (EMA, 2012), RTRT is a release strategy that provides assurance of the intended quality of the product by using the information collected during the production process, through product knowledge and on process understanding and control. Importantly, there are two key issues that need to be resolved for RTRT to be applied. First of all, the availability of guidelines provided by the European Medicines Agency (EMA) and Food and Drug Administration (FDA) on RTRT is limited. This entails that at the moment it is unclear which conditions need to be met in order to get RTRT approved by authorities (EMA/FDA) and what approach can be used to apply a RTRT strategy. Secondly, the relation between the process parameters and the quality of the output (antigen) is not entirely clear. The question still remains which production criteria need to be fulfilled to ensure the desired result for a dichotomous quality attribute or how critical process parameters can be used to predict the outcome of numeric quality attributes. Besides the challenges for RTRT application, MSD and other bio-pharmaceutical companies struggle to identify product candidates for RTRT application and encounter difficulties in processing RTRT data in such a way that they can benefit from it with regards to process monitoring and, demand and supply alignment.

Therefore, more research should be conducted on the guidelines for RTRT and the required validation to get approval, so a framework can be formulated that will function as guideline for RTRT application. Also, it should be determined which process parameters influence the quality attributes and how the relation between process parameters and quality attributes can be modelled. Based on that information, criteria for the relevant process parameters can be formulated that ensure that the quality of the output or for numeric quality attributes a prediction model can be formulated based on the relations between process parameters and the quality attribute. The formulation of these RTRT approaches and further validation of the proposed approaches enables the application of RTRT. In addition, a method should be constructed that assists in identifying products that are suitable for RTRT. Lastly, an approach for processing RTRT data should be proposed that formulates appropriate analyses that can be used to obtain insights in process behavior.

1.4 Research Goals

Section 1.3 describes the problem experienced within MSD. To address this problem first more insight is provided on the current process flow at MSD illustrated in Figure 1.3. For this research we are interested in RTRT application for both numeric and dichotomous quality attributes as the approaches for RTRT differ between these attribute types. It is decided to focus this research on yield, a numeric quality attribute, and purity, a dichotomous quality attribute, since these attributes are both measured before downstream processing for the product in scope. Section 1.5 elaborates on the motivation for this limitation in scope. The upstream process, downstream process and quality control are explained in Section 1.2. Yield is measured after the upstream process is completed. This yield result provides the necessary input for the downstream process. Therefore, the yield test needs to be conducted and the result has to be known before downstream processing can start. Purity is also determined after the upstream process is completed

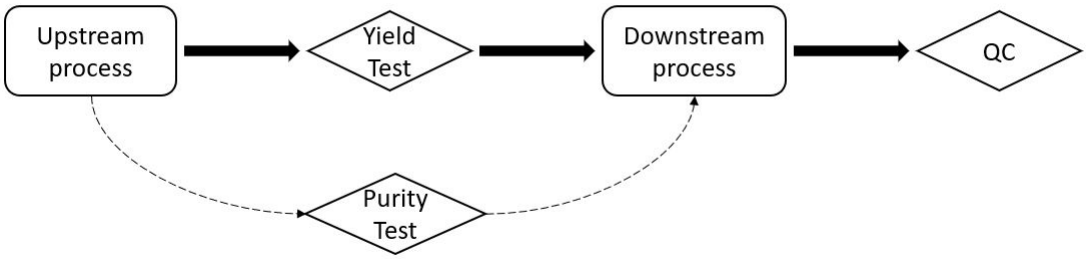


Figure 1.3: Current process flow with upstream process, downstream process and quality control (QC) in relation to the yield and purity test

however since this test takes a couple of days downstream processing is started before the test result is known. The quality attributes measured by QC are not considered in this research but in the future RTRT may also be applied for these tests as most gain is in replacing these tests because they have the longest leadtime.

With applying RTRT, yield and purity are based on data from process parameters measured during the upstream process as a result the result for these quality attributes is available real-time and the leadtime for obtaining the quality attribute results becomes zero. Figure 1.4 illustrates the process flow based on RTRT application for yield and purity. Changing the quality control strategy from quality end testing to a RTRT strategy, or any change to the production process for that matter, requires besides an appropriate formulation of the RTRT approach also the approval of European Medicine Agency (EMA) or, for the USA approva,l of the Food and Drug Administration (FDA). As this research is conducted within the Netherlands, we focus on the EMA. The role of the EMA is to facilitate the development of medicines and to monitor and control the safety and effectiveness of medicines, all with the higher purpose to protect human and animal health. The monitoring and controlling of the safety of medicine is done on two fronts: (1) with the evaluation of an application for a new medicine including its quality control processes and (2) by the control of major changes to any of the processes, production and/or quality related. Replacing the traditional quality control process by a RTRT approach is seen as a major change to a quality related process. This means that formulating a RTRT approach is the first step towards RTRT application though actual implementation as shown in Figure 1.4 requires an extensive approval process. This study will address the problem discussed in Section 1.3 to the point where the approval process from the EMA starts. This process often takes several months/years which exceeds the timeline of this research.

This study aims to tackle the problem discussed in Section 1.3 by formulating and answering the following main research question that will be answered by deriving six sub questions. This main research question addresses both the need for a guideline of RTRT application and the formulation of a RTRT approach and the need for direction towards broader RTRT application and using the gained insights from RTRT.

How can real time release testing be applied for quality attributes yield and purity within

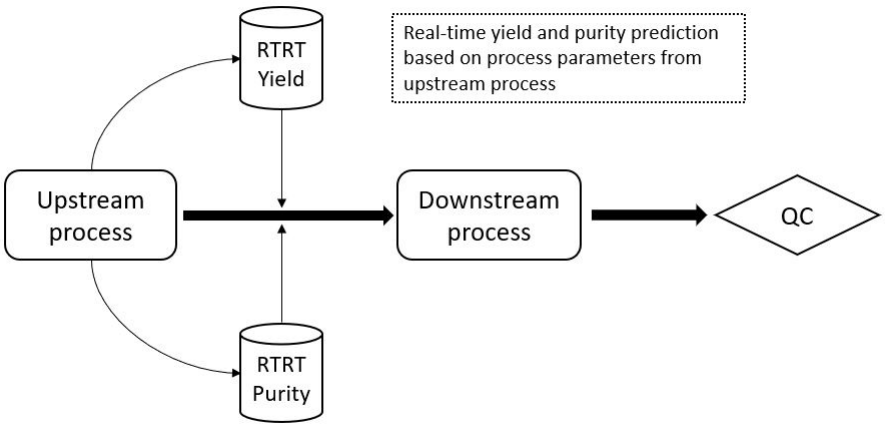


Figure 1.4: Process flow based on RTRT with upstream process, downstream process and quality control (QC) in relation to the yield and purity test

the production processes at MSD ensuring adequate product quality and how can real time release be further expanded and benefited from?

Currently, quality attributes are measured using quality testing. With the introduction of RTRT, quality tests can be replaced by RTRT approaches that determine the values for the quality attribute by using process parameters measured during the production process. The EMA has provided some guidelines and requirements for RTRT that need to be met in order to obtain approval for implementing this strategy and replacing traditional quality testing. However, these guidelines and requirements are limited and do not provide a great level of detail. In addition, as RTRT is a novel strategy limited to no literature is available describing a methodology for RTRT application or formulating a RTRT approach; the limited work on RTRT does not discuss the method used to come to the formulated RTRT approach. For this research, but also literature has expressed a need for a methodology that describes which steps need to be conducted when formulating a RTRT approach such that the requirements provided by the EMA are met and when completing all steps of the methodology the process for gaining approval for implementing RTRT can start (Torres, 2017). The aforementioned is addressed by the first sub-question that focuses on developing a method that can be used as guide towards applying RTRT not only for this research but for RTRT application in general. Chapter 2 will answer this research questions by reviewing existing literature on RTRT and combining the insights gained to formulate a method for RTRT.

1. What general methodology can be used for introducing RTRT?

After formulating the method for RTRT, the next step is analyze the process parameters measured in the production process. RTRT is a control strategy based on process parameters measured during the process that are found critical for the analyzed quality attribute. These can be either process parameters measured off-line only once during the process or process parameters that are measured on-line over the whole duration of the process, also referred to as real-time measured process parameters. Real-time measured process parameters do not have a single value for a particular batch but have a set of longitudinal data including millions of values for this parameter. This longitudinal data needs to be processed in

order to be used for RTRT application. Growth curve modelling is a method proposed in literature for processing longitudinal data however this method is limited in its applicability. As a result, various real-time measured process parameters cannot be analyzed with this method and can therefore not be used as critical process parameter for RTRT. For RTRT, broader applicability is desired because it allows for more process parameters to be analyzed and function as input for the RTRT model. Therefore, sub-question 2 is formulated which will answered in Chapter 3.

2. How can real-time process parameters be processed to obtain potentially critical process parameters that can serve as input for RTRT approaches?

Obtaining the critical process parameters for RTRT is addressed by sub-question 2, the next step is formulating the RTRT approach. According to EMA (2012), the RTRT approach for a dichotomous quality attribute such as purity can be formulated using acceptance and rejection criteria for the critical process parameters. If the value for the critical process parameter meets the acceptance criteria, the batch is classified as pure. To the best of our knowledge there is no literature that discusses how these acceptance and rejection criteria can be determined and what the relation is with quality assurance. In addition, no studies exist that capture the relation between the acceptance and rejection criteria and the consequences on purity classification in a mathematical way. Having a justification for the criteria of the RTRT approach is mentioned by the EMA as one of the requirements for RTRT. The formulation of the acceptance and rejection criteria is a trade-off between quality risk and waste risk each having financial and quality impact. Understanding this trade-off and capturing the relation between quality assurance and, acceptance and rejection criteria is essential for the justifying the formulation of the RTRT approach for purity. In order to understand this trade-off and define a RTRT approach for purity, sub-question 3 is formulated. This question is addressed in Chapter 4.

3. How can the purity of a batch be classified using acceptance and rejection criteria for process parameters?

The RTRT approach for a numeric quality attribute, in this research yield, estimates the quality attribute value using process parameters critical to that quality attribute. The formulation of a RTRT for a numeric quality attributes using process parameters is a subject rarely discussed in literature. The work dedicated to this subject estimates the quality attribute using mathematically complex models that are difficult to interpret. In the requirements from the EMA it is explicitly mentioned that the understanding of the relations between the quality attribute and the process parameters used for RTRT prediction needs to be demonstrated. These mathematically complex models do not allow for demonstrating this understanding as the relations within these models are data-driven and cannot be explained by biological behaviour. To the best of our knowledge there has not yet been a literature work that discusses a RTRT approach that is approved by the EMA or the FDA for yield or any relating numeric quality attribute. The formulation of a RTRT approach for yield that is build upon the relations between the yield and relevant process parameters is difficult due to the complex biological processes happening during the production process. Sub-question 4 is formulated to address this challenge and formulate a RTRT approach that demonstrates the understanding of the relationships between the included process parameters and yield.

This sub-question is discussed in Chapter 5.

4. How can yield be predicted using a model that is based on the relations between process parameters and yield?

The methodology for RTRT introduction and the formulation of the RTRT approaches is discussed for the product in scope. In the future, MSD is interested in expanding RTRT to wider range of products. In the requirements for RTRT it is mentioned that a product is suitable for RTRT if the production process for this product is able to consistently deliver the desired quality (EMA, 2012). To this moment, it is unclear how it can be determined if products meet this requirement. Studies on RTRT have not focused on developing a methodology for proving if a product meets this requirement or , for case studies, on describing how the product in scope has been chosen. To address aforementioned, sub-question 5a is formulated which is answered in Chapter 6. This chapter will explain how products can be selected based on meeting the requirement for RTRT provided by the EMA.

5a. How can products that are suitable for RTRT application be identified?

Eventhough implementing RTRT requires the approval of the EMA, MSD is interested in how they can benefit from the opportunities that arise when using RTRT as quality control strategy. These opportunities are especially in the field of process monitoring because RTRT allows for obtaining quality results in real-time which provides direct feedback on the performance of the process. As RTRT application is limited to none, it is unclear to MSD how they can process RTRT data is such a way that they can benefit from it and improve their process monitoring. This challenge is addressed by sub-question 5b and answered in Chapter 6. In this chapter a tool is designed for processing RTRT data with the purpose of improving process monitoring.

5b. How can RTRT application be used to improve process monitoring?

To summarize, the linkage between the 6 sub-questions is illustrated in Figure 1.5. Sub-question 1 provides a general methodology where sub-question 2 discusses a specific step of this methodology. Sub-questions 3 and 4 use the methodology of sub-question 1 to formulate a RTRT approach for purity and yield, respectively. Sub-question 5a and 5b discuss RTRT from a future perspective by identifying RTRT product candidates and explaining how process monitoring benefits arising from RTRT can be captured.

1.5 Scope

The scope of this research is limited to the RTRT application for one product. The production processes differs from one product to another, especially for the downstream process. It is decided to focus on one product because RTRT is a new concept and in-depth process and product understanding are essential for this research. The biological process at MSD is very complex so in order to fully understand the production process for a specific product takes time. By focusing on one product, there will be more time to gain process and product understanding. This research can serve as guideline for RTRT application for other products in the future. The product is selected based on both the stability of the production process and the availability of historical batch data.

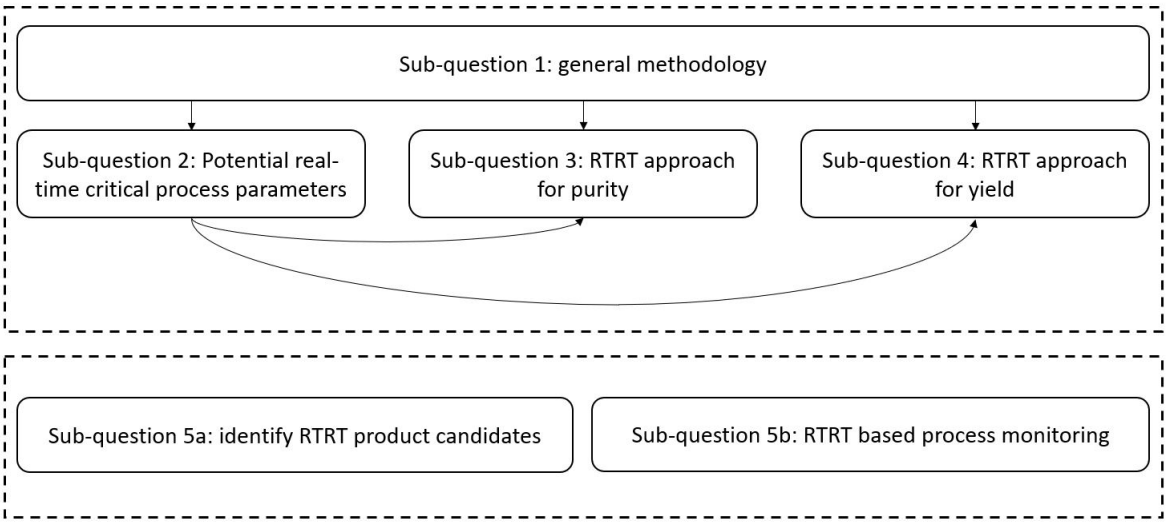


Figure 1.5: Schematic overview of linkage between the sub-questions

In addition, it is decided to limit this research to one production line on which the product in scope is produced because the process characteristics can differ between production lines. The decision for the production is based on data availability.

Moreover, the research is limited to RTRT for quality attributes that are related to upstream process characteristics and are measured before the start of the downstream process which, for the product in scope, are quality attributes yield and purity. It is decided to only focus on the upstream process because the number of process parameters measured in the downstream process is limited and research on quality attributes only related to upstream processing does not require resources from the quality control department.

Furthermore, studies have shown that medium used in the upstream process has influence on the cell growth. The quality of the raw materials used for producing the medium is precisely measured however the quality some of the raw materials, for example animal based products, cannot be measured precisely. Animal based products or other living materials will always show some variation and their quality can only be roughly measured. Therefore, the exact quality of the medium will stay undefined. For this reason, the medium and the influence of its quality on the quality of the output is out of scope. One must remember that this implies that there will be some unexplained variation in output quality.

1.6 Outline

This research starts with formulating a methodology for RTRT which is presented in Chapter 2. Than, Chapter 3 describes growth curve modelling as method for analyzing real-time measured process parameters. Subsequently, Chapter 4 discusses the RTRT approach for purity. In this chapter a mathematical model is constructed that determines the optimal control limits based on the trade-off between quality risk and waste risk. In Chapter 5 the RTRT approach for yield is described and the prediction model for estimating yield based on process parameters is presented. Furthermore, Chapter 6 presents a dashboard that functions as both a decision-tool for identifying product candidates for RTRT and as process-monitoring

tool for processing RTRT data to seize the benefits from RTRT application. Lastly, the conclusion and recommendation are discussed in Chapter 7 and 8, respectively.

Chapter 2

Framework for Real Time Release Testing

This chapter provides background information on RTRT relevant for this research. Although the research on RTRT is limited, there are some authority documents and case studies discussing this control strategy. The requirements and guidelines mentioned in these documents and the methods explained in casestudies can be valuable input for this research. It can provide insight in which steps to conduct and what aspects of RTRT and model validation have to be discussed to comply with authority requirements. Based on these insights, a general framework can be constructed that explains a broad methodology for the application of RTRT.

First, Section 2.1 discusses a literature review on the requirements and guidelines provided by authorities on RTRT or relevant concepts. Subsequently, Section 2.2 elaborates on casestudies discussing RTRT that can be used as inspiration for this research. At last, a framework is presented in Section 2.3 that can be used as methodology for the application of RTRT. Also, this section explains which parts of this framework are discussed in this research.

2.1 Requirements and Guidelines

The European Medicines Agency (EMA) and the Food and Drug Administration have provided several documents discussing the requirements for using RTRT, or parametric release, to provide assurance of product quality. In 2012, a guideline for RTRT was published by EMA addressing the guidelines for the application of RTRT (EMA, 2012). This guideline was based on a revision of an earlier guideline on Parametric Release. Unlike the guideline of Parametric Release which has its main focus on terminally sterilized products, the guideline in question focuses on active substances, intermediates and finished products in the context of chemical and biological processing. According to EMA (2012), the framework for RTRT is provided by the demonstration of enhanced product and process understanding, the formulation of a pharmaceutical quality system and the integration of quality risk management principles. The last two referring to quality management are further defined in ICH 8, 9 and 10 which will discussed later on

in this section. A firm should be able to demonstrate its understanding of the production process and of the relationships between the process parameters, material attributes and quality attributes present in this process. These relationships should be justified and verified by sound scientific data. Based on this acquired knowledge, a RTRT strategy can be formulated that uses a set of in-process test and controls for monitoring quality attributes and controlling the process parameters. Moreover, EMA states that in case of failure (or trending towards failure) an investigation should be started to identify the cause, and product end testing cannot be used as substitution for RTRT when, according to RTRT acceptance criteria, failure occurs.

EMA (2012) does not specify in detail which requirements apply for the submission of RTRT as manufacturing processes for substances and dosage forms differ in whether they are discrete or, partially or wholly continuous. The general criteria for authorization of RTRT mentioned by EMA are:

- Identification of critical quality attributes using development studies
- Execution of risk-based development program
- Development and implementation of control strategy
- Execution of validation of manufacturing process
- Demonstration that process requirements for approval/rejection are based on the defined acceptance criteria in development studies
- Justification of the acceptance criteria and demonstration of relation between end-testing and RTRT supported by comparative test results
- Formation of procedures for reporting and required actions to be performed approval/rejection
- Demonstration of acceptable quality of the applied technologies
- Demonstration that RTRT is equivalent or better in providing quality assurance compared to product end testing
- Formulation of contingency plan

The validation of the real time release test method should be supported by substantial comparative data at commercial scale, for example parallel testing, to demonstrate the relation between the real time release test method with acceptance criteria and the end product testing method. The contingency plan should be included in the control strategy and specify alternative testing methods, such as end-testing, or monitoring approaches in case of equipment failure. In addition, EMA recommends a close collaboration with both inspectors and assessors as their input plays an important role in the approval process. The role of the assessor is the assessment of product related issues and the role of inspectors is focused on system related issues.

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is a organisation that brings together regulatory authorities, such as the EMA, and companies active in the pharmaceutical industry to establish guidelines for the scientific and technical features of

pharmaceuticals (EMA, 2004). As stated in EMA (2012), the ICH guidelines that are relevant for RTRT are ICH 8, 9 and 10. ICH 9 and 10 are more complementary and not directly relevant for constructing an approach or framework for RTRT but may be useful in later stages when RTRT is implemented. In the guideline ICH 8, EMA (2004) discusses the criteria for pharmaceutical development and elaborates on a systematic approach to development. This guideline also elaborates on the connect between product development and RTRT. With pharmaceutical development the goal is to create a quality product and develop a manufacturing process for this product that can consistently deliver the specified product quality. Pharmaceutical development studies can also be used to enhance product knowledge and process understanding which facilitates the realization of RTRT. ICH 8 suggests the use of a systemic approach to development as such approaches support regulators in understanding a company's applied method and can improve the delivery of the intended product quality. The approach that is proposed is the Quality by Design (QbD) approach: " a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management. " (Yu et al., 2014). At minimum, a pharmaceutical development should consist of: (1) the definition of the quality target product profile, (2) the identification of the CQAs of the drug product and the CQAs of the drug substance and excipients, (3) the selection of a suitable manufacturing process and (4) the definition of a proper control strategy. Additional to these aforementioned elements, the QbD approach mentions the following elements:

- A systematized analysis of the formulation and manufacturing process:
  - Identification of the material attributes and process parameters that affect the CQAs
  - Formulation of the relationships that link the material attributes and process parameters to the CQAs
- Establish a control strategy based on a combination of enhanced product and process understanding and the assessment of risk (this strategy for control could be RTRT)

RTRT is mostly related to the formulation of a control strategy as a strategy for control could be the implementation of RTRT however it is also linked to other elements of the pharmaceutical development process and QbD approach. For example, for RTRT the critical quality attributes should be defined and a systematized analysis of the manufacturing process is needed to identify the relations between process parameters and material attributes, and critical quality characteristics.

FDA (2010) has written a guideline on parametric release for terminally sterilized products. Parametric release is a concept which is almost identical to RTRT, therefore the requirements discussed on this document can also be used as input for the framework in this research. The FDA approval for parametric release will be based on the firms ability to address the risks to sterility and the reliability of the control strategy. The risk assessment should be in line with the principles of EMA (2006b) and focus on the risk of not meeting the minimum required probability of non-sterile unit for each unit of every batch. Overall, the documentation for parametric release has to include similar elements as mentioned in the EMA (2012). The document of FDA (2010) adds that a sterility test cannot overrule the outcome of the RTRT; if the

acceptance criteria are not met, the product will not be released. In addition, Tirumalai & Porter (2005) have discussed requirements for the validation and documentation for parametric release. They emphasize the justification of the critical process parameters and the operational ranges, the establishment of the relation that links a critical process parameter to a certain quality attribute and the verification that the in-process requirements for approval and rejection are based on the acceptance criteria and verified by validation studies.

2.2 Casestudies

Case studies may also mention criteria for RTRT and the methodology used in these case studies can be a source of inspiration for this research. The article of Torres (2017) describes a case study on RTRT for a drug substance manufacturing process by Biogen; the product end-testing for the quality attributes appearance, purity and biological activity have been eliminated for the drug substance by the use of RTRT. In this case study, a combination of increased understanding of the product and process, and quality risk management is used to implement the RTRT strategy. The enhancement in process understanding has been accomplished by development studies and by real-time data obtained from production. According to Torres (2017), QbD and ICH quality guidelines played a key role in the realization of RTRT because enhanced product and process knowledge, crucial for RTRT, are fundamental in QbD and these ICH guideline. Unfortunately, the methodology used to conduct this case study has not been discussed in detail. In general, the steps in this research are: (1) select testing points based on their governance or control of quality attributes, (2) determine the action limits for control and (2) gain manufacturing experience to test performance. The case study of Singh (2015) discussing a tabletting manufacturing process at Bristol-Myers Squibb also used the QbD approach for the realization of RTRT. In addition, the QbD approach is implemented by Pawar et al. (2016) with as goal understanding the changes in process parameters on tablet dissolution to RTRT can be used instead of end-product testing. First, the process parameters that have impact on quality have been determined whereafter a principal components analysis has been conducted with these process parameters as input. Subsequently, a multivariate linear model is constructed to investigate the relation between the eigenvalues of the principal component analysis and the dissolution profiles. With this regression model, the dissolution profiles can be predicted without performing a destructive test using parameters measured during the production process.

2.3 Framework for Real Time Release Testing

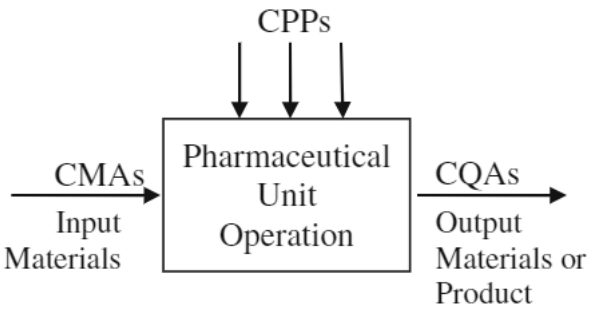
The insights from Section 2.2 and 2.1 can be used to construct a general framework for RTRT. This framework can be used to formulate the methodology for this study or help to inspire future work on RTRT.

The framework illustrated in Figure 2.2 and includes all steps necessary to comply with the known requirements for RTRT. The framework is based on the QbD approach discussed in ICH 8 (EMA (2004)). This approach is complemented with information on QbD obtained from other sources and with the documentation on requirements for RTRT. Also, as shown in Figure 2.2, the framework is iterative meaning it is possible to adapt or extent an earlier step based on knowledge obtained in a later stage.

Starting with the Quality Target Product Profile (QTPP), QTPP, often skipped for existing products, provides a summary of all probable characteristics of a drug that relate to quality bearing in mind safety and efficacy and should be achieved to reproducibly deliver a quality drug (Gandhi & Roy, 2016).

From this QTPP, a subset of quality attributes critical (CQA) to quality is selected based on the extent to which variation in this attribute out of its acceptable range can result in a drug of unacceptable quality (Umesh, 2012). CQA's are potentially altered by changes in the input parameters or process parameters (Gandhi & Roy, 2016). A proper method for identifying CQAs is the CTQ flowdown used in the DMAIC methodology where safety and efficacy are linked to lower level performance indicators. As the RTRT strategy should be at least as accurate as the current testing method, it is advised to conduct a measurement system analysis for the identified CQAs (Runje, Novak, & Razumić, 2017).

As discussed by Gandhi & Roy (2016), the CQAs are potentially altered by changes in input materials and process parameters. The goal of the risk assessment is to identity the material attributes and process parameters that are critical to the CQA of interest (see Figure 2.1). Critical meaning a realistic change in the attribute or parameter can result in noncompliance with the quality standard. For that reason such parameters should be controlled in order to produce a product of the desired quality or can be used to predict the quality of a product. The risk assessment starts with identifying all material attributes and process parameters present in the production process. These attributes and parameters are ranked based on their influence on the CQAs. This ranking can be done by using prior knowledge or by analyzing historical data or initial experimental data. It is also important to discuss why certain attributes and/or parameters are not critical, for example, because they show no variation. As the critical material attributes and process parameters have been identified, the operational ranges should be determined and the measurement systems should be evaluated. Also, the potential interactions between critical material attributes (CMAs) and/or critical process parameters (CPPs) should be researched (Gandhi & Roy, 2016).



$$CQAs = f\,(CPP_1, CPP_2, CPP_3 \ldots CMA_1, CMA_2, CMA_3 \ldots)$$

Figure 2.1: The link between the critical material attributes, critical process parameters and the critical quality attributes of a unit of operation (Yu et al., 2014)

Following the risk assessment is the design space which maps the multivariate relations between the CMAs and the CQAs, and the CPPs and the CQAs. This mapping is based on prior knowledge, experiments and historical data and can be presented graphically, with operational ranges or with models and mathematical

equations. A design space can be modelled separate for each unit of operation or can be for the whole process; the former is easier to develop and the latter provides more operational flexibility.

A control strategy is formulated in order to consistently produce a product of the intended quality and should be based on process and product understanding. In this control strategy, appropriate acceptance criteria for the CPPs and CMAs, which are in-process requirements for the approval of the product, are formulated for which it is be assumed that the product if of the intended quality. These criteria are based on the relationships defined in the design space; validation studies are needed to verify these acceptance criteria. Another strategy is to build a model that can predict quality using CPPs and CMAs that way concluding whether or not a product is of the intended quality based on the outcome of the model. It is important to integrate quality risk management in the control strategy by showing which risks can be controlled and which risks cannot. The risks that cannot be controlled should be either mitigated or accepted if they are acceptable in terms of severity and probability of occurrence. Furthermore, the control strategy should provide proof on the relationship between the end-testing currently used and the RTRT acceptance criteria or the prediction model showing that RTRT strategy is equivalent or superior to end-testing. The measurement system analysis for the CQA can be provide the basis for proving the equivalence or superiority of the RTRT control strategy. Important to note, the control strategy can include a combination of RTRT and the traditional end-product testing. For instance, one CQA is controlled using a test that is conducted on the end-product, whereas the other CQA is controlled using in-process requirements for the CPPs (that relevant for that specific CQA).

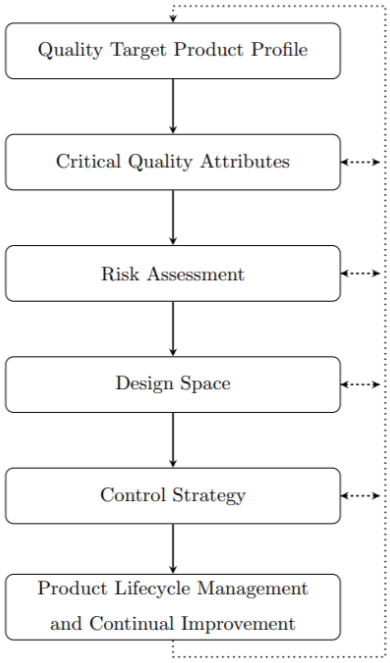


Figure 2.2: Framework for RTRT Realization

Product life cycle management and continual improvement includes activities such as model maintenance and parallel testing. The latter is when end-testing and RTRT control are simultaneously used to evaluate the performance of the RTRT control strategy and to proof that the the conclusions based on end-testing are compliant with those based on RTRT.

The scope of this research is limited to the steps of risk assessment, design space, control strategy and, product life cycle management and continual improvement steps. The QTPP is already documented for the product in scope and critical quality attributes are known.

Chapter 3

Growth Curve Modelling for Real-Time Measured Process Parameters

This chapter discusses the formulation of the growth model which is used to obtain potential critical process parameters from longitudinal data of real-time measured process parameters. The purpose of the growth model is used to quantify characteristics of the curves of real-time measured process parameters where the curves resemble the bacterial growth curve. These characteristics, also called curve parameters, can be used as potential critical process parameters for the yield model and the purity model. First, Section 3.1 explains the linkage between the growth curve model and, the yield and purity model. Subsequently, Section 3.2 discusses the relevant literature on growth curve models and the contributions of this study to this field of research. The mathematical model is presented in Section 3.3. Than, Section 3.4 describes the industry case study conducted at MSD used to validate the growth curve models in practice. At last, Section 3.5 presents the conclusion and discusses the insights gained.

3.1 Introduction

With the evaluation of treatment effects, developmental or analyzing dairy data there is a focus on analyzing change over time. A suitable method for describing change over time is growth curve modelling, also called latent curve analysis. Growth curve modelling is a statistical method used for longitudinal data to describe change over time that allows for the comparison between individuals that way capturing changes in structure between individuals and similarities between groups of individuals (Frey, 2018). The growth curve model quantifies curve parameters to describe characteristics of change. This method has a wide range of applications across various industries; from population size and biomass for public health to fungal growth for biological processes (Dasgupta, 2017).

In this research, growth curve modelling is used to obtain potential critical process parameter from

real-time measured process parameters that can serve as input for the RTRT models to predict quality attributes, yield and purity. The method and modelling efforts towards the application of RTRT can differ between tests, for instance the type of model used or the outcome of the model. The outcome of the yield test is predicted with a prediction model where the relation between process parameters critical to yield and yield is used to estimate the yield of a batch. For purity, a model is constructed that defines the optimal control limits for the process parameters critical to purity that are used to classify the purity of a batch. The binary outcome value of the purity test depends on whether or not all critical process parameters fall within their specified control limit. In both models critical process parameters are used for the prediction of a test outcome; for yield as independent variables for the prediction model and for purity to build control limits. These critical process parameters can be parameters measured off-line at certain time points during the process or can be derived from real-time on-line measured process parameters. Real-time on-line measured parameters often provide more insight in process behaviour compared to process parameters measured off-line because the real-time measured process parameters provide insight in how the process parameter changes over time. The case study of Torres (2017) also emphasizes the use of real-time process parameter data for RTRT. An example of a process parameter that is often measured real-time is optical density which measures the biomass in a fermentation process where changes over time in parameter value provide information on growth. The longitudinal data of the real-time measured process parameters cannot be directly used as model input for yield and purity, instead, characteristics of curve have to be derived. This is where growth curve modelling comes into play; a growth curve model can be used to derive the characteristics of the curve. These curve characteristics can serve as critical process parameters for the yield and purity model (see Figure 3.1).

Growth curve modelling is a method widely used for various applications however the existing models are limited in their model flexibility. The models can only be used for parameters for which the standard curve behaviour is observed which limits the possibilities in terms of analysis and input parameters for RTRT tremendously. Especially, since in practice, parameters often have slight deviations from the standard curve for example because they are indirectly controlled. Therefore, there is a need for growth curve models that can be applied to a wider range of process parameters. This chapter will first define the shortcomings of existing models and based on these insights formulate alternative growth curve models. These growth curve models will take into account the behaviour of real-time measured process parameters often observed in industry that as for now could not yet be modelled using growth curve modelling. These alternative growth models are validated using an industry case study where the model fit is evaluated and the added value is determined by comparing with existing models. Alternative growth models allow for more process parameters to be analyzed and to serve as input for RTRT models. This can substantially increase the quality and prediction performance of RTRT models and will get us closer to actual RTRT implementation. In addition, this growth modelling can be used for various other purposes such as process optimization and process monitoring. For instance, quantifying characteristics of process parameter curve enables the comparison of batches with low yield and high yield based on data analysis instead of visual inspection.

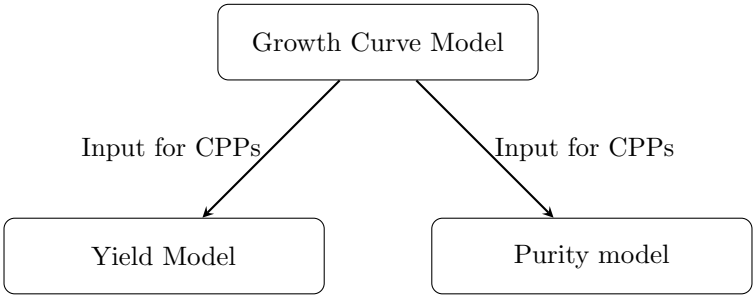


Figure 3.1: The linkage between the growth model, the yield model and the purity model

3.2 Literature Study

An essential part in the field of bacteriology when optimizing or predicting a particular quality or process outcome is the identification of the critical process parameters influencing this outcome (EMA (2012); FDA (2010); Tirumalai & Porter (2005)). Two type of critical process parameters can be distinguished: (1) critical process parameters derived from on-line real-time measured process parameters and (2) critical process parameter measured off-line at a certain time point(s) during the production process. This section will focus on the existing literature on process parameters measured real-time and the application of growth curve modelling.

According to Loutfi et al. (2018), many process parameter that are measured real-time often resemble a curve similar to the well-known bacterial growth curve. Investigators have come up with a variety of adequate mathematical models to describe this growth curve and translate the measured data of process parameters into a limited number of biologically relevant curve parameters (Zwietering, Jongenburger, Rombouts, & van 't Riet, 1990). Peleg & Corradini (2011) and Zwietering et al. (1990) have studied the modeling bacterial growth curves and have distinguished two types: (1) empirical algebraic expressions and (2) growth rate models. The empirical models are not derived from basic principles or mechanistic considerations therefore the model does not explain why a certain pattern is observed. The only purpose of the model is to describe and quantify the observed pattern with the highest mathematical convenience and best goodness of fit. For these models, the assignment of intuitive meaning to the model coefficients is seen as highly important hence the number of adjustable parameters should held minimal; this is called the parsimony principle. The growth rate model is, in contrast to the empirical model, based on an existing mathematical model namely the continuous logistic equation. Independent of type, empirical algebraic expression or growth rate model, the mathematical parameters of the growth curve model, for most models three, can be translated into biologically relevant parameters which described the characteristics of the growth curve; the maximum specific growth rate (μ_m) , the lag time (λ) and asymptotic value (A). The meaning of these three biologically relevant parameters is independent of the type of model used although the formulas translating the mathematical process parameters into the biologically relevant parameters differ. There exist some models that have four mathematical parameters instead of three, in most cases, this fourth parameter has no biological meaning.

Several studies have focused on the comparison of the different growth curve models evaluating their applicability in terms of fit and interpretation in practice. Zwietering et al. (1990) and Z. Lu et al. (2005)

regarded the Gompertz model, an empirical algebraic expression type model, as the best model in terms of fit and ease of use. However, Zwietering et al. (1990) commented that the the difference between the logistic function and Gompertz model is small. In addition, based an extensive literature study, Zwietering et al. (1990) concluded that the Gompertz model is the most widely used growth model in both industry-based and research-based studies. In contrast, Annadurai, Babu, & Srinivasamoorthy (2000) concluded by comparing the expected and measured optical density values of the Pseudomonas Putida bacteria that the logistic model was best to model the growth pattern in comparison to the Gompertz model and Richards model. Peleg & Corradini (2011) elaborate on the article of Zwietering et al. (1990) by stating that the Gompertz model is the most suitable in terms of model flexibility that way allowing for adaptions based on unique parameter characteristics. Moreover, according to Peleg & Corradini (2011), the use of a growth rate type model named the Verhulst model is limited in practice because this model often does not show a smooth transition between the bacterial growth phases.

The existing literature on growth curve modelling knows two main limitations. First of all, most of the existing growth models are limited to the modelling of the lag phase, exponential growth phase and stationary phase. Often the death phase is not considered because this phase is rarely observed in production process as it negatively impacts the process output and is therefore undesirable from a business perspective. However, when observed it can impact quality attributes that model process output hence it may be relevant for RTRT purposes to model this phase. Chatterjee, Chatterjee, Majumdar, & Chakrabarti (2015) is one of the few studies that has incorporated this last phase of the bacterial curve and did so by introducing a fourth biologically relevant parameter, the death rate (d), for the Gompertz model. However, the shortcoming of the model of Chatterjee et al. (2015) is that it does not provide a smooth transition for processes for which the stationary phase is short to none existent. Secondly, the application of growth curve models is limited to direct measures of relative population size, for example viable cell count, though many other process parameters measured in bio-pharmaceutical production processes however not directly measuring population size also resemble a bacterial growth curve (Mytilinaios, Salih, Schofield, & Lambert (2012); Dalgaard & Koutsoumanis (2001)).

Another relevant stream of literature discusses the possibilities for process improvement using the insights gained from modelling the growth curve though this field is not widely studied yet. For example, Fujikawa et al. (2004) developed a logistic model to predicted the bacterial growth curves for various temperatures histories which enables a better understanding of the impact of storage and transportation process for food products with respect to food quality assurance. In addition, Zou, Li, & Wang (2013) defined the optimal harvesting method and maximum sustained yield using the Gompertz stochastic differential equation.

In summary, this study has several contributions to the field of life sciences, in particular to the stream of bacteriological growth modelling. First of all, existing literature is limited to the application of the growth curve models for parameters directly measuring the population size. This work will adapt existing growth curve models to allow for the model to be applied to a wider range of process parameters. In addition, the model developed in this study includes the death phase by modelling this phase as a separate stochastic

process, that way extending the limited research on the inclusion of this phase in the growth curve model. By developing a growth curve model that is broader in its applicability in the bio-pharmaceutical industry, more opportunities in the area of process improvement and process optimization arise as process behaviour can be quantified and operational management methodologies can be applied.

3.3 Mathematical Model

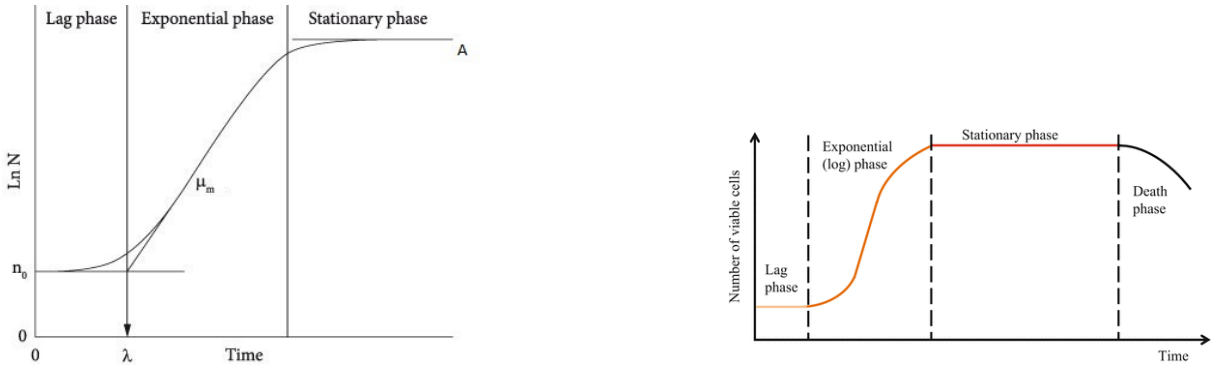
In this section, three mathematical models are presented that can be used to determine relevant curve parameters for real-time measured process parameters. These curve parameters can be used as input for models that predict quality characteristics of a product. The mathematical models are based on an existing growth curve model from literature that will be adapted and expanded in order to fit various process parameters. From the literature review in Section 3.2, it can be concluded that the Gompertz model or the logistic model would be most suitable as starting point for this mathematical model. These models appear to have the best model fit (R-square above 95%). Additionally, it is mentioned that, compared to other growth curve models, the Gompertz model is most suitable in terms of model flexibility. Model flexibility is important for this research because the basic model will be adapted so it can be applied to a wider range of process parameters types. Therefore, it is chosen to use the Gompertz model as basic model for this analysis. This is supported by data from the industry; the data set provided by MSD on real-time measured process parameters shows that in most cases the Gompertz model was superior compare to the logistic model in terms of model fit.

3.3.1 The Gompertz Model

Benjamin Gompertz developed a sigmoid function, now called the Gompertz model, which models growth being the slowest at the beginning and the end of a time period with a period of rapid increase in-between. The explanation of why this behaviour occurs is not provided by the model. When developed, the model was intended for human mortality however nowadays the model is modified for the application of biological production processes. The standard Gompertz model is presented in Equation 3.1 where y(t) represents the relative population size at time t and a, b and c are mathematical parameters.

$$y(t) = a * e^{-e^{b-c*t}} \tag{3.1}$$

The mathematical parameters from Equation 3.1 can be translated to three biologically relevant parameters; the maximum specific growth rate (μ_m) , the lag time (λ) and asymptotic value (A). Figure 3.2a illustrates the three parameters for a standard growth curve with the growth phases for the standard growth curve presented in 3.2b. The maximum specific growth rate represents the measure of growth intensity of the bacteria and is given by the slope of the line in the exponential growth phase. The lag time indicates the moment of transitioning from the lag phase into the exponential growth phase. The asymptotic value speaks for itself although this value can be tricky when during the process the stationary phase is not reached. In this case, A represents the extrapolated value which can be unrepresentative for the behaviour of the process. Equation 3.2 presents how the mathematical parameters of the Gompertz model can be translated to the three biologically relevant parameters. The derivation of these translation formulas can



(a) Standard Bacterial Growth Curve Model with Curve Parameters

(b) Schematic View of a Typical Microbial Growth

Figure 3.2: Curve Parameters of Growth Model with View of Included Growth Phases

be found in Appendix 8.

$$\mu_m = \frac{a * c}{e} \qquad \lambda = \frac{b-1}{c} \qquad A = a \tag{3.2}$$

3.3.2 Gompertz Model with Death Phase

The standard Gompertz model can be applied to situations where the bacterial growth curve includes the lag phase, the exponential phase and the stationary phase. The death phase is not included in the standard Gompertz model because in production processes this phase is avoided by stopping the process in the exponential growth phase or stationary phase since it negatively affects the process output. However, for RTRT modelling this phase can be interesting for predicting the outcome of quality attributes measuring process output or biomass. Modelling process parameters for which in some cases a death phase is observed requires expanding the existing literature on growth curve modelling.

Limited literature is available on incorporating the death phase in growth models and, if available, death phase is incorporated by adding a death rate to the existing Gompertz model (Chatterjee et al., 2015). However, when the transition from the exponential phase into the death phase is only intermittent by a very short stationary phase, adding a death rate creates problems with model fit. Therefore, the existing Gompertz model is extended by adding a second stochastic process that models the death phase behaviour. The resulting model is presented in Equation 3.3. The first stochastic process modelling the lag phase, exponential phase and stationary phase using the Gompertz model and the second stochastic process modelling the death phase. This makes the model more generally applicable in a real-life production environment; it can be applied to process with a short or none existent stationary phase or for processes that only for some batches observe a death phase. The death phase will be modelled with an decreasing exponential function; this is supported by the articles of Bailey (2018) and Todar (n.d.). The parameters d_1 and d_2 are used to model the exponential function for the death phase (see Equation 3.3). The mathematical parameters d_1 and d_2 can be explained biologically; d_1 represents the value of y at which the death phase starts and d_2 illustrates the rate of death. The correction of t with t_{death}, the time at which the stationary phase transitions into the death phase, ensures that d_1 models the y-intercept at the start of the death phase.

Differences in curves are observed between batches for a particular fermentation process produced with the same production settings. These differences can be explained by the biological nature of the process, for example, differences in the quality of the components of the medium can result in a higher or lower growth rate. As a result, some fermentation batches only show a lag phase, exponential phase and stationary phase where other batches also present a death phase. Using t_{death} allows for the model to be used also when no death phase is observed. Mathematically, if a death phase is observed, the derivative should become negative at a certain moment during the fermentation process; this insight is used to determine t_{death}. If no death phase is observed, t_{death} is set to a value bigger than T where T represents the duration of the process as a result only the Gompertz model is used to determine y(t).

$$y(t) = \begin{cases} a * e^{-e^{b-c*t}}, & \text{if } t < t_{death} \\ a * e^{-e^{b-c*t}}, & \text{if } t = t_{death} \\ a * e^{-e^{b-c*t}} + d_1 * (1 - d_2^{t-t_{death}}), & \text{if } t > t_{death} \end{cases} \tag{3.3}$$

3.3.3 Gompertz Model for Indirectly Controlled Process Parameters

In a production environment, three type of process parameters can be distinguished: (1) uncontrolled process parameters, (2) directly controlled parameters and (3) indirectly controlled parameters. Directly controlled process parameters are often kept at a constant level therefore do not resemble a bacterial growth curve. Uncontrolled process parameters can most often be modelled with the standard Gompertz model or with the Gompertz model with a death phase. The curve of an indirectly controlled process parameters resembles the bacterial growth curve till a certain point in time, from that point on a stabilization at a certain level is observed. A process parameter is indirectly controlled when this particular process parameter is influenced by a directly controlled process parameter. For example, pH is controlled by adding lye and acid such that the pH of the solution is between set control limits. pH does not resemble a bacterial growth curve but the total lye usage, which is influenced by the control mechanism of pH, does. The total lye usage will show typical bacterial growth behavior until pH stabilizes and stays between its control limits. From this point, lye stabilizes at its maximum level and a transition between the standard Gompertz model to a constant value is observed. These type of process parameters, starting with a typical growth curve and transitioning into a constant value, can be modelled as presented in Equation 3.4 using combination of the standard Gompertz model and a variable d representing the maximum value. The Gompertz model also has a maximum value reached in the stationary phase, however the main difference between the model presented in Equation 3.1 and Equation 3.4 is that latter allows for an abrupt transition between exponential growth and the stationary value whereas for the standard Gompertz model this transition is smooth.

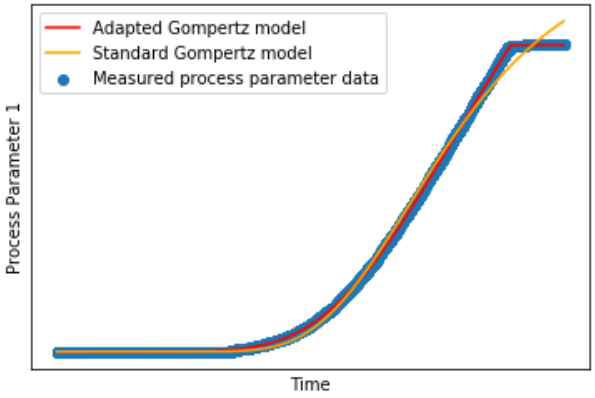
$$y(t) = min(a * e^{-e^{b-c*t}}, d) \tag{3.4}$$

3.4 Numerical Analysis

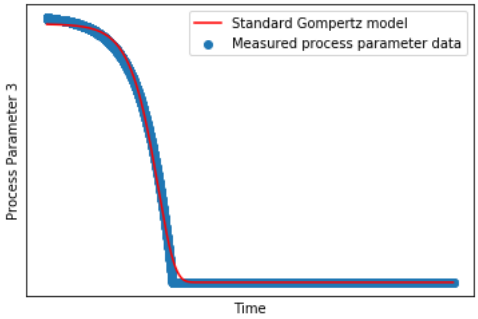
This section discusses a numerical case study at MSD Animal Health. The goal of this case study is to determine the potential applicability of the adapted Gompertz models and to get a deeper understanding on how these models can be used in practice.

MSD has provided us with a substantial data set of historical batch data of the last two years that includes all process parameters documented and measured during the production process. For this analysis, only process parameters that are measured real-time are relevant. These real-time measured process parameters can be divided in three groups: (1) uncontrolled process parameters, (2) directly controlled process parameters and (3) indirectly controlled process parameters. Data shows that directly controlled process parameters are kept at a certain constant level therefore we are only interested in the uncontrolled and indirectly controlled parameters. From the parameter overview of MSD, three process parameters have been identified as relevant; these parameters are measured real-time and resemble the behaviour of a bacterial growth curve. Process parameter 1 is an indirectly controlled parameter for which the typical two-stage curve is observed; first the typical bacterial growth curve than a strict transition to a constant value. The Gompertz model expressed in Equation 3.4 fits this behavior best because this model is build for indirectly controlled parameters and will therefore be used for the analysis of this process parameter. The second process parameter is uncontrolled and resembles the typical bacterial growth curve including the death phase. From the curve it is observed that the exponential growth phase transitions almost directly into the death phase. Because of the presence of the death phase, this process parameter can be modelled best with Equation 3.3. Process parameter 3 resembles a unique variant of a bacterial growth curve. The curve shows the decline in the process parameter value through the various phases of bacterial growth. This process parameter can be modelled with the standard Gompertz model presented in Equation 3.1 because the curve of this process parameter does not illustrate a death phase and the parameter is uncontrolled. The flexibility of the Gompertz model allows for the modelling of such curves however the application for this unique variant is hardly discussed in literature.

Figure 3.3 and Figure 3.4 show the curves for the three process parameter explained in the previous paragraph. The blue line corresponds to the measurement data for that specific process parameter. The red line illustrates the model fit for the adapted models except for process parameter 3 where this line represents the fit of the standard Gompertz model. In addition, the model fit for the standard Gompertz model is shown by the orange line for process parameter 1 and 2. For process parameter 2, an additional line is shown representing the model from Chatterjee et al. (2015) that includes the death phase by correcting the standard Gompertz model with adding a death rate. As can be observed from the graphs, the adapted models presented in this study fit the data well and are superior to existing models. For process parameter 1, the standard Gompertz model overestimates the parameter value from the moment in time that the process parameter stabilizes to a constant value. As can be seen from Figure 3.3a, the adaption to the standard Gompertz model enables a strict transition between the bacterial growth behaviour and stabilization to a constant value hereby improving the model fit. The comparative analysis of the three models for process parameter 2 illustrates that till the start of the death phase the fit of the



(a) Growth curve modelling for indirectly controlled process parameter 1



(b) Growth curve modelling on a special variant of the bacterial growth curve for process parameter 3

Figure 3.3: Growth curve modelling for process parameters observed in a production process

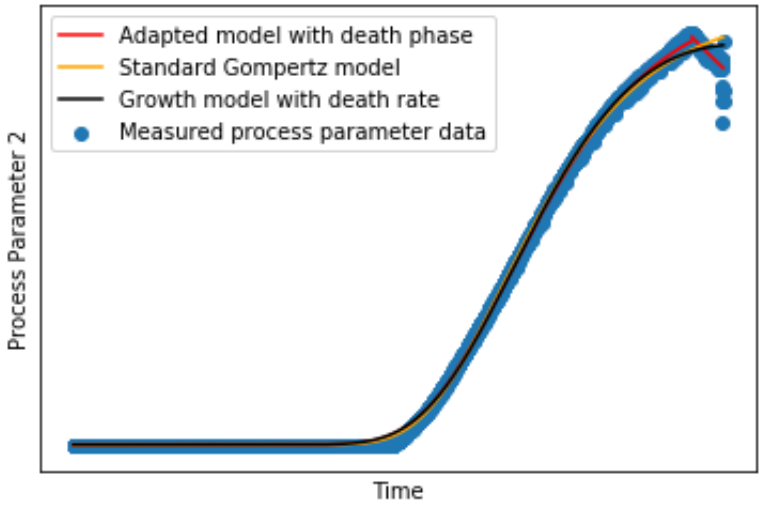


Figure 3.4: Growth curve modelling with the presence of a death phase for process parameter 2

three models is similar. However, at the transition to the death phase both the standard Gompertz model and the model from Chatterjee et al. (2015) overestimate the value of the process parameter whereas the adapted Gompertz model from Equation 3.3 follows the behaviour smoothly. This overestimation of the model from Chatterjee et al. (2015) decreases when the death phase duration becomes longer. The quantified differences in goodness of fit are estimated with R-square and can be found in Appendix 8.

Additionally, the impact of the improved fit of the growth curve models on the prediction accuracy of the yield model is analyzed. This analysis shows that the prediction error of the yield model decreases by using the adapted growth curve models instead of the standard Gompertz model. The extensive analysis of this improvement is shown in Appendix 8.

3.5 Conclusion and Discussion

In the bio-pharmaceutical industry, many characteristics of the process, referred to as process parameters, are measured during the production processes. Information on these process parameters allows for

quality guarantees, process analysis in case of events, process improvement and process optimization. The more insight gained from these process parameters the more opportunities arise for analyzing and improving the process. In particular, process parameters measured real-time, for example optical density, provide valuable insight in process behaviour as the change in process parameter value over time can be observed. Quantifying this change over time and translating this into relevant curve parameters requires curve modelling. As these parameters often resemble the typical bacterial growth curve model so-called growth curve modelling can be used. Although many different growth curve models exist, the application possibilities in terms of process parameter types is fairly limit. This study addresses this problem by adapting an existing growth curve to fit different types of real-time measured process parameters.

The real-time process parameters measured during a bio-pharmaceutical production process can be uncontrolled process parameters or indirectly controlled process parameters, including all bacterial growth phases or just some of the phases. The existing growth curve models are limited to the modelling of uncontrolled parameters that do not include the death phase. By making multiple adaptions to the Gompertz model, a well-known growth curve model, the applicability of the modelling with growth curve models broadens to more different type of process parameters observed in the production environment. First, the adaption for including the death phase is introduced. The adapted model includes two stochastic processes: the Gompertz model for modelling the lag phase, the exponential growth phase and the stationary phase and a decreasing exponential model for modelling the death phase. Separating the stages of growth from the stage of death allows for a wider range in applications and more flexible usage. The model can be used for processes that have a short to none existent stationary phase and for products that, due to biologically inherent randomness, only observe the death phase for some batches. Secondly, the standard Gompertz model is adapted for process parameters that are indirectly controlled, meaning the process parameter is influenced by another process parameter that is controlled by, for example, keeping the value between set control limits. This is often observed in the bio-pharmaceutical industry as the processes are highly controlled to provide quality assurance. For the modelling of these process parameters a combination of the standard Gompertz model and a constant representing the maximum value is used. This allows for an instant transition between a bacterial growth curve type of behaviour and a fixed stabilization level. The models were validated by conducting an industry case study at MSD. This analysis shows that the fit of the adapted models is superior in comparison to the existing growth curve models. For indirectly controlled process parameters, the transition from bacterial growth to a constant value can be modelled more accurately with the adapted model. As a result, quantitative information on the moment of transitioning and the value of the constant can be obtained. The adapted model for process parameters presenting a death phase was found to be more accurate in defining the start of the death phase and modelling the decrease in the process parameter value than existing models. Using these adapted models allows for more process parameters to be analyzed with growth curve modelling. As a result, more process parameters can be analyzed in a quantitative matter providing the required input in the form of critical process parameters for RTRT models. This both improves the accuracy of these models and allows for more quality attributes to be considered for RTRT application. Additionally, expanding the portfolio of process parameter that can be modelled with a growth curve model creates opportunities

for process analysis and process improvements. Moreover, it contributes to existing literature on growth curve model where not only the bio-pharmaceutical industry but also other industries such as the food industry can benefit from.

Future work could test these models and their applications in other industries, for example the food industry. In the food industry, growth curve models are used the monitoring of beer production but also for quality assurance during food transportation. It would be interesting to see what benefits can be gained in this industry by being able to monitor more different types of process parameters. Additionally, it would be interested to explore how these insights gained for the growth curve models can be used for process improvement projects.

Chapter 4

Real Time Release Approach for Purity

This chapter presents how RTRT can be applied to the purity test which is a dichotomous quality attribute that determines whether or not bacteria other than the bacteria cultivated are present in the batch. The goal is to identify acceptance and rejection criteria in the form of control limits for the process parameters critical to purity that can be used to classify a batch either as pure or contaminated. The growth curve model is used to find potential process parameters critical to purity. The control limits are defined by finding the optimal trade-off between quality risk and waste risk, measured using false negative and false positive misclassifications. First, Section 4.1 explains the quality attribute purity and discusses the current testing method and the RTRT strategy aimed for. Than, in Section 4.2 an overview on the relevant stream of literature is provided and the contributions of this research are defined. Subsequently, Section 4.3 present the mathematical model used for defining the optimal control limits for the process parameters critical to purity. Section 4.4 focuses on the validation of the mathematical model using an industry case study at MSD. At last, the conclusion and managerial insights are discussed.

4.1 Introduction

The criteria for a batch to be classified as pure is two-fold, only one type of bacterial colonies should be present and this type should resemble the colony of interest. If a batch contains bacterial colonies of a type other than the type that is cultivated, a batch is classified as contaminated. The presence of a contaminating bacteria can cause undesired reactions when injecting the vaccine in an animal causing a health risk to the animal. A contamination can be caused by various factors for example improper cleaning of the fermentors or insufficient tubing connections. Currently, purity is tested by conducting a traditional purity test with an outcome on nominal scale; zero when a batch is impure and one when a batch is pure. With this testing method, a sample is taken from the fermentor after the main fermentation process is completed and this sample is incubated on an agar plate for a prescribed period of time. After this time, the agar plate is inspected on the presence of bacteria growth. Additionally, a small part of the sample is

taken before incubation and analyzed microscopically to check of the presence of the bacteria cultivated. If no growth of contaminating bacteria is observed and the microscopic inspection is acceptable, the batch is classified as pure.

The traditional purity test had disadvantages both in terms of cost and quality. First of all, the purity test has to be conducted in specialized laboratories that require cleaning protocols and occupy valuable production space. Secondly, the test has to be executed by employees who have had the proper training required for conducting the test. Thirdly, the results of the purity test are only known after the incubation period has passed. The production process is continued during that time meaning the batch is further processed to downstream eventhough the results of the purity test are not yet known. This is a business risk; if the batch appears to be contaminated, the value added in downstream is lost and the batch is disposed. The cost allocated to the specialized laboratories and the personnel costs increase the manufacturing cost of the product. Also, the continuation of the process eventhough purity results are not known can lead to unnecessary cost resulting from further processing a contaminated batch and can cause quality risks as contaminating bacteria are introduced to downstream processing.

With the introduction of RTRT, there is no need for purity testing as a result the cost for a specialized lab environment and personnel are eliminated. Also, the classification of the batch, pure or contaminated, is known directly after the main fermentation process is completed or in some cases even during the main fermentation. The latter is the case when during the process it is observed that a certain critical process parameter is outside its defined control limit meaning the batch is contaminated. To prevent further cost, the fermentation process can be aborted.

The RTRT strategy uses acceptance and rejection criteria, better known as control limits, for process parameters critical to purity to classify the purity of a batch. A batch is classified as pure if all critical process parameters are within their specified control limit. The formulation of these control limits is a trade-off between waste and quality risk; when the control limits are too wide the probability of having a quality issue increases whereas when the control limits are too narrow the probability of having unnecessary waste increases. Chapter 2 explains that for RTRT understanding of the process and of the relations between process parameters critical to purity and the quality attribute purity should be demonstrated and verification of these relations using data should be provided. Moreover, the acceptance criteria for purity classification should be justified and the relation between RTRT approach and end testing method should be established. In this chapter, a model is constructed that evaluates the trade-off between the two types of misclassifications for different combinations of control limits to find the optimal combination in terms of waste and quality for the process parameters critical to purity. This provides the required justification for the acceptance criteria for purity classification. An industry case study at MSD is conducted to validate the model and compare the model approach with the traditional purity testing method. In addition, the process of identifying the critical process parameters and formulating the relations between these critical process parameters and purity is demonstrated.

4.2 Literature Review

This research is connected with two streams in literature: life sciences and operations management. Section 4.2.1 discusses the relevant studies related to the field of life sciences. Section 4.2.2 elaborates on the relevant work in the field of operations management.

4.2.1 Relevant Literature in Life Sciences

Numerous articles in the field of life sciences discuss the causes, consequences, detection and prediction of contamination during the fermentation processes. Contamination can have significant impact on the outcome of a process, for example the entire batch is unusable.

A considerable amount of literature on contamination addresses the identification and elimination of the causes of contamination. Blackwell (2017) explains how contaminations originate and provides procedures that can be used for limiting and eliminating the sources of contamination. Allman (2020) further elaborates on the sources of contamination considering the process steps and materials relevant in a fermentation process and suggests troubleshooting methods for each of these sources. Also, the linkage between preventive maintenance for fermentors and associated materials and valves, and microbial contamination is discussed as potentially beneficial for reducing contamination issues (Suvarna, Lolas, Hughes, & Friedman, 2011).

In addition, research is conducted on the quality and the improvement potential of the traditional purity analysis. The study of Remund, Dixon, Wright, & Holden (2001) quantifies the percentage of misclassifications (incorrectly rejecting or accepting a seed based on purity) resulting from numerous assessing methods for purity, that way defining confidence intervals on purity assurance per method and identifying the optimal method for testing purity. Ramirez-Arcos, DiFranco, McIntyre, & Goldman (2017) conducted a validation study for a bacterial screening protocol to determine whether or not this method is efficacious for identifying Gram-negative bacteria using the percentage of false positives and false negatives for different Gram-negative bacteria with varying contamination levels. de Souza Liberal , da Silva Filho, de Morais, Simoes, & de Morais (2005) even succeeded to identify a detection limit that tells us the minimal concentration of contaminated cells in the fermentation that should be presented for a contaminated to be detected.

Eventhough most industries still use the conventional purity testing methods, increasing interest is shown in predicting purity based on process parameters measured during the fermentation process or specific process analytical tools. Sun et al. (2019) has formulated a soft-sensing model using Gaussian process regression that predicts the seed purity based on the viscosity of the fermentation broth. Similarly, M. Wang, Han, Sun, & Chen (2021) used the soft-sensing model method to build a prediction model for purity using the correlation between the online measured sugar content of the broth and the purity of a batch. J. Yang, Chen, & Jin (2015) explains a method for in-time contamination detection where contamination is predicted by fusing and extracting latent information from changes in the trend of real-time measured process parameters. A more general fault detection method is presented by Mid & Dua (2018); faults are detected and diagnosed by modelling a certain type of fault as a state of the art

variable that can be estimated by an explicit function of kinetic model parameters, such as the growth rate coefficient of yield obtained from the dissolved oxygen or temperature curve. Basílio et al. (2008) and Sue, Obolonkin, Griffiths, & Villas-Bôas (2011) describe monitoring the presence of particular yeast specifies based on their unique polymerase chain reaction (PCR) or gas chromatography-mass spectrometry (GC-MS)-based metabolic fingerprint, respectively. Subsequently, studies have identified new measurement tools for detecting purity; Sanaeifar, ZakiDizaji, Jafari, & Guardia (2017) explain the e-nose as tool for early detection of contaminations in food production and Elmroth, Valeur, Odham, & Larsson (1990) describe the use of gas chromatography to diagnose the presence of contaminating organisms. The model presented in this studies differs from the studies in this stream as these studies model purity as a numeric variable instead of a dichotomous variable.

Another relevant, although small, stream of research focuses on real time release testing for purity attributes. For instance, the study of Laursen, Frederiksen, Leuenhagen, & Bro (2010) demonstrates purity analysis using a multivariate statistical process control tool that can, based on principal component analysis, monitor subtle changes in a chromatography. The article discusses how this method can, in the future, replace the visual investigation of the chromatography for a real-time release strategy. In a similar context, Walch et al. (2019) have formulated a model based on an array of online detector measurements providing real-time predictions for purity replacing laboratory measurement with a real-time release method. Nevertheless, the model proposed in this research differs from the aforementioned studies for two reasons: (1) the RTRT approach is not based on control limits and (2) the purity measure in this research is dichotomous in contrast to these other studies where purity is numeric and represents the amount of contaminating bacteria present in the batch.

In summary, the prediction and detection of contaminations and seed purity is extensively discussed in literature and studies have presented prediction models for contamination using real-time measured or kinetic parameters or in some case even fingerprinting. However, identifying control limits for process parameters to identify the purity of a seed or batch has not yet been discussed. In addition, none of these existing studies model or predict purity as a dichotomous quality attribute. To the best of our knowledge, this study is the first work that formalizes control limits for process parameters to predicted the purity of a batch. Moreover, this work will provide a deeper understanding in optimizing control limits for purity detecting by quantifying the trade-off between the two types of misclassifications, false positives and false negatives.

4.2.2 Relevant Literature in Operations Management

In the era of data analysis, more industries started using Operations Management (OM) approaches to improve processes and increase their efficiency. Although results have proven the benefits of OM methodologies, the bio-pharmaceutical industry still lacks to seize the opportunities these sophisticated analytical tools and models can bring (Kaminsky & Wang, 2015).

Most studies in the field of bio-pharmaceuticals using OM methodologies focus on decision making in the supply chain, such as, planning and scheduling (Limon & Krishnamurthy (2020);Martagan et al. (2020); Leachman, Johnston, Li, & Shen (2014);Vieira, Pinto-Varela, Moniz, Barbosa-Póvoa, & Papageorgiou

(2016); Vieira, Pinto-Varela, & Barbosa-Póvoa (2019)), investments strategies (Farid, Washbrook, & Titchener-Hooker (2008)), strategy alignment (Holder, Devpura, Lee, & Chandran (2018);Gurău (2004)) and, risk management and operational decision making (Wang, Xie, Martagan, Akcay, & Corlu (2019); Ma (2011); Abbasian et al. (2021)). For example, Limon & Krishnamurthy (2020) developed a tool that proposes an effective schedule for biomanufacturing projects by solving a series of mixed-integer linear programming models while incorporating user feedback. Another stream in OM studies investigates the impact of regulations (Lim, Zhou, Washbrook, Titchener-Hooker, & Farid (2004); La Rosa & Liberatore (2014)) and, strategic and RD alliances (Shakeri & R (2017); Shin, Kim, & Park (2015); Mazzola, Bruccoleri, & Perrone (2015); M. Chang et al. (2019)). Moreover, limited OM research is available on the improvement and optimization of manufacturing practices and operational decisions (Martagan, Krishnamurthy, & Maravelias (2016);Koca, Martagan, Adan, Maillart, & van Ravenstein (2020); Martagan (2015); Martagan et al. (2020); McGillicuddy, Floris, Albrecht, & Bones (2018)). In the context of improving manufacturing operations, Martagan et al. (2016) proposed a stochastic model that integrates both operational and cell-level dynamics to identify the optimal harvesting policy in fermentation operations, balancing the risk of failure and output quantity. However, none of these studies discuss the concept of real time release testing or predicting the product quality using control limits for critical process parameters.

In addition, the OM literature stream of machine maintenance shows relevance for this research where the field of time-based and condition-based preventive maintenance are most relevant to this study (Scarf (1997); Golmakani & Fattahipour (2011);Banjevic, Jardine, Makis, & Ennis (2001);Heng et al. (2009)). For example, Chen & Trivedi (2005) presents a semi-Markov decision making tool that determines the optimal maintenance policy combing time-based and condition-based maintenance. The authors used this tool to jointly optimize the inspection rate, covering the time-based aspect, and the maintenance policy in regards to the failure limit, covering the condition-based nature. In a more recent study, Shafiee & Finkelstein (2015) proposed an optimization model that determines the optimal group maintenance time T* at which a planned group preventive maintenance action is conducted for a multi-unit degrading system. The objective of this model is to minimize the long-run maintenance cost while considering additional cost resulting from maintenance actions, such as set-up costs. More general, Ahmad & Kamaruddin (2012) provide a general overview of both time-based and condition-based maintenance discussing the application, implementation, challenges, and modelling of each maintenance technique. Nevertheless, this study differs from the existing research in this field. First of all, the control limits in this research are optimized by considering the trade-off between waste and quality whereas in condition-based maintenance these are directly obtained from production data. Secondly, this study considers the thresholds for multiple parameters in contrast to the time-based strategy that only determines the optimal T or the condition-based strategy that optimizes the thresholds separably when not obtained from data.

Another stream in OM research relevant to this study focuses on quality and reliability in processes by monitoring trends with control charts (Aebtarm & Bouguila (2011a); Glushkovsky (1994); C. Zhang, Xie, & Goh (2006); M. Zhang, Peng, Schuh-Renner, Megahed, & Woodall (2013); Xie, Goh, & Ranjan (2002); Topalidou & Psarakis (2009); H. Wang (2009); Aebtarm & Bouguila (2011b); Ahangar & Chimka

(2015); X. Lu, Xie, Goh, & Lai (1998a)). For example, X. Lu, Xie, Goh, & Lai (1998b) developed a Shewhart-type control chart for multi-attribute products that uses the weighted sum of the counts of non conforming units regarding all attributes relevant for quality compliance. Additionally, an example is presented by Xie, Goh, & Kuralmani (2000) where the traditional 3-sigma rule is challenged by developing an optimization method for the geometric control chart that maximizes the average run time at normal process level. Notwithstanding, this study defines the optimal control limits for process parameters with regards to consequence on quality in terms of waste and quality non-compliance while the aforementioned studies use standardized statistical methods for determining the control limits or neglected the relation to quality compliance.

Moreover, the body of research where the stream of preventive maintenance and process control interact has relevance to this study. The interaction between process control and preventive maintenance strategies is discussed in the work of Tagaras (1988). In this article, a joint optimization is proposed to find the optimal combination of the sample size, width of the control limits and the frequency of preventive maintenance. Another example is the study of Rivera-Gómez, Gharbi, Kenné, Montaño-Arango, & Corona-Armenta (2020) where an integrated model is constructed that optimizes production, preventive maintenance and quality control. Nevertheless, the model discussed in this research differs from the existing models as, instead of combining policies, it jointly optimizes the control limits of multiple process parameters.

To summarize, this study has multiple contributions to the field of operations management. In this research, a optimization model is developed that defines the optimal control limits considering the consequences of misclassifications using the probabilities on having a false negative or false positive misclassification. To the best of our knowledge, this is the first model that jointly optimizes the control limits of multiple parameters with respect to type 1 and type 2 error probabilities. Another contribution to the stream of RTRT is that this research expands the literature on dichotomous variables. The existing literature on RTRT is mostly focuses on numeric quality attributes and is very limited with regards to dichotomous quality attributes. In addition, the studies on dichotomous attributes lack to present policies that use control limits for classification. In this work, a model is developed can be used to find the optimal control limits for classifying dichotomous quality attributes and is validated with an industry case study.

In conclusion, this work has contributions to the field of life sciences and operations management. In this research a joint optimization model is developed that can propose the optimal RTRT strategy for dichotomous quality attributes such as purity. Additionally, this research addresses a gap in research with regards to RTRT application to dichotomous quality attributes and presents a case study from industry that present the use and potential of this model. Using RTRT has high potential for reducing cost and increasing output which makes this study highly relevant for the bio-pharmaceutical industry.

4.3 Mathematical Model

The goal of this research is to replace the traditional purity test with a RTRT strategy. The RTRT strategy evaluates the values of certain process parameters with the defined control limits to determine whether or

not that batch is pure. This section will discuss the mathematical models that can be used to determine the optimal control limits for parameters that are critical to purity. First, it is important to know which process parameters are critical to purity. Subsequently, mathematical models can be formulated that can determine the optimal control limits for the process parameter identified as critical to purity. In Section 4.3.2, a preliminary model is formulated where the objective, decision variables, input variables and constraints for this model are extensively discussed. This model uses just one process parameter to determine the purity of a batch. In addition, further analysis is conducted on the mathematical properties of the model. The preliminary model is expanded to a model that classifies the purity of a batch based on two process parameters. This model is presented in Section 4.3.4.

4.3.1 Process Parameters Critical To Purity

The identification of the process parameters critical to purity is crucial for collecting the required input for the mathematical models. Starting with the question what does "critical to purity" mean, parameters that are critical to purity have a significantly different value for contaminated batches compared to pure batches. These can be process parameters that are directly measured in the production process or parameters that are derived from process parameters measured real-time, for example curve parameters. The analysis for identifying these critical parameters starts with gaining insight in what process parameters are measured and how these parameters are related to purity. Brainstorming with process experts, studying literature on contamination behaviour and/or collecting an overview of the process parameters measured during the process and their measurement procedure are methods to gain insight on which process parameters are potentially critical. Using this acquired knowledge, process parameters that are potentially critical can be listed and further analyzed. Data analysis can be used to supported or rejected the hypothesis of a process parameter being critical to purity. This analysis can include both a graphical comparison of the distribution and value of a potentially critical process parameter for contaminated batches and pure batches, and more in-depth statistical analyses such as a two-sample t-test.

The bacterial growth curve modelled explained in Chapter 3 plays an important role in the identification of the process parameters critical to purity. The growth curve model quantifies curve parameters for real-time measured process parameters for which the curve resembles that of a bacterial growth curve. Real-time process parameters are measured over the whole duration of the process and can therefore provide valuable insight in the behaviour of a process and with this in the purity of a batch. A contamination often causes the behaviour of the process to differ from the normal process behaviour. Normal process behaviour is the behaviour of a pure batch that has the desired final product quality. In the bio-pharmaceutical industry, this is called a golden batch which refers to the golden batch principle. The golden batch principle is the analysis of batches and comparing the behaviour of these batches to that of the 'golden batch' to estimate the quality of that batch or detect abnormalities.

Real-time process parameters can facilitate the analysis of differences from golden batch (normal process) behaviour by comparing the curves. The analysis can be based on visual inspection, but even stronger is a quantitative analysis. This is where growth curve modelling as explained in Chapter 3 comes in play; this model can quantify the characteristics of a curve of a real-time measured process parameter allowing

for quantitative comparison of the curves of contaminated batches and pure batches. This comparison can determine whether or not a real-time measured process parameter is critical to purity but can also tell which specific curve parameter of this real-time measured process parameter is critical. This latter is input for the mathematical models that determine the optimal control limits for these curve parameters when found critical. To determine the purity of batch, the value of the critical process parameter is compared to its defined control limits. Within these limits the process behaviour is seen normal and in line with the behaviour of a golden batch. For example, process parameter 1 has a value of 8 and its defined control limits are 7.8-8.2 hence we can conclude the batch is pure based on this parameter. However, a real-time process parameter does not have one specific value but thousands or maybe even millions data points; a data point for time step over the duration of the process. Therefore, it is needed to quantify curve parameters and use these as input for the models instead of the real-time process parameter itself.

4.3.2 Preliminary Model

The preliminary models classifies the purity of a batch based on the value of one process parameter. This section will present the objective function, decision variables, input variables and constraints of the mathematical model used to determine the optimal control limit for this process parameter.

Objective: In the field of machine maintenance defining the threshold for a parameter used for condition based maintenance is a trade-off between conducting preventive maintenance too early resulting in higher maintenance cost or conducting corrective maintenance as a result of machine break-down. This insight is used to define the control limits for the RTRT approach for purity. Formalizing the control limits is a trade-off between quality risk, classifying an impure batch as pure, and waste risk, classifying a pure batch as impure. For the RTRT model, the objective is to identify the control limit for the critical process parameter that minimizes the total cost resulting from quality risk and waste risk; in other words the cost resulting from the misclassification of batches. The total cost for misclassification can be subdivided in (1) cost (or cost ratio) resulting from false positives modelled as the cost C_{FP} times the probability of having a false positive ($p_{FP}(k)$) and (2) cost (or cost ratio) resulting from false negatives modelled as the cost C_{FN} times the probability of having a false negative ($p_{FN}(k)$), this way quantifying waste risk and quality risk, respectively. Both probabilities, $p_{FP}(k)$ and $p_{FN}(k)$ are dependent on k where k represents the control limit coefficient (see Figure 4.1), in other words the width of the control limit.

Inspired by modelling condition based maintenance optimization problems, the probability distributions for having a false negative or having a false positive are constructed using the probability of a batch being pure ($p(v)$), the probability of occurrence for a certain parameter value ($f(h)$) and the boundary conditions for classifying a batch as pure or impure which are dependent on k, x and the measurement domain. The section on variables will further elaborate on the aforementioned variables.

The probability on having a false positive, a pure batch classified as impure, is the sum of the probability of having a pure batch for all process parameter values outside the control limits and can be calculated with the formula presented in Equation 4.1. Batches are classified as impure when the process parameter has a value that falls outside the control limits. Summing all $p(v)$'s for the parameter values within the measurement domain but outside the control limit and multiplying this probability with the probability

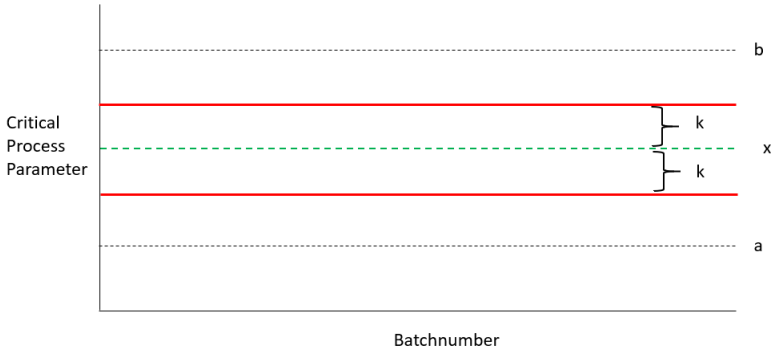


Figure 4.1: Example of Control Chart Parameters

of this parameter value occurring will result in the probability of having a false positive.

$$p_{FP}(k) = \int_{h=a}^{x-k} p(h) * f(h)dh + \int_{h=x+k}^{b} p(h) * f(h)dh \tag{4.1}$$

The probability of having a false negative, impure batch classified as pure, can be calculated by summing the probability of having an impure batch (1- $p(v)$) for all process parameter values within the control limits and multiplying the probability with the probability of occurrence for this parameter value. This calculation is presented mathematically in Equation 4.2.

$$p_{FN}(k) = \int_{h=x-k}^{x+k} (1 - p(h)) * f(h)dh \tag{4.2}$$

$$\underset{k\geq 0}{\text{Minimize}}\ p_{FN}(k) * C_{FN} + p_{FP}(k) * C_{FP} \tag{4.3}$$

Variables

Decision Variables:

The control limit is modelled with two parameters, x and k, with k being the decision variable in this model (see Figure 4.1). x and k represent the control line and the control limit coefficient for critical process parameter, respectively. The assumption has been made that x is equal to the mean of the critical process parameter. This assumption could be relaxed in a later stage of the modelling process by introducing the optimization of x. The decision variable k determines the width of the control limit and should therefore be strictly positive ($k \geq 0$). The width of the control limit influences probability of having a certain misclassification (false positive or false negative). The wider the control limit, the higher the probability on having a false negative increasing the quality risk but the lower the probability on having a false positive decreasing the waste risk. A smaller control limit will cause an increase in the probability of having a false positive but will cause the probability of having a false negative to decrease.

Input Variables:

The input variables are variables from which the function or value is retrieved from an external source or determined from an external analysis. This model has four input variables: (1) the probability of a batch being pure, (2) the measurement domains of the parameter (3) the average process parameter value modelled as x and (4) the probability that the parameter has a certain value.

The probability distribution for a batch being pure when the critical process parameter has value v is modelled with $p(v)$. This probability distribution is determined using a binary logistic regression and is dependent on the relation between the critical process parameters and the probability of a batch being pure. Equation 4.4 shows the distribution function for a critical process parameter with a negative relation to the probability of a batch being pure. The function for a critical process parameter with a positive relation is shown in Appendix C.

$$p(v) = \frac{e^{\alpha-\beta*v}}{1+e^{\alpha-\beta*v}} \qquad \forall\, v \in a, ..., b \tag{4.4}$$

where α and β are constants that are determined by maximum likelihood estimation and a and b are the boundaries of the measurement domain.

The boundaries of the measurement domain determine the boundary values for the false positive and false negative probability function and are used for modelling $p(v)$. The minimum and maximum parameter value (v) over all data points determine the lower limit, a, and upper limit, b, of the measurement domain for the process parameter in question. These boundaries are respected to prevent extrapolation; parameter behaviour outside these boundaries is unknown and could potentially differ from the behaviour within these boundaries. x, representing the control line, is set to the average value for the process parameter.

The probability that the critical process parameter has value h, $f(h)$, shows the likeliness of a certain parameter value occurring. Both the uniform and the normal distribution can be considered to model this probability. The uniform distribution assumes that each parameter value has an equal probability of occurring; this is often not the case in a manufacturing environment. The normal distribution would be a better resemblance of the actual parameter behavior in a manufacturing environment. Most parameter values are close to the mean hence the probability of occurrence of these values is high, the further from the mean the less likely that value will occur resulting in a lower probability of occurrence. However, important to consider is that a contaminated batch probably has parameter value that is not close to the normal mean but on the outsides of the measurement domain. Resulting from the definition for a parameter critical to purity meaning the value for contaminated is significantly different from that of a pure batch. Using the normal distribution could cause less weight to be given to a parameter value from a contaminated batch as a result the model accuracy would decrease. Therefore, it is assumed that $f(h)$ follows an uniform distribution. An additional benefit of using the uniform distribution is that it simplifies the modelling. The assumption of the uniform distribution is later relaxed to evaluate the impact of the distribution of $f(h)$ on the optimal k.

Constraints

Within the measurement domain, the behavior of the critical process parameter is known whereas outside this domain the behavior can differ. A k causes the control limits to exceed the measurement domain could result in classification inaccuracy. For this reason, control limits ($x - k$; $x + k$) should not fall outside the measurement domain of the data set. This is ensured by formulating the constraints presented

in Equations 4.5 and 4.6.

$$x - k \geq a \tag{4.5}$$

$$x + k \leq b \tag{4.6}$$

Complete Model

For the complete model, it is assumed that $f(h)$ follows a uniform distribution and that the critical process parameter has a negative relation with the probability on a batch being pure and x is the middle of the measurement domain ($\frac{a+b}{2}$).

$$\underset{k\geq 0}{\text{Minimize}} \left(\int_{h=a}^{x-k} \frac{e^{\alpha-\beta*h}}{1+e^{\alpha-\beta*h}} * \frac{1}{b-a} * dh + \int_{h=x+k}^{b} \frac{e^{\alpha-\beta*h}}{1+e^{\alpha-\beta*h}} * \frac{1}{b-a} * dh \right) * C_{FP} + \tag{4.7}$$

$$\int_{h=x-k}^{x+k} \left(1 - \frac{e^{\alpha-\beta*h}}{1+e^{\alpha-\beta*h}}\right) * \frac{1}{b-a} * dh * C_{FN}$$

Subject to:

$$x - k \geq a \tag{4.8}$$

$$x + k \leq b \tag{4.9}$$

4.3.3 Mathematical Properties Preliminary Model

The mathematical properties are evaluated for the preliminary model. Especially convexity is extremely important as the absence of convexity can result in a local optimal for k resulting in a sub-optimal RTRT approach. In some cases, boundary conditions have to be formulated for the included parameters in order to guarantee convexity. Furthermore, the interpretation of the probabilities and the relations as described theory have to be ensured. For example, probabilities should always be between zero and one and the probability in a false positive should decrease in k. If this is not the case, modelling issues can arise and boundary conditions have to be formulated to ensure that these properties hold. Boundaries, when needed, will only be formulated for parameters x and k. The α and β of the $p(v)$ are determined using binary logistic regression and are therefore assumed to be given. The same applies for parameters a and b modelling the measurement domain as these are dependent on the available data set.

Firstly, the mathematical properties of the probability distribution of a batch being pure, $p(v)$, are analyzed. The probability distribution can differ between critical process parameters and is dependent on the direction (positive or negative) between the critical process parameter and $p(v)$. The analysis provided in this section is based on a critical process parameter having a negative relation with the $p(v)$. The analysis of the mathematical properties for a critical process parameter with a positive relation with $p(v)$ can be done in a similar way.

Starting with the property that a probability should always be a value between zero and one. The formula for $p(v)$ ensures that this property holds therefore no boundary conditions have to be formulated. Another property evaluates the relation between the critical process parameter and the function $p(v)$. In case of a negative relation between a process parameter and the probability of being pure, the first derivative should be smaller or equal to zero. As shown in Equation 4.10, this property holds for all values of v. α and β are positive numbers hence the first derivative is always smaller or equal to zero.

$$\frac{dp(v)}{dv} = \frac{-\beta * e^{\alpha+\beta*v}}{(e^{\alpha} + e^{\beta*v})^2} \leq 0 \tag{4.10}$$

Secondly, the mathematical properties of the probability distribution for having a false positive and having a false negative are examined. Starting with false positive, this function consists of two parts, one part modelling the situation where the parameter value is above the upper control limit and the other modelling the situation where the parameter value is below the lower control limit. These parts are analyzed separately whereafter the function is evaluated as a whole. Equation 4.15 represents the part of the probability of having a false positive where the parameter value is below the lower control limit. To ensure that this probability is between zero and one, boundaries equations 4.11 and 4.12 have been formulated. If x and k are respecting these boundaries it can be assumed that the probability property (probability between zero and one) holds.

$$k < \frac{\alpha - \beta * a + \beta * x - Log[e^{\alpha}]}{\beta} \tag{4.11}$$

$$k > \frac{-\alpha + Log[-e^{\beta*x} + e^{\alpha-\beta*b+\beta*x} + e^{\beta*a-\beta*b+\beta*x}]}{\beta}$$

$$x > \frac{-\alpha + \beta * k + \beta * a + Log[e^{\alpha}]}{\beta} \tag{4.12}$$

$$x < \frac{\alpha - Log[-e^{-\beta*x} + e^{\alpha-\beta*k-\beta*b} + e^{-\beta*k+\beta*a-\beta*b}]}{\beta}$$

The property also has to hold for Equation 4.16 modelling the probability of having a false positive where the parameter value is above the upper control limit. The resulting boundary conditions for x and k for which the property is holds are presented in Equation 4.13 and 4.14.

$$k < \frac{-\alpha + \beta * b - \beta * x + Log[e^{\alpha}]}{\beta} \tag{4.13}$$

$$k > \frac{\alpha - Log[-e^{\beta*x} + e^{\alpha-\beta*a+\beta*x} + e^{\beta*x-\beta*a+\beta*b}]}{\beta}$$

$$x < \frac{-\alpha - \beta * k + \beta * b + Log[e^{\alpha}]}{\beta} \tag{4.14}$$

$$x > \frac{\alpha - Log[-e^{\beta*k} + e^{\alpha+\beta*k-\beta*a} + e^{\beta*k-\beta*a+\beta*b}]}{\beta}$$

$$\int_{h=a}^{x-k} p(h) * f(h)dh \tag{4.15}$$

$$\int_{h=x+k}^{b} p(h) * f(h)dh \tag{4.16}$$

Than, the formula for false positive as a whole can be analyzed. The boundaries on x and k for which the probability property holds are given in Equation 4.18 and 4.17. The probability on having a false positive is between zero and one for all x and k values respecting these boundaries.

$$x > \frac{Log[-\frac{e^{-\beta*k}(e^{\alpha+2\beta*k+\beta*a}-e^{\alpha+\beta*b}-e^{\beta*a+\beta*b}+e^{2\beta*k+\beta*a+\beta*b})}{e^{\beta*a}-e^{\beta*b}}]}{\beta} \tag{4.17}$$

$$x < \frac{Log[\frac{e^{\alpha-\beta*k}(-e^{\alpha}+e^{\alpha+2\beta*k}-e^{\beta*a}+e^{2\beta*k+\beta*b})}{e^{\beta*a}-e^{\beta*b}}]}{\beta}$$

$$k < \frac{1}{\beta}Log[\frac{1}{2(e^{\alpha}+e^{\beta b})}e^{-\beta a}(-e^{\beta(a+x)}+e^{\beta(b+x)}+ \tag{4.18}$$

$$\sqrt{4(e^{2\beta(a+b)}+e^{\alpha+2\beta a+\beta b}+e^{\alpha+\beta a+2\beta b}+e^{2\alpha+\beta(a+b)})+e^{2\beta(a+x)}+e^{2\beta(b+x)}-2e^{\beta(a+b+2x)}})]$$

$$k > \frac{1}{\beta}Log[0.5e^{-\alpha}(\frac{e^{\beta*x}(e^{\beta*a}-e^{\beta*b})}{e^{\alpha}+e^{\beta*b}}+\sqrt{\frac{e^{2\beta*x}(e^{\beta*a}-e^{\beta*b})^2+4e^{2\alpha}(e^{\alpha}+e^{\beta*a})(e^{\alpha}+e^{\beta*b})}{(e^{\alpha}+e^{\beta*b})^2}})]$$

Additionally, the behavior of the probability distribution function for false positive should correspond to the relationship between false positives and k. From intuition, it is expected that if the control limits become wider it is less likely that a pure batch is classified as contaminated. $p_{FP}(k)$ decreases in k; for a larger k, the number of process parameter values considered decreases causing the sum of probabilities to decrease as well. For the function to be in line with this logic, the following property should hold: $\frac{dp_{FP}(k)}{dk} \leq 0$. The function for the first derivative ensures that for all x and k values this property holds.

The convexity property for the probability on having a false positive can be expressed in a mathematical equation: $\frac{d^2p_{FP}(k)}{dk^2} \geq 0$. Whether or not the second derivative becomes negative depends on x, and its interaction with a and b. $\frac{d^2p_{FP}(k)}{dk^2} \geq 0$ is not influenced by the value of k as it either holds for all k values or for none of the k values. Equation 4.19 models the boundary condition for x for which for convexity can be ensured.

$$x < \frac{\alpha}{\beta} \tag{4.19}$$

Continuing with the probability distribution function for having a false negative, the false negative function should satisfy the probability property, having a result between zero and one. Equation 4.20 and 4.21 present the boundary conditions for x and k for which this property holds.

$$x < \frac{Log[\frac{e^{\alpha+\beta*k}(-e^{\beta*a}+e^{\beta*b})}{e^{\beta(2k+a)}-e^{\beta*b}}]}{\beta} \tag{4.20}$$

$$k < \frac{Log[0.5e^{-\beta(a+x)}(-e^{\alpha+\beta*a}+e^{\alpha+\beta*b}+\sqrt{4e^{\beta*a+\beta*b+2\beta*x}+e^{2\alpha}(e^{\beta*a}-e^{\beta*b})^2})]}{\beta} \tag{4.21}$$

The first derivative of the false negative formula should be in line with intuition formed on the relation between the probability on having a false negative and k. The intuition is that if the control limits become wider it is more likely that a contaminated batch is classified as pure. $p_{FN}(k)$ increases in k; for a larger k, the number of process parameter values considered increases causing the sum of probabilities to also increase. In mathematical terms, this means that the first derivative should be bigger than or equal to zero $\frac{dp_{FN}(k)}{dk} \geq 0$. The formula for the first derivative for false negatives ensures that this holds for all values of x and k.

Moreover, the function for false negatives should satisfy the convexity property meaning $\frac{d^2p_{FN}(k)}{dk^2} \geq 0$. Similar to the probability function for false positives, there will only be a boundary condition for x, not for k because the second derivative is either positive or negative over the whole range of k. The boundary condition for x is illustrated in Equation 4.22

$$x < \frac{\alpha}{\beta} \tag{4.22}$$

Thirdly, the sum of $p_{FN}(k)$ and $p_{FP}(k)$ is analyzed; this sum models the behavior of the objective function. For the model it is assumed that weights are applied to the two types of misclassification: $p_{FP}(k)$ and $p_{FN}(k)$ are multiplied with γ and $(1-\gamma)$ where γ represents the weight and has a value between 0 and 1. Both functions, $p_{FP}(k)$ and $p_{FN}(k)$, should be convex for the sum to be convex and have an optimal k*. Convexity can be determined by taking the second derivative which should be strictly positive. The probability functions for false negative and false positive have been analyzed and for both functions the convexity property holds if the boundary condition x¡$\frac{\alpha}{\beta}$ is respected.

$$\gamma * p_{FP}(k) + (1-\gamma) * p_{FN}(k) \tag{4.23}$$

where γ and (1-γ) model the weights for false positives and false negatives.

4.3.4 Two Parameter Model

The preliminary model can be expanded to a model that uses two process parameters instead of one to classify the purity of a batch. Using two process parameters should increase the accuracy of the classification as it is evaluates multiple characteristics of the process. The increase of classification accuracy can be explained using the reasoning behind the Swiss Cheese model. The Swiss Cheese model shown in Figure 4.2 is a model well-known in safety research (Larouzee & Le Coze, 2020). The slices of the Swiss cheese represent a organisations defences against failure and the holes illustrate weaknesses in individual parts of a system. These holes differ in position and size representing different type of failures. An accident occurs when an arrow can be drawn straight through the holes. Translating this to purity classification, one slice represents one process parameter and a hole models the case where a batch is contaminated but still respects the control limits of that process parameter. When using multiple parameters, the probability of having a trajectory of holes (straight line through the holes) that leads to a misclassification of batch decreases as the number of slices increases.

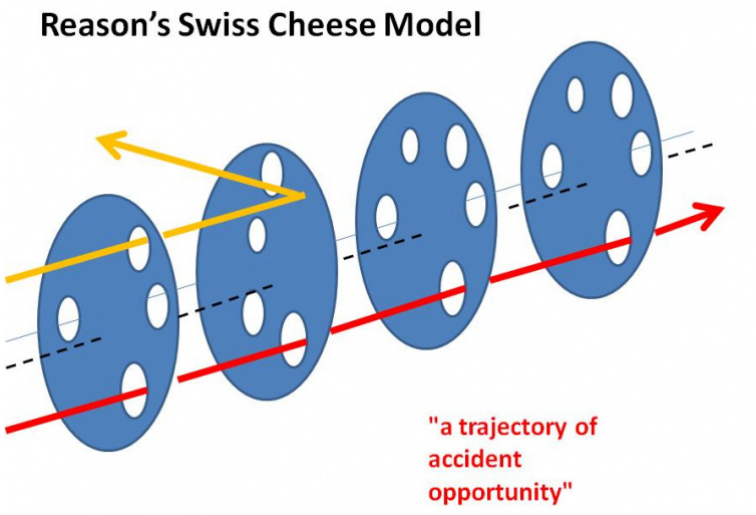


Figure 4.2: Swiss Cheese Model

Objective: The objective is to identify the combination of control limits for both process parameters that minimizes the total cost resulting from quality risk and waste risk. As two process parameters are modelled, the effect of the interaction between the two control limits on the probability of having a certain type of misclassification should be considered in the modelling process. In other words, the control limits cannot be determined independently. The cost for misclassification is modelled with $p_{FP}(k_1, k_2)$ and $p_{FN}(k_1, k_2)$ with weights C_{FP} and C_{FN}. These weights can represent absolute costs or cost ratios. The probability on having a false negative or false positive is dependent on k_1 and k_2 where k_1 and k_2 resemble the control limit coefficients for process parameter 1 and process parameter 2.

$$\underset{k_1 \geq 0, k_2 \geq 0}{\text{Minimize}}\ p_{FN}(k_1, k_2) * C_{FN} + p_{FP}(k_1, k_2) * C_{FP} \tag{4.24}$$

The probability on having a false negative or false negative with two process parameters can be modelled by integrating over the two process parameters. The probability on having a false negative is fairly similar to the probability function with one process parameter; the main difference is the second integral that has been added to integrate over two process parameters. The equation for this probability is presented in Equation 4.25.

$$p_{FN}(k_1, k_2) = \int_{q=x_1-k_1}^{x_1+k_1} \int_{h=x_2-k_2}^{x_2+k_2} (1 - p(q, h)) * f_1(q) * f_2(h) * dq * dh \tag{4.25}$$

The probability on having a false positive can be calculated by modelling the three scenarios in which a false positive occurs: the batch is pure and, (1) process parameter 1 and 2 are both outside their control limits or (2) process parameter 1 is outside its control limits and process parameter 2 is within its control limits or (3) process parameter 2 is within its control limits and process parameter 2 is outside its control limits. These scenarios differ in the boundary conditions for the integrals. Equation 4.26 presents the formula for the probability of having a false positive. The derivation of this equation can be found in Appendix B.

$$p_{FP}(k_1,k_2) = \int_{q=a_1}^{x_1-k_1}\int_{h=a_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1+k_1}^{b_1}\int_{h=a_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1-k_1}^{x_1+k_1}\int_{h=a_2}^{x_2-k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1-k_1}^{x_1+k_1}\int_{h=x_2+k_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh \quad (4.26)$$

Variables

Decision Variables:

The control limits of critical process parameter 1 and 2 are modelled with k_i and x_i with i $\in$ {1,2}. x_i, the control line, is set at the mean of the process parameter and k_i, the control line coefficient, models the width of the control limit and is the decision variable of this model.

Input Variables:

The probability distribution for a batch being pure is modelled with $p(v_1, v_2)$ and dependent on the values of both process parameters where v_1 and v_2 represent the value for process parameter 1 and 2, respectively. $p(v_1, v_2)$ can include an interaction term between process parameter 1 and process parameter 2, in that case $\beta_3 > 0$, otherwise β_3 is equal to zero. The function for $p(v_1, v_2)$ as presented Equation 4.27 is an example of how the function can look like. The relations between the process parameters and the probability of a batch being pure can be positive or negative, changing the plus and minus signs.

$$p(v_1,v_2) = \frac{e^{\alpha-\beta_1*v_1+\beta_2*v_2+\beta_3*v_1*v_2}}{1-e^{\alpha-\beta_1*v_1+\beta_2*v_2+\beta_3*v_1*v_2}} \quad (4.27)$$

The boundaries for the measurement domains for the process parameters are modelled with a_1 and b_1, and a_2 and b_2 where a_i and b_i with i $\in$ {1,2} are the lower and upper limit of the measurement domain. The x_i with i $\in$ {1,2} is set to the average process parameter value for process parameter i.

$f_i(h)$ with i $\in$ {1,2} is specified for each process parameter and models the probability that process parameter i has value h. It is assumed that the probability functions $f_1(h)$ and $f_2(h)$ are independent of each other. Both the normal and uniform distribution can be used to model this probability. The model will be build upon the assumption that $f_i(h)$ is uniformly distributed to ensure enough weight on the parameter values resulting from a contamination batch and to simplify modelling. This assumption could be relaxed in a later stage to determine the impact of this assumption.

Constraints:

The constraint of the measurement domain on the boundaries of the control limits are formulated for each process parameter individually. The control limits of a particular process parameter should fall within the measurement domain to avoid extrapolation (see Equations 4.28 till 4.28).

$$x_1 - k_1 \geq a_1 \quad (4.28)$$

$$x_1 + k_1 \leq b_1 \quad (4.29)$$

$$x_2 - k_2 \geq a_2 \quad (4.30)$$

$$x_2 + k_2 \leq b_2 \tag{4.31}$$

Complete Model

$$\underset{k_1 \geq 0, k_2 \geq 0}{\text{Minimize}} \left(\int_{q=x_1-k_1}^{x_1+k_1} \int_{h=x_2-k_2}^{x_2+k_2} (1 - p(q,h)) * f_1(q) * f_2(h) * dq * dh \right) * C_{FN} \tag{4.32}$$

where x_i is the middle of the measurement domain ($\frac{a_i+b_i}{2}$).

$$\left(\int_{q=a_1}^{x_1-k_1} \int_{h=a_2}^{b_2} p(q,h) * f_1(q) * f_2(h) * dq * dh + \int_{q=x_1+k_1}^{b_1} \int_{h=a_2}^{b_2} p(q,h) * f_1(q) * f_2(h) * dq * dh + \right.$$

$$\left. \int_{q=x_1-k_1}^{x_1+k_1} \int_{h=a_2}^{x_2-k_2} p(q,h) * f_1(q) * f_2(h) * dq * dh + \int_{q=x_1-k_1}^{x_1+k_1} \int_{h=x_2+k_2}^{b_2} p(q,h) * f_1(q) * f_2(h) * dq * dh \right) * C_{FP}$$

Subject to:

$$x_1 - k_1 \geq a_1$$

$$x_1 + k_1 \leq b_1$$

$$x_2 - k_2 \geq a_2$$

$$x_2 + k_2 \leq b_2$$

4.4 Numerical Analysis

The purity model is validated by conducting a numerical analysis based on a bacteriological production process at MSD Animal Health. The two main goals of this analysis is (1) to show the process of identifying the process parameters critical to purity classification and (2) to validate the purity models for both one and two process parameters and show the valuable insights that can gained from these models. The identification of the critical process parameter also illustrates the connection between growth curve modelling and the purity model.

MSD has collected historical batch data, including process parameter data and quality data, from the last two years for one product on a particular production line. However, the numerous batches included in this data set are all non-contaminated batches. To generate data on contaminated batches, an experiment is conducted during which multiple batches have been contaminated varying the degree of contamination and the moment of contaminating. The batches are contaminated with the bacteria that appears to be the most occurring contaminating bacteria for contamination events in production. The contamination is done either at the beginning of the fermentation process or just before the end of the fermentation process. It is chosen to test these two moments as it is most likely that, if a contamination takes place, it will be either at the beginning or at end of the production process. Especially a contamination near the end of the process is an interesting scenario because detecting such type of contamination could be more challenging compared to a contamination at the start of the process. The influence of the degree of contamination on the detectability is evaluated by injecting either a low, middle or high concentration of the contaminating bacteria. The amount of bacteria injected can influence the growth rate of the contaminating bacteria and with this the detectability. Also, control batches have been included in the experiment which are

batches that are not contaminated (pure batches). For the experiment, process settings and raw materials are kept the same as for commercial production although corrected for the difference in scale. For data analysis purposes, the data sets are normalized to correct for this scale difference.

The RTRT approach for purity has been formulated in close collaboration with the stakeholders at MSD. Weekly meetings took place with the lead of the production department to gain understanding of the production process and product, to reflect on proposed models and to interpret results. In addition, regular meetings with operators provided additional process insight and product knowledge. Moreover, during monthly meetings with higher management the progress was shared and feedback from a management perspective was provided. The main purpose of these meetings was to understand what managerial insights would be interesting for MSD, also for potential future RTRT efforts. Furthermore, regular stakeholder meetings with the qualified person from the quality department responsible for releasing the products in scope and the regulatory affairs team responsible for the managing the dossier for the products in scope gave valuable insights on the results and the requirements for the approach and validation. The qualified person challenged certain modelling aspects from a quality perspective and provided support in the interpretation of validation results. The regulatory team provided information on validation requirements for dossier changes when introducing RTRT. In addition, meetings with the department responsible for conducting the experiments were held to design the experiment and reflect on the experimentation results.

The project on the RTRT approach for purity started in April 2020. From April 2020 to June 2020 understanding of the production process and product were developed through regular visits at the production lines and meetings with operators. During June 2020-October 2020, the model was build with various iterations using the input from stakeholder meetings. In the same period, in close collaboration with experts at BTS the experiment was designed. In October 2020 and December 2020 the first and second experiment at BTS for obtaining contaminating data took place. During January 2021-March 2021 the model was validated using the data from the experiments. The results and approach for purity have been shared during Report Out which is a company wide presentation where various people expressed their recognition on the potential and impact on this research for batch release.

First, Section 4.4.1 discusses the identification of the process parameters critical to purity. This information is used for the model validation for the preliminary model discussed in Section 4.4.2 and for the model validation for the two parameter model discussed in Section 4.4.3. In Section 4.4.2 and 4.4.3, the construction and validation of the model is discussed. In addition, a sensitivity analysis is conducted and the model performance in compared to benchmark models. Lastly, the model is compared to the current purity test.

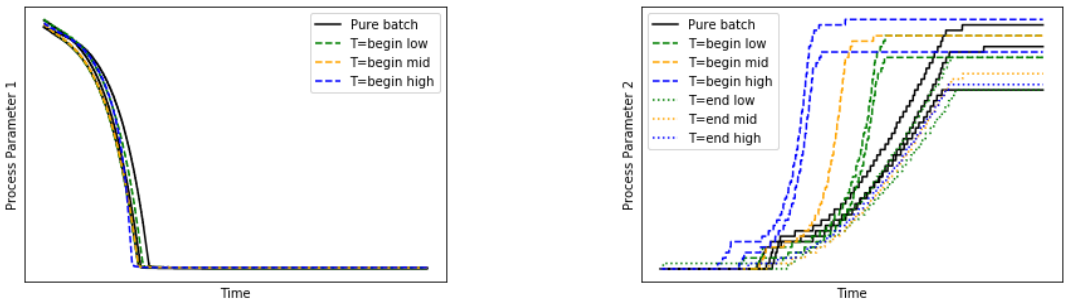
4.4.1 Identification of Process Parameters Critical to Purity

Before the models can be constructed and validated, it is essential to determine which process parameters are critical to purity and can serve as input for the models. The process parameters critical to purity are process parameters for which the value differs significantly between contaminated batches and pure batches. Defining these process parameters is done combining the insights from data analysis and expert

knowledge; data analysis focuses on identification based on quantitative grounds and expert knowledge focuses on the theory of biological relations. For the one parameter model, this step is especially important because only one parameter is used therefore it has to be determined which parameter is most critical for identifying purity.

First, expert knowledge is leveraged to form an initial thought on potentially relevant process parameters. This initial thought can be further verified and specified using data analysis. From conversations with process experts it can be concluded that especially process parameters measured real-time are interesting for the classification of purity. These parameters provide information on the behaviour of the process whereas offline process parameters only provide information on the state of the process at the time of the measurement. This is supported by the case study of Torres (2017). For this reason, it is decided to focus on real-time measured process parameters. In this case study, there are two of such parameters, referred to as process parameter 1 and 2. For the analysis of process parameter 1 only the pure batches and the batches contaminated at the start of the process are considered because this parameter is limited to providing information on the lag phase and early part of the exponential growth phase. From Figure 4.3a, it shows that no significant differences can be observed between non-contaminated and contaminated batches. This, in combination with the knowledge that this parameter cannot be used to detect contamination later during the process, leads us to conclude that process parameter 1 is not suitable for defining the purity of a batch. Process parameter 2 is presented in Figure 4.3b. In contrast to process parameter 1, significant differences are observed when comparing the curves of batches contaminated at the beginning of the process and pure batches. Batches contaminated at the beginning of the process have a shorter lag phase, meaning the curve starts increasing at an earlier point in time, and the rate at which the curve increases is higher compared to the pure batches. Batches contaminated at the end of the process are harder to detect when comparing to pure batches. This is to be expected because at the moment of contamination the process parameter is already stabilized at its maximum value. This stabilization is the result of the depletion of nutrients; the bacteria used for contaminating does not have the essential nutrients to grow. Therefore, its growth is limited to none causing no observable changes to the curve. In conclusion, process parameter 2 is critical to purity classification for contaminations early in the process however contaminations late in the process are difficult to identify. As no other real-time process parameters have been measured during this experiment due to limitations in scale, it is decided to use process parameter 2 for purity classification however keeping in mind the limitation for contaminations late in the process.

The data for process parameter 2 for one batch includes a data point for every 10 seconds which together forms a curve. Defining control limits for every single data point is not feasible but instead, since it is required for RTRT to have numeric control limits, control limits can be formulated for specific characteristics of the curve. The growth model presented in Chapter 3 can be used to translate the characteristics of a curve in quantified curve parameters such as growth rate. Process parameter 2 is an indirectly controlled parameter and can therefore be modelled best with the adapted model for indirectly controlled parameters presented in Equation 3.4. This parameter is also included in the numerical analysis



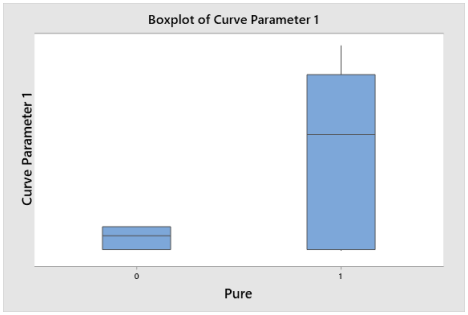
(a) The curve of process parameter 1 for the pure batches and early contaminated batches

(b) The curve of process parameter 2 for all experimental batches

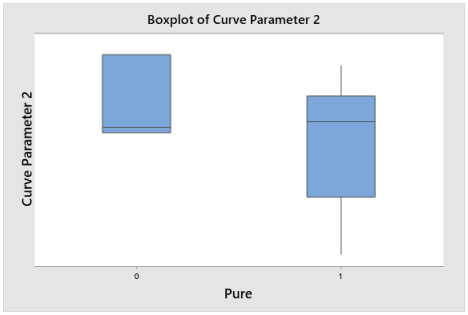
Figure 4.3: Data analysis on experimental data for predicting purity with two real-time measured process parameters

for growth curve modelling where the accurate fit of this model was proven. The model generates the following curve parameters; the growth rate, the asymptotic value, the lambda and the maximum value (constant). Data analysis can provide more insight in which of these four parameters are critical to purity, or in other words, have a significant different value for pure batches in comparison to contaminated batches. According to Allman (2020) growth rate and lambda could potentially be critical to purity. The contaminating bacteria is often a faster growing bacteria compared to the cultivating bacteria causing the growth rate to be higher and the growth to start earlier when a contamination takes place. However, this is often not the case for contaminations late in the process because lambda and growth rate are curve parameters that model characteristics of the lag phase and exponential growth phase and in case of a late contamination the process is often in stationary or death phase.

Data analysis is used to verify the expectations from experts and literature. Due to confidentiality the names of the curve parameters are changed to curve parameter 1, curve parameter 2, curve parameter 3 and curve parameter 4. First, a graphical analysis is conducted whereafter the curve parameters are analyzed in a more quantitative matter using statistical analysis. Starting with the graphical analysis, Figure 4.4 and 4.5 illustrate a boxplot including contaminated batches (Y=0) and pure batches (Y=1) for each curve parameter. The boxplots for curve parameter 3 and curve parameter 4 show no significant difference between contaminated and pure batches; the position and width of the boxplots are similar. The curve parameter 1 and curve parameter 2, on the other hand, do show an observable difference in terms of the position and width of the boxplot. Pure batches have, on average, a lower value for curve parameter 1 compared to contaminated batches. For curve parameter 2, it can be observed that, although the median is not significantly different, pure batches range more to higher parameter values in contrast to pure batches that range more to small parameter values. For a more in-depth analysis it is chosen to exclude the batches contaminated at the end as it has been found that these batches are hard to distinguish using the curve data. The results for this analysis are presented in Figure 4.6. Especially the difference in curve parameter 1 becomes even more significant; the parameter value for batches contaminated at the start of the process is substantially higher compared to that of the pure batches. For curve parameter 2, the parameter value for batches contaminated at the start of the process is lower compared to the parameter value for pure batches although this difference is not substantial. Furthermore, the boxplot for curve

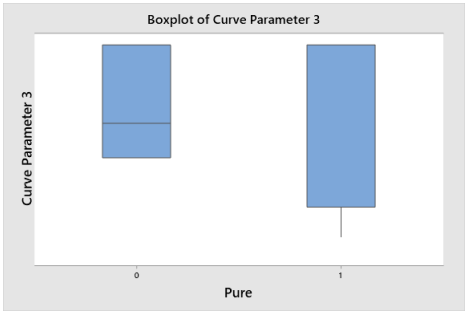


(a) Boxplot for curve parameter 1

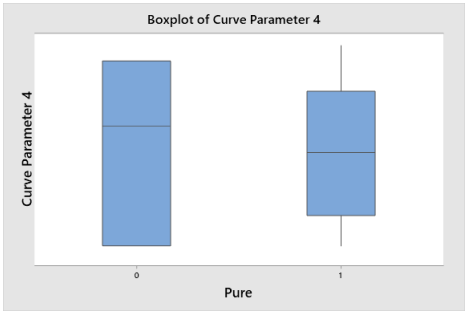


(b) Boxplot for curve parameter 2

Figure 4.4: Graphical analysis of contaminated batches (Y=0) and pure batches (Y=1) for curve parameters 1 and 2



(a) Boxplot for curve parameter 3



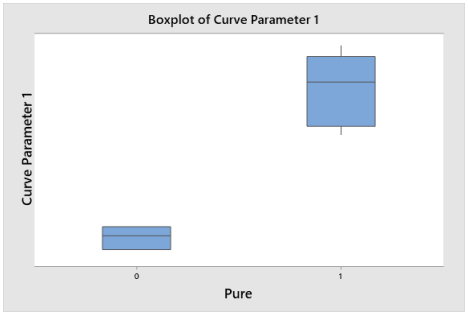
(b) Boxplot for curve parameter 4

Figure 4.5: Graphical analysis of contaminated batches (Y=0) and pure batches (Y=1) for curve parameters 3 and 4

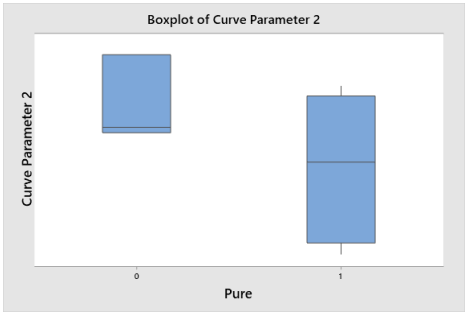
parameter 2 in Figure 4.6 shows a lower median and more even spread compared to the boxplot in Figure 4.4.

A statistical analysis can provide more insight in whether or not there is a significant difference between contaminated and pure batches. This is especially relevant for curve parameter 2 since no clear conclusion can be drawn from the graphical analysis. A 2-sample t-test is performed with as null hypothesis 'the means do not differ significantly'. Including all batches, the analysis showed that for the curve parameter 1 the mean for contaminated batches is significantly different compared to the mean for pure batches (p-value=0.033). For the other three curve parameters, the means found not significantly different. The null hypothesis was supported with p-values 0.283 for curve parameter 2, 0.933 for curve parameter 3 and 0.936 for curve parameter 4. Repeating the t-test with including the pure batches and the batches contaminated at the start of the process gave a similar result. The p-values where slightly lower but the conclusions remained the same.

In conclusion, data analysis has shown that curve parameter 1 is most critical to purity. Curve parameter 2 can potentially be insightful when using two parameters for purity classification but evaluating purity solely on this parameter would not be advised. These result are supported by the study of Allman (2020). Nevertheless, batches contaminated close to the end of the process remain hard to distinguish from pure batches. This is because curve parameters 1 and 2 represent characteristics of the lag and exponential growth phase and a contamination near the end of the process affects, in most cases, only the phase where the process parameter has already reached its constant value.



(a) Boxplot for curve parameter 1



(b) Boxplot for curve parameter 2

Figure 4.6: Graphical analysis of pure batches (Y=1) and batches contaminated (Y=0) at the end of the process for curve parameters 1 and 2

4.4.2 Model Validation for Preliminary Model

This section present the preliminary model used to determine the optimal control limits for purity classification based on one process parameter. First, Section 4.4.2 discusses the input for the model. In addition, this section explains in more depth the construction of the model for defining the optimal control limits and the mathematical properties of the model are evaluated. Subsequently, Section 4.4.2 presents a sensitivity analysis and comparison to two well-known benchmarks in the literature stream of quality and reliability in processes. At last, Section 4.4.2 compares the model results to the existing testing method.

Construction and Validation of Preliminary Model

The preliminary model uses the value of one process parameter to classify a batch as either pure, if the value is between the control limits, or contaminated, otherwise. From Section 4.4.1 can be concluded that the curve parameter 1 of the curve of process parameter 2 would be the most suitable parameter for purity prediction with the preliminary model. The first step is to describe the relation between curve parameter 1 and purity whereafter this relation can be expressed in the form of a probability distribution function. Curve parameter 1 is negatively related to purity meaning a pure batch has a lower parameter value in comparison to a batch that is contaminated. This implies that the probability of having a pure batch, expressed as $p(v)$, decreases in the parameter value of curve parameter 1. Moreover, it tells us that the control limit for this curve parameter is one-sided with an upper bound. Both biology and the results from the experiments conclude that curve parameter 1 increases when a contamination occurs. Having a lower bound for curve parameter 1, meaning a parameter value below a certain threshold is indicative for a contamination, is contradicting with the statements from literature and the observations from the experiment.

The probability distribution function for curve parameter 1 can be determined using logistic binary regression analysis. A logistic regression model uses a set of predictor variables (X) to model the dichotomous response variable (Y), often denoting the occurrence of an event (Harrell, 2015). The method of maximum likelihood is used to determine the regression parameters. The potential x variables for the distribution function for purity based on curve parameter 1 is (1) the parameter value (x) and (2) the quadratic term of the parameter value (x^2). A quadratic term for curve parameter 1 is not considered

for the regression analysis because this type of behavior is not in line with the theory and observed experimental results. Equation 4.33 models $p(v)$ representing the probability of having a pure batch dependent on curve parameter 1. The first expression can be written as a linear relation by taking the Ln of the odds ratio $(p(v)/(1-p(v)))$ for easier interpretation. β_0 is the interception with the y-as; if the parameter value is equal to zero, $p(v)$ is equal to $e^{\beta_0}/(1+e^{\beta_0})$ resulting in a value close to one. The impact of a change in curve parameter 1 on the probability of having a pure batch is modelled with β_1. A increase of curve parameter 1 by 1 decreases the Ln of the odds ratio by β_1. The fit of the model is evaluated by comparing the p-value with the 95 % significance level; the p-value is 0.007 which is below the threshold of 0.05. From this can be concluded that association between curve parameter 1 and purity is statistically significant and the model fit is acceptable.

$$p(v) = e^{\beta_0-\beta_1*v}/(1+e^{\beta_0-\beta_1*v}) = e^{5.94-2.65*v}/(1+e^{5.94-2.65*v}) \tag{4.33}$$

$$Ln(p(v)/(1-p(v))) = \beta_0 - \beta_1 * v = 5.94 - 2.65 * v$$

The boundaries for the measurement domain, modelled as a and b, can be determined from the data and are set equal to the minimum (0.59) and maximum (4.38) value for curve parameter 1, respectively. The center line of the control limit, x, is equal to the average value for curve parameter 1 (1.13). The probability of having a parameter value equal to h, $f(h)$, is assumed to be uniformly distributed hence can be modelled with Equation 4.34. This assumption is validated numerically with several other distributions; the managerial insights remained the same with another distribution assumptions for $f(h)$.

$$f(h) = \frac{1}{b-a} = \frac{1}{4.38-0.59} \tag{4.34}$$

The last input for the model is the weights for false positive and false negative, C_FP and C_FN. There are different methods for setting the weights: (1) equal weights for false positive and false negative can be assumed or (2) ,more in line with practice, setting the weights in a way that it represents the impact in terms of monetary cost and quality risk. However, the cost and quality risk as a result of a misclassification are hard to quantify. For the model construction and validation, it is assumed that the weights are equal. This assumption will be relaxed in the sensitivity analysis where various weights are applied. This will provide insight in the impact of the weights on the optimal control limit.

Continuing with further specifying the model, the preliminary model as described in Section 4.3.2 assumes a two-sided control limit (upper and lower limit) whereas for curve parameter 1 an one-sided control limit is assumed. This adaption requires changing the boundaries of the integrals for the probability on having a false positive or false negative. The resulting expressions are presented in Equation 4.36 and 4.35. $p(h)$ and $f(h)$ are input for the model and have been discussed in the previous section.

$$p_{FP}(k) = \int_{h=x+k}^{b} p(h) * f(h) dh \tag{4.35}$$

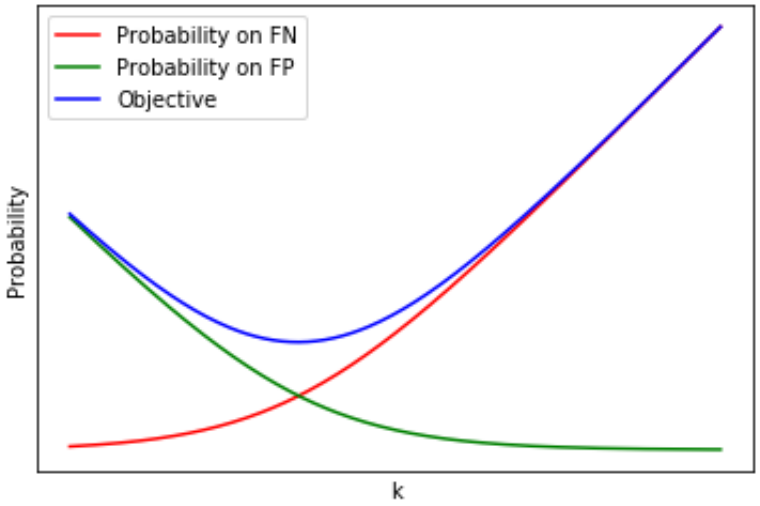


Figure 4.7: Preliminary model for curve parameter 1 with equal weights for false positives and false negatives

$$p_{FN}(k) = \int_{h=a}^{x+k} (1 - p(h)) * f(h)dh \tag{4.36}$$

For this section it is assumed that impact of the consequences resulting from false positive and false negative misclassifications are equal. The results are presented in Table 4.1 and Figure 4.7. Figure 4.7 illustrates the curve for the probability of having a false negative (p_{FN})), the probability of having a false positive (p_{FP})) and the sum of these probabilities. When k becomes larger, the upper limit becomes higher classifying batches with a higher value for curve parameter 1 also as pure although the probability of a batch being pure decreases in curve parameter 1. As a result, p_{FN}) increases because more impure batches will be classified as pure and p_{FP}) decreases as less batches will be falsely identified as impure. The former has a negative influence on quality whereas the latter has a positive influence on quality but results in unnecessary waste. The sum of p_{FP}) and p_{FN}) is the objective function of the model and is minimized to determine the optimal trade-off between the probability of having a false negative and the probability of having a false positive. The minimal objective value results from a k of 1.12 resulting in an upper control limit of 2.25; this k is defined as k*. The performance of this control limit in terms of misclassifications can be evaluated using the probabilities p_{FP} and p_{FN} or by calculating the number of misclassifications in the data set assuming a control limit of 2.25. Starting with the former, p_{FP} and p_{FN} are approximately equal meaning the probability of misclassifying a pure batch as impure is the same the probability of misclassifying a impure batch as pure. The analysis of misclassifications using the data set, shows that there were no false positives however the batches that are contaminated at the end of the process were classified as pure. This observation was to be expected as the value for curve parameter 1 of these batches was not significantly different from that of pure batches for earlier explained reasons.

Moreover, the mathematical properties are evaluated for the obtained k* and corresponding parameter configuration. Section 4.3.3 discusses the mathematical properties for the preliminary model and formulates boundary conditions for x and k for which these properties hold. However, these boundary conditions

are based on two-sided control limits so have to be adapted for the model based on an one-sided control (only upper limit). The false positive distribution function changes to the form presented in Equation 4.16 which models the false positives for process parameter values exceeding the upper limit. From the boundary conditions formulated for this equation (Equation 4.14 and 4.13) it can be concluded that the probability property holds for the parameter configuration with k^* is 1.12. The first derivative of Equation 4.16 ensures that the probability of having a false positive decreases in k for all x and k values therefore it can be concluded that this property holds. Moreover, it is found that the second derivative ensures that the convexity property of the false positive distribution function holds for all x and k values. Continuing with the false negative distribution function, the boundary conditions for x and k are defined for the probability distribution for false negative based on only an upper limit and show that with the given parameter configuration and k^* the probability property holds. The first derivative of the probability distribution ensures that for all x and k values the probability of having a false negative increases in k. The same applies to the convexity property which holds for all values of x and k. As both false negative and false positive distribution functions are convex it follows from Theorem **??** that the objective function is also convex.

Sensitivity Analysis and Benchmarking

The model with the assumption of equal weights produced a reasonable control limit for curve parameter 1 however the assumption of equal weights is not in line with reality. Therefore, this assumption is relaxed and the influence of the weights on the optimal control limit is investigated by conducting a sensitivity analysis. The impact of the weights is investigated by using different cost ratio combinations and evaluating the influence on k^*. At first, it is important to understand the relationship between cost and a certain misclassification. The process consists of the upstream and the downstream process. Purity is determined at the end of the upstream process after the main fermentation process. A batch is classified as pure if the value for curve parameter 1 is below the control limit; a pure batch will continue to the downstream process where further value adding activities that result in extra cost are conducted. A batch classified as contaminated will not continue to the downstream processes but will be disposed as waste. Translating this to misclassifications, a false negative batch will continue to the downstream process where the sterility test, conducted at the end of the downstream process, will identify that this batch is contaminated. This batch cannot be used for further processing so will be disposed as waste. Hence, the cost for this classification includes the sunk cost from the upstream process and extra cost from the downstream process (and quality control testing). Also, important to keep in mind is that quality risks can involved with further producing a contaminated batch. A false positive batch is classified as contaminated at the end of the upstream process hence only the sunk cost from the upstream process are allocated. Although, it is good to keep in mind that throwing away a pure batch is unnecessary waste resulting in lost profit. In terms of weights, it can be recognized that false negatives should have a higher weight compared to false positives. Quantifying the precise weights remains difficult but based on the cost description for misclassification a rough estimation of the cost ratios can be made. Table 4.1 presents seven different ratios for $\frac{C_{FN}}{C_{FP}}$ with the corresponding model results. In the bio-pharmaceutical industry false negative misclassifications often have higher consequences in terms of costs and quality risk compared to

false positives therefore a ratio bigger than 1 is the most representative of real-life practice. The scenarios with a ratio smaller than 1 are fairly unrealistic but can provide valuable insights. From Figure 4.9 and Table 4.1 can be observed that the optimal k decreases when the ratio between false negatives and false positives ($\frac{C_{FN}}{C_{FP}}$) increases. This is in line with expectation; the risk on having a contaminated batch be classified as pure (false negative) increases in k. p_{FP} increases in ($\frac{C_{FN}}{C_{FP}}$) whereas p_{FN} decreases in ($\frac{C_{FN}}{C_{FP}}$). This is a result of the decreasing k when the ratio between false negatives and false positives becomes larger. Moreover, the objective value becomes smaller as the ratios become more extreme.

The change in k^* does not translate directly to the number of false negatives and false positives found in the data set provided by MSD when using the RTRT approach. For almost all scenarios, zero batches are classified as false positive and batches that are contaminated at the end of the process are classified as false negatives. This could be a result from the limited amount of data on contaminated batches; only data of contaminations occurring either at the beginning or the end of the process are included in this data set. The detection of contaminations occurring midway during the process is likely to be more sensitive to k^* although it is unlikely that a contamination will occur at this time. Bram van Ravenstein, lead from the bacteriological processing department, commented on contaminations occurring and mentioned that when such contamination occur this is observed in the visual inspection by operators therefore it is expected that these contaminations influence process parameters and are detected by the RTRT model. This supports the statement that k most likely influences the detection of contaminations occurring midway in during the process.

The results obtained from the optimization model are compared with a benchmark well-known in process control literature, the six sigma method ((Joghee, 2017)). This method is used to increase the performance of production processes and improve quality by decreasing process variation. The name of the method reveals how the control limits are constructed namely six times the standard deviation from the center line (see Figure 4.8). Table 4.2 shows the results and presents the percentage improvement resulting from using the preliminary model presented in this study. The k* obtained from using the Six Sigma method is significantly higher compared to the k^* resulting from the preliminary model and even exceeds the boundaries of the measurement domain. The probability of having a false negative is high and as explained in the previous paragraph especially a false negative has substantial negative consequences when occurring in a production environment. “Factors of Safety and Reliability in Geotechnical Engineering” (n.d.) discusses similar benchmark named the Thee Sigma Rule which constructs the control limits at three times the standard deviation from the center line. This method still underperforms compared to the preliminary model however the k^* is more close to the k^* obtained from the optimization model compared to the Six Sigma method. In addition, the control limits of the two benchmarks are used for classifying the batches included in the data set of MSD where it is found that with the 6 sigma method only 1 of the 9 contaminated batches is detected. For the three sigma method the misclassification are the same as for the optimization model; the contaminations introduced late in the process are not detected.

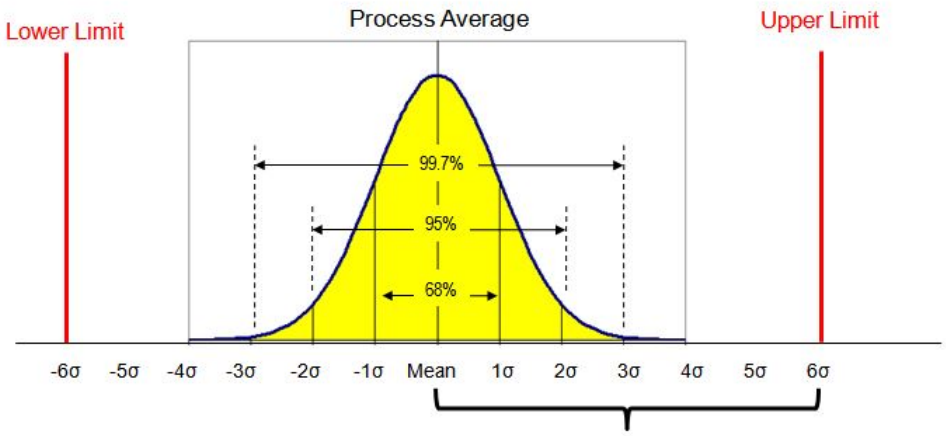


Figure 4.8: Graphical representation of the Six Sigma method

$\frac{C_{FN}}{C_{FP}}$	p_{FP}	p_{FN}	Obj	k*
0.25	0.0219	0.1590	0.0494	1.64
0.5	0.0400	0.1084	0.0628	1.38
1 (Base)	0.0685	0.0682	0.0683	1.12
1.5	0.0913	0.0495	0.0662	0.96
2	0.1095	0.0390	0.0625	0.85
2.5	0.1243	0.0324	0.0587	0.77
3	0.1379	0.0274	0.0550	0.70

Table 4.1: Sensitivity analysis results for the influence of the weights for false negatives and false positives on the model performance for the preliminary model

Model	p_{FP}	p_{FN}	Obj	k*	Improvement Compared to Preliminary Model
Preliminary Model	0.0685	0.0682	0.0683	1.12	-
6 sigma Method	0	0.6156	0.3078	3.4524	77.79%
3 sigma Method	0.0177	0.1779	0.0978	1.7262	30.11%

Table 4.2: Quantitative comparison of preliminary model to the 6 sigma and 3 sigma method

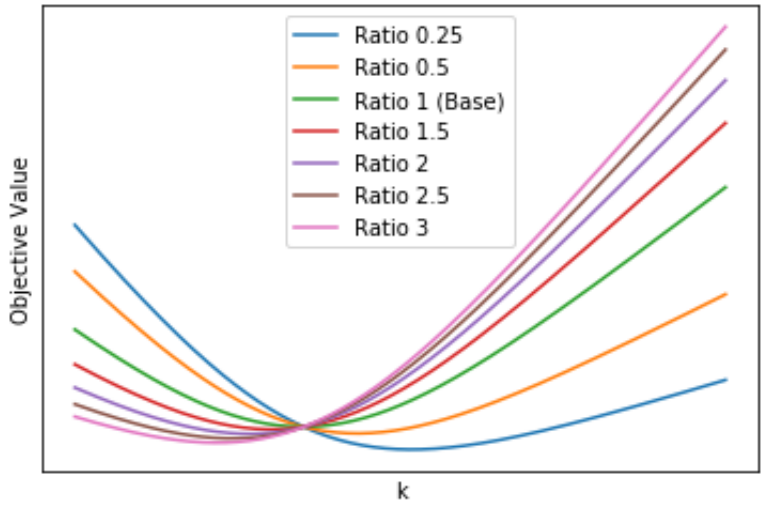


Figure 4.9: The objective value plotted for the different ratios between C_{FP} and C_{FN} for the preliminary model

Comparison to Traditional Purity Test

In the requirements on RTRT is described that the RTRT strategy should be superior or equivalent to the existing testing method. The traditional testing method for purity is taking a sample and incubating this sample to determine if there is any growth from a bacteria other than the bacteria cultivated. An experiment has been conducted for this research where batches are contaminated in a controlled way to study the effect of contamination on various process parameters. This experiment is also used to test the performance of the traditional purity test. At end of the experiment a sample is taken from each batch included in the experiment and tested for impurities. The results of this analysis shows that batches contaminated near the end of the process are hard to detect using the traditional purity test especially when the degree of contamination is low. Only batches contaminated late in the process with a high degree of contamination are detected by the purity test. The RTRT strategy also shows difficulty with detecting batches that are contaminated at a late stage of the process even when the degree of contamination is high. The comparison of the RTRT strategy to the traditional purity test leads us to conclude that both testing methods have difficulty detecting a contamination occurring at a late stage of the process. However, the traditional test outperforms the RTRT strategy as it is able to detect batches contaminated late in the process but with a high degree of contamination.

4.4.3 Model Validation for Two Parameter Model

The model with two process parameters classifies a batch as either pure or impure based on the value of two process parameters. The control limits of both parameters have to be respected, meaning both parameter values have to be within their control limit, for a batch to be identified as pure. First, Section 4.4.3 explains the process parameters for classification and additional input parameter configuration. This section also discusses adaptions to the mathematical model, presents the optimal control limits and explains the evaluation of the mathematical properties. In Section 4.4.3, a sensitivity analysis is conducted where the effect of multiple factors is analyzed and the percentage improvement of the model compared

to a benchmark from literature is calculated. Lastly, the comparison to the existing testing method is discussed in Section 4.4.3

Construction and Validation of Two Parameter Model

According to Section 4.4.1, curve parameter 1 and 2 would be most suitable for purity prediction when using the two parameter model. In this section was found that the suitability of curve parameter 2 for purity classification is questionable. Analysis with the two parameter model confirmed that curve parameter 2 is not sensitive enough to contaminations to be valuable for purity prediction. For this study, curve parameter 2 is manipulated by using the contamination data of another product. A distribution for curve parameter 2 is determined using the contamination data of this other product and used to generate parameter values for curve parameter 2.

The relation between curve parameter 1 and purity has already been explained in Section 4.3.2. Curve parameter 2 has a positive relation with purity; a lower parameter value is an indication for a potential contamination. The relation between curve parameter 2 and purity indicates the use of a lower (one-sided) control limit. Most often a contamination results in both a value for curve parameter 1 and a lower value for curve parameter 2 which indicates that there may be an interaction between the two parameters. This could be included in the probability distribution model for determining the probability of a batch being pure. Logistic binary regression is used to build the probability distribution function with as dependent variable the probability of a batch being pure and two independent variables curve parameter 1 and 2. The quadratic terms will be excluded for both parameters since this is not in line with the theory on bacteriological behaviour during contamination. As mentioned earlier, the interaction between curve parameter 1 and 2 could be relevant to include in the regression function. Therefore, two regression analyses will be conducted; the first analysis will not consider the interaction term and the second analysis will. The resulting models are compared using the goodness of fit measure Akaike's Information Criterion (AIC). This measure evaluates how well the model fits the data while penalizing for over-fitting; the lower the AIC value the better the goodness of fit Konishi & Kitagawa (2008). The regression model without the interaction term has an AIC of 43.02 and the model with the interaction term has a 45.01. From this can be concluded that the probability distribution function for the purity of a batch without the interaction fits the data best. Equation 4.37 models the probability of having a pure batch with independent variables curve parameter 2 ($p(v_1, v_2)$). The interception with the y-as is modelled with β_0. β_1 and β_2 represent the impact of curve parameter 1 and 2 on purity, respectively. Increasing the value of curve parameter 2 by 1 results in an increase of β_2 in the Ln of the odds ratio for the probability of having a pure batch.

$$p(v_1, v_2) = e^{\beta_0 - \beta_1 * v_1 + \beta_2 * v_2} / (1 + e^{\beta_0 - \beta_1 * v_1 + \beta_2 * v_2}) = e^{-2.12 - 1.57 * v_1 + 0.348 * v_2} / (1 + e^{-2.12 - 1.57 * v_1 + 0.348 * v_2}) \tag{4.37}$$

$$Ln(p(v_1, v_2)/(1 - p(v_1, v_2))) = \beta_0 - \beta_1 * v_1 + \beta_2 * v_2 = -2.12 - 1.57 * v_1 + 0.348 * v_2$$

Further input parameters are the measurement domains for curve parameter 1 and 2, the probability functions for having a certain process parameter value and the values for the center control line. The

measurements domain, a_i and b_i for i $\in \{1, 2\}$, are equal to the minimum and maximum for curve parameter 1 (0.59;4.38) and curve parameter 2 (10.84;27.41). The uniform distribution is used to model the probability distribution functions for having a certain parameter value (f_i for i $\in \{1, 2\}$) using the values for a_i and b_i. The assumption of using an uniform distribution is validated by testing several other distributions however the managerial insights remained the same. The center control line (x_i for i $\in$ {1,2}) is set equal to the average value for curve parameter 1 (1.13) and curve parameter 2 (20.42).

Continuing with further specifying the model, the model formulated in Section 4.3.4 assumes a two-sided control limit however both parameters have a one-sided control limit; curve parameter 2 only has a lower limit and curve parameter 1 only an upper limit. Therefore, the probability functions for false negatives and false positives need to be adapted as presented in Equations 4.38 and 4.39

$$p_{FN}(k_1, k_2) = \int_{q=a_1}^{x_1+k_1} \int_{h=x_2-k_2}^{a_2} (1 - p(q,h)) * f_1(q) * f_2(h) * dq * dh \tag{4.38}$$

$$p_{FP}(k_1, k_2) = \int_{q=x_1+k_1}^{b_1} \int_{h=a_2}^{x_2-k_2} p(q,h) * f_1(q) * f_2(h) * dq * dh + \tag{4.39}$$

$$\int_{q=x_1+k_1}^{b_1} \int_{h=x_2-k_2}^{b_2} p(q,h) * f_1(q) * f_2(h) * dq * dh + \int_{q=a_1}^{x_1+k_1} \int_{h=a_2}^{x_2-k_2} p(q,h) * f_1(q) * f_2(h) * dq * dh$$

The model construction as presented above can now be used to determine the optimal combination of control limits. This optimal combination depends on the weights allocated to false positive and false negative misclassification. For the initial validation of the model, it is assumed that these weights are equal. This assumption will be relaxed in the sensitivity analysis.

This model has, in comparison with the preliminary model, two instead of one decision variables, k_1 and k_2. The optimal combination of k_1 and k_2 depends on the influence of the certain k_i on the number of false positives and false negatives. For example, the control limit of curve parameter 1 may have more influence on the number of misclassifications compared to the control limit of curve parameter 2. This translates to the optimal combination of control limits; allowing curve parameter 2 to have a wider control limit while controlling curve parameter 1 more strictly can result in a lower objective value. The contour plot presented in Figure 4.10 illustrates this trade-off. The ranges for k_1 and k_2 have been determined using Equation 4.40. The colors in the contour plot resemble the objective value. The optimal combination of control limits is the combination of k_1 and k_2 that results in the lowest objective value, the area with the dark blue color. The optimal combination of control limits is reached with k_1 is 2.22 and k_2 is 5.44. The resulting probability for having a false negative and false positive are 0.1426 and 0.1064, respectively. The number of false negative and false positive misclassifications in the data set provided by MSD are evaluated and show that misclassifications occur for batches contaminated at the end of the production process.

Furthermore, the mathematical properties for the obtained model outcome and parameter configuration are evaluated. The boundary conditions formulated in Section 4.3.3 are based on the preliminary model therefore these boundaries cannot be used for the two parameter model. Another method for evaluating

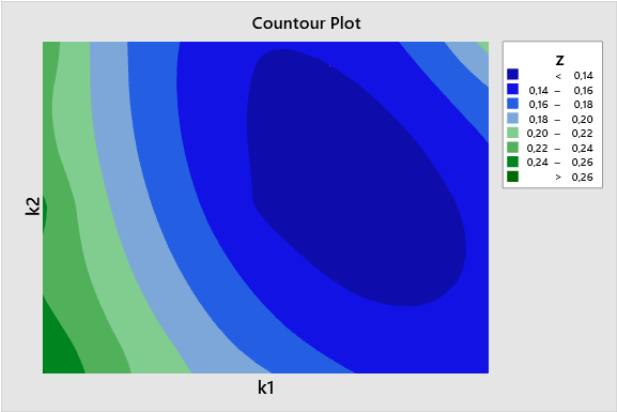


Figure 4.10: Contour plot for k_1 and k_2 for the two parameter model assuming equal weights for false negatives and false positives

these properties is to check numerically by plotting the probability distribution function and derivatives for false positive and false negative. From this numerical analysis can be concluded that all properties (probability property, decreasing or increasing in k and convexity) discussed in Section 4.3.3 hold for the given parameter configuration.

$$x_i - k_i \geq a_i \qquad for\ i\ \in\ \{1,2\} \tag{4.40}$$
$$x_i + k_i \leq b_i \qquad for\ i\ \in\ \{1,2\}$$

Sensitivity Analysis and Benchmarking

In the sensitivity analysis, the effect of three factors in the model are evaluated: (1) the ratio between the weight of false negative misclassification and the weight of false positive misclassification, (2) the regression coefficient of curve parameter 1 and (3) the regression coefficient of curve parameter 2. The outcome of the model for different scenarios considered in the sensitivity analysis is compared to a benchmark in order to quantify the improvement gained from using the optimization model proposed in this study compared to using the benchmark model. Condition-based maintenance is mentioned as relevant literature stream for this research and is consulted for finding a suitable benchmark model. The study of Phillips, Cripps, Lau, & Hodkiewicz (2015) proposes a method for defining the control limits for condition-based maintenance policy using on the binary logistic regression cut-off value. The cut-off value is adapted based on the weights for specificity and sensitivity, in other words the weight on false negative and false positive misclassifications. This model determines the control limits for multiple process parameters separately whereas the two parameter model jointly optimizes the control limits of multiple parameters.

The results of the sensitivity analysis are presented in Tables 4.3, 4.4 and 4.5. First, the results of the sensitivity analysis on the weight of false negative compared to that of false positive are discussed. Table 4.3 shows that both $k*_1$ and $k*_2$ decrease as the ratio $\frac{C_{FN}}{C_{FP}}$ increases. As the consequences of having a false negative increase, the requirements for a batch to be classified as pure becomes stricter by making the control limits smaller. As expected, the probability of having a false positive increases as the ratio increases whereas the probability of having a false negative decreases. The objective value decreases as

the ratio becomes more extreme, either a very low or very high ratio. The two parameter optimization model is compared to the benchmark from maintenance where the improvement in performance by using the optimization model is expressed by improvement potential (IP); IP is calculated by using Equation 4.41. The improvement potential decreases in the ratio $\frac{C_{FN}}{C_{FP}}$ meaning the gain resulting from using the optimization model becomes smaller if more weight is given to false negatives in comparison to false positives. In the bio-pharmaceutical industry, often more weight is given to false negative misclassifications compared to false positive misclassifications because false negatives have higher consequences in regards to costs and quality risk. However, in this weight lost profit, resulting from a false positive misclassification, is often not considered because lost profit is hard to quantify. In addition, the optimization model provides a more balanced solution compared to the benchmark model meaning there is a better trade-off between p_{FN} and p_{FP}. For example with ratio 3 the optimization model has a p_{FN} of 0.0303 and a p_{FP} of 0.2672 whereas the benchmark model has a p_{FN} of 0.0107 and a p_{FP} of 0.3745. The p_{FN} of the benchmark model is lower compared to that of the optimization model however opposite is true for the p_{FP}. Taking into account lost profit, it is more beneficial to have a lower p_{FP} while still obtaining a low objective value and p_{FN}. This principle shows that the improvement potential of the optimization model may be even more than quantified by the IP measure in the tables.

$$IP = \frac{Obj_{benchmark\ model} - Obj_{optimization\ model}}{Obj_{benchmark\ model}} * 100\% \tag{4.41}$$

Continuing to the analysis of coefficient of curve parameter 1, increasing the coefficient represents a higher influence of the parameter value on the probability of a batch being pure. Table 4.4 shows that increasing β_1 causes both $k*_1$ and $k*_2$ to decrease. Moreover, p_{FP} increases in β_1 whereas p_{FN} and IP decrease in β_1. Curve parameter 1 should be more strictly controlled to prevent false negative misclassifications when the influence of curve parameter 1 on the purity of a batch increases. A lower upper limit is required which is accomplished by decreasing $k*_1$. In the optimization model, $k*_2$ decreases as well which can be explained by the mechanism between $k*_1$ and $k*_2$, and the probability of having a misclassification. Only decreasing $k*_1$ would require a higher decrease in k for the same impact on misclassifications in comparison to decreasing both $k*_1$ and $k*_2$. In other words, the decrease in k is spread out over $k*_1$ and $k*_2$ which generates a more balanced control strategy in terms of trade-off between p_{FP} and p_{FN} and requires a less extreme decrease in $k*_1$. The increase in p_{FN} and decrease in p_{FP} when β_1 decreases is a result of the broader control limits for both parameters. In comparison to the benchmark, improvement gained by using the optimization model decreases as β_1 increases. Evaluating the p_{FN} and p_{FP} of the benchmark, it can be observed that a higher β_1 results in a more balanced solution, meaning a better trade-off between p_{FP} and p_{FN}. One of the advantages of using the optimization model is a more balanced solution hence the power of the optimization model decreases when the solution of the benchmark becomes more balanced.

Similarly, the strength of the relation between curve parameter 2 and the probability of a batch being pure, β_2, is varied and the impact of the optimal solution is analyzed. The results of this analysis are presented in Table 4.5. The $k*_1$, $k*_2$ and p_{FP} increase and p_{FN} decreases in β_2. An increase in β_2 represents a

	Optimization Model						**Benchmark Model**				
$\frac{C_{FN}}{C_{FP}}$	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$	IP	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$
0.25	0.0088	0.3802	0.0831	3.11	9.43	45.53%	0.1650	0.1026	0.1525	1.64	7.99
0.5	0.0678	0.2135	0.1164	2.65	7.35	30.21%	0.2195	0.0612	0.1667	1.38	6.90
1 (Base)	0.1426	0.1064	0.1245	2.22	5.44	20.19%	0.2785	0.0335	0.1560	1.12	5.80
1.5	0.1894	0.0679	0.1165	2.00	4.42	16.27%	0.3138	0.0227	0.1391	0.96	5.18
2	0.2223	0.0488	0.1066	1.85	3.75	14.19%	0.3394	0.0167	0.1243	0.85	4.71
2.5	0.2471	0.0377	0.0975	1.74	3.27	12.82%	0.3588	0.0131	0.1119	0.77	4.35
3	0.2672	0.0303	0.0895	1.66	2.88	11.93%	0.3745	0.0107	0.1017	0.70	4.07

Table 4.3: Sensitivity analysis for the influence of the ratio between the weight for false negatives and the weight for false positives on the two parameter model

	Optimization Model						**Benchmark Model**				
Compared to $\beta_{1,base}$	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$	IP	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$
-0.3	0.1098	0.1366	0.1232	2.96	5.77	30.30%	0.3214	0.0321	0.1768	1.40	5.80
-0.2	0.1271	0.1253	0.1262	2.68	5.66	28.50%	0.3203	0.0327	0.1765	1.30	5.80
-0.1	0.1374	0.1154	0.1264	2.44	5.55	22.76%	0.2942	0.0331	0.1637	1.20	5.80
0 (Base)	0.1426	0.1064	0.1245	2.22	5.44	20.19%	0.2785	0.0335	0.1560	1.12	5.80
+0.1	0.1438	0.0986	0.1212	2.03	5.33	18.11%	0.2623	0.0337	0.1480	1.03	5.80
+0.2	0.1422	0.0917	0.1170	1.86	5.21	16.40%	0.2459	0.0339	0.1399	0.96	5.80
+0.3	0.1387	0.0857	0.1122	1.71	5.09	14.97%	0.2299	0.0340	0.1320	0.89	5.80

Table 4.4: Sensitivity analysis for the influence of the strength of the relation between curve parameter 1 and the probability of a batch being pure (β_1) on the two parameter model

stronger relation between curve parameter 2 and the probability of a batch being pure. As this is a positive relationship meaning a higher parameter value for curve parameter 2 means a higher probability a batch is pure, increasing β_2 keeping $k*_1$ and $k*_2$ unchanged results in a lower number of false negatives and a higher number of false positives. For this reason, $k*_1$ and $k*_2$ are increased to reverse this effect of β_2. In the previous paragraph, it is explained why both $k*_1$ and $k*_2$ are increased; the increase in k spread out over both variables results in a more balanced strategy. The increase in p_{FN} and decrease in p_{FP} when β_2 increases is results from the broader control limits for both parameters. The comparison of the optimization model to the benchmark model shows that IP increases in β_2. Similar to the observation for β_1, this is most likely cause by the more balanced solution; the difference between p_{FP} and p_{FN} is less extreme when β_2 becomes smaller.

Comparison to Traditional Purity Test

Similar to the preliminary model, the model performance is compared to the performance of the traditional testing method to evaluate if the requirement for RTRT stating "RTRT is equivalent or better in providing

	Optimization Model						**Benchmark Model**				
Compared to $\beta_{2,base}$	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$	IP	p_{FP}	p_{FN}	Obj	$k*_1$	$k*_2$
-0.09	0.1394	0.0716	0.1055	1.17	3.39	0.14%	0.1444	0.0669	0.1057	1.12	3.38
-0.06	0.1462	0.0840	0.1151	1.52	4.22	3.84%	0.1862	0.0532	0.1197	1.12	4.27
-0.03	0.1479	0.0956	0.1218	1.87	4.88	11.23%	0.2322	0.0421	0.1372	1.12	5.08
0 (Base)	0.1426	0.1064	0.1245	2.22	5.44	20.19%	0.2785	0.0335	0.1560	1.12	5.80
+0.03	0.1285	0.1167	0.1226	2.57	5.90	29.64%	0.3217	0.0268	0.1743	1.12	6.46
+0.06	0.1043	0.1272	0.1158	2.92	6.28	39.29%	0.3595	0.0218	0.1907	1.12	7.07
+0.09	0.0714	0.1369	0.1042	3.25	6.67	49.13%	0.3916	0.0179	0.2048	1.12	7.62

Table 4.5: Sensitivity analysis for the influence of the strength of the relation between curve parameter 2 and the probability of a batch being pure (β_2) on the two parameter model

quality assurance compared to product end testing" is met. From the analysis it was found that the two parameter models also shows difficulty with detecting batches contaminated at a late stage of the process. This can be supported as the two parameter model uses parameters both modelling a characteristic of the lag phase and exponential growth phase. Most batches contaminated at a late stage have already reached the stationary stage therefore do not influence the values of the parameters included in the model. In conclusion, the traditional testing method is superior to the two parameter model because the purity test is able to detect batches with a high degree of contamination contaminated near the end of the process whereas the two parameter model cannot detect these batches.

4.5 Conclusion and Discussion

In the bio-pharmaceutical a lot of attention is given to ensuring the safety and efficacy of the products. Safety and efficacy is a subject that refers to various quality characteristics of a product either focused on the product's ability to produce the desired effect (protection against diseases or curing a disease) or the likelihood and severity of adverse effects. The quality attribute purity is a measure of safety that determines whether undesired bacteria are present in the product. Currently, purity is measured by conducting a test in which an agar plate is incubated and later inspected on the presence of undesired bacteria. This is a dichotomous test meaning the test has two possible outcomes: pure (1) or contaminated (0). However, this test knows some disadvantage; it takes time, often a day, before the test result is known, testing is costly due to personnel costs and costs for the specialized lab and the test only takes a small sample which may not in all cases be representative for the whole batch. This study describes a novel control strategy known as RTRT that classifies the purity of a batch based on process parameters and eliminates the need of purity testing.

The RTRT strategy for a dichotomous quality attribute formulates control limits for the process parameters critical to that attribute; for purity, if the value of the process parameters are within their defined control limit the batch is classified as pure, otherwise the batch is classified as contaminated. However, defining a RTRT strategy involves challenges for instance identifying the process parameters critical to purity.

Additionally, the misclassification of batches can result in lost profit due to a pure batch being misclassified as contaminated, manufacturing cost for batches that are being disposed as waste after being processed downstream and quality risks resulting from further processing a contaminated batch. The trade-off between the risk of having a false positive classification, a pure batch classified as contaminated, and the risk of having a false negative classification, a contaminated batch classified as pure, should be optimized when defining the control limits. The wider the control limits the higher the probability of having a false negative, the smaller the control limits the higher the probability of having a false positive.

To address these challenges, the process of identifying the process parameter critical to purity using a combination of expert knowledge and data analysis is presented. Moreover, two models have been developed; an one-parameter optimization model that classifies purity based on one process parameter and a joint optimization model that classifies purity based on two process parameters, and boundary conditions on variables for the relevant mathematical properties are formulated. The model jointly optimizing the control limits of multiple process parameters links to the concept of the Swiss cheese model; preventing misclassification by using various checks based on different process characteristics. Both models were validated with an industry case study at MSD Animal Health. Results proved that contamination early in the process can be detected accurately using a RTRT approach. Comparing our model to existing models and methods, it is shown that our optimization models could lead to significant improvements in terms of misclassification costs especially with the joint optimization model that classifies purity based on multiple process parameters. Using additional process parameters would potentially increase the improvement gained by the use of this model but more importantly, according to the philosophy of the Swiss Cheese model, it would provide a higher level of quality assurance. From the analysis is was concluded that process parameters with a strong positive relation to purity lead to higher improvements. The opposite is true for process parameters with a negative relation to purity; the improvement percentage decreases in the strength of the relation. This indicates for introducing additional process parameters that have a strong relation to purity. In addition, it was found that the improvements resulting from the model may be even higher than quantified in this case study when considering lost profit as misclassification cost for false positives. The comparison to the benchmark shows that the joint optimization model provides a better balance between false positive and false negative misclassifications resulting in less lost profit. Moreover, the analysis found that the improvement increases as the weight on false positive compared to false negative misclassifications increases. However, in industry often false negatives have the higher consequences in terms of costs and quality risks in comparison to false positives. The comparison of the RTRT strategy to the traditional purity test led us to conclude that the purity test is superior to the RTRT strategy although the difference in performance is limited to the detection of contaminations late in the process. When implementing RTRT, an option would be to include a risk, such as not detecting a certain type of contamination, in the contingency plan.

This study is one of the first to illustrate how operations management methods can be used to formulate a RTRT strategy and bring benefits to the quality control of bio-pharmaceutical production processes. In the future, RTRT can eliminate cost for testing, shorten the production leadtime and increase quality

assurance.

Future work could expand the joint optimization model to n process parameters. Increasing the number of process parameters for classifying purity would increase the quality assurance of the RTRT strategy as more characteristics of the process are evaluated. Additionally, it would be interesting to optimize the center line of the control limit. This was not relevant for the case study of this research due to the one-sided control limit but could potentially lead to further improvements for the formulation of two-sided control limits. Moreover, testing various contaminating bacteria can provide more insight in detecting overall purity based on different type of contaminations. Purity measures the absence of all possible contaminating bacteria and for this research only one particular bacteria is considered. Other contaminating bacteria can have different growing characteristics which can influence the detectability with a RTRT approach. For example, a slow growing contaminating bacteria can be harder to detect. This would especially be valuable information when applying for RTRT implementation and replacing the traditional purity test. Lastly, future work can focus on expanding this joint optimization model to other fields such as condition-based maintenance. In condition-based maintenance conditions such as vibration are measured to determine when predictive maintenance has to be conducted. The joint optimization of the thresholds for these conditions can result in a more cost efficient maintenance policy.

Chapter 5

Real Time Release Approach for Yield

In this chapter, the RTRT strategy for yield is discussed. Yield is defined as the biomass at the end of the production process. This biomass is formed due to the presence of biological processes that cause bacteria to multiply. For the purpose of RTRT, the goal is to predict yield by using characteristics of the fermentation process measured through process parameters. The process parameters that are indicators for the amount of yield are called critical process parameters and will be used to formulate a prediction model. First, the current yield test and the general RTRT approach for numeric quality attributes is described in Section 5.1. Than, a literature review is conducted in Section 5.2 where both the field of bacteriology and the field of operations management are discussed. Subsequently, the mathematical model for prediction yield is presented. Chapter 5.4 summarizes the case study conducted at MSD evaluating the use of the prediction model in a production environment. At last, in Chapter 5.5, the conclusion are drawn and the managerial insights are presented.

5.1 Introduction

Yield is a quality attribute that evaluates the efficacy of a product by measuring the biomass at the end of the production process. The biomass amount is an measure of process output and indicative for the performance of the process. The amount of yield produced for a certain batch can in some cases even influence the production planning; if yield is very low, an additional batch needs to be produced in order to ensure enough supply for the customer, if yield is very high, a batch can be removed from the planning to prevent keeping excess stock.

Yield can be tested using various testing methods conducted either after upstream processing or at the end of the process after downstream process. Examples of yield testing methods are the wet cell weight test or the ELISA test. For testing methods executed after downstream processing the test is conducted at the quality control lab. The tests conducted after the main fermentation is completed (upstream processing) are most often executed by the production department itself in a specialized lab environment.

Yield testing has some disadvantages regarding cost and leadtime. The cost allocated to the maintenance, cleaning and allocated space of the labs and the personnel cost for trained employees conducting the test can be a substantial portion of the manufacturing cost of a product. Additionally, the tests conducted at the quality control lab have significant leadtime before the yield results are available. As a result, production departments are not able to react on the yield results of a previous batch before starting the production of a new batch. For example, yield is low due to a problem in the raw materials, when the yield results are known before starting the production of a new batch the cause of this low yield can be determined and suitable actions can be initiated that prevent the new batch from having a low yield as well. In addition, the leadtime for results prevents quick response to batches with high or low yields that require changes in the production schedule.

With the introduction of RTRT, the cost for specialized lab and personnel are eliminated and the leadtime for results becomes zero which allows for quicker response to the outcome of the test. The scope for this research is limited to a yield test executed directly after the main fermentation process by the production department itself influencing mainly the cost for specialized lab and personnel. This decision is based on the limited amount of data collected during downstream processing. When more data is collected downstream, the insights from this research can be expanded to RTRT for yield tests conducted after downstream processing.

The RTRT strategy should be formulated based on the understanding of the relationships between process parameters, material attributes and the quality attribute yield. This understanding is used to build a prediction model for yield using process parameters measured during the fermentation process critical to yield and modelling the relation between these process parameters and yield. These process parameters can be seen as an indicative or indirect measure for the amount of biomass at the end of the main fermentation process. The performance of the prediction model is tested for a bacterial production process at MSD Animal Health where the outcome of the quality test for yield is compared to the outcome of the prediction model. This case study provides verification of the relations between the process parameters and yield using production data at commercial scale and demonstrates whether or not the RTRT is equivalent or superior to the traditional testing method for yield.

5.2 Literature Review

Two streams of literature are relevant to this research: the stream of life sciences and the stream of operations management. Section 5.2.1 explains the studies discussed in the field of life sciences relevant to this study. Subsequently, Section 5.2.2 focuses on the studies relevant in the field of operations management.

5.2.1 Relevant Literature in Life Sciences

Yield, or more generally biomass, is an important measure of process performance in the bio-pharmaceutical industry. Various efforts have been made to develop off-line and in-line measurements tools for biomass or parameters that are indirect measures for the amount of biomass and to increase the process yield.

Most of the existing work focuses on the indirect measurement both in-line and off-line, and the prediction

of biomass in the bio-pharmaceutical industry. Using predictive or analytical models based on process parameters measured in-line or measured off-line using sampling has been done for a long time. The study of Hall et al. (1996) proposed a partial least-squares regression model that can estimate the biomass of Escherichia Coli fermentation using the results of a near-infrared spectroscopy conducted on a sample of the broth. Similarly, S.Sivakesava, Irudayaraj, & Ali (2001) explain the use of infrared spectroscopic techniques for rapid off-line measurement for estimating optical density which can serve as indirect measure for biomass. Nowadays, online sensors are developed that can measure various characteristics of a process allowing for indirect estimation of biomass without taking a sample. For instance, Jenzsch, Simutis, Eisbrenner, Stückrath, & Lübbert (2006) examine estimation techniques based on the relationship between the parameter to be monitored, for instance biomass, and variables that are measured with online sensors during the process. The article compares three methods, (1) multivariate linear regression, (2) principal component analysis and (3) artificial neural networks for modelling of these relationships using data from cultivation experiments. In the same line of work, the work of Kiviharju, Salonen, Moilanen, & Eerikäinen (2008) discusses the use of industrially applicable probes for the measurement of optical density, infrared spectroscopy and fluorescence as a way of permitting real-time estimation of the microbial biomass. Nevertheless, this research differs from the existing work as it does not assume linear relations between yield and predictive variables. Additionally, in contrast to to the models in literature, the model is theory driven instead of data driven allowing for generalisation of the model for various bacteriological processes.

Another relevant stream in literature aims to improve the performance of fermentation processes by increasing yield or reducing the variation in yield. The study of Koca, Martagan, Adan, Maillart, & van Ravenstein (2020) discusses the application of the bleed-feed technique, a production strategy that allows for eliminating intermediate set-ups of the bioreactor, to increase yield. Another example of an efficiency initiative to improve yield is presented by Martagan et al. (2020). This research defined the critical process parameters for yield and conducted a design of experiments to find the optimal parameter configuration resulting in a substantial increase in yield. In addition, Papapetridis et al. (2018) explore engineering strategies with as goal to improve ethanol yield; these strategies focused on improving growth rate and fermentation reactions by reducing of eliminating by-product.

Although real-time release testing is a novel strategy, some studies discuss the application of this approach to yield testing. For example, the study of Wechselberger, Sagmeister, & Herwig (2013) presents a general method for quantifying biomass and growth rate real-time in fed-batch culture that is based on the element balancing approach. In addition, Yardley, Kell, Barrett, & Davey (2000) discusses the use of passive electrical properties of biological materials that are produced at a constant rate during growth in order to assess the amount of biomass. This research differs from the aforementioned studies; the model developed in this research is based on the biological understanding of the relations between process parameter and yield using a separate stochastic process for each growth phase whereas the model in literature are based on complex data-driven relations that cannot be generalised to other products.

In summary, methods for predicting biomass using indicative process parameters measured either in-line or

by sampling are extensively discussed. Also, research on improvement initiatives for increasing biomass and efficiency are widely studied. However, the application of RTRT for biomass has barely been mentioned in available literature and is hard to generalize for a wider application. This study will propose a general theory-driven model for the prediction of biomass that can be applied to various bacteriological processes and is based on the intuitive biological behaviour of biomass creation.

5.2.2 Relevant Literature in Operations Management

Studies have shown the substantial benefits in terms of cost reduction and output improvement from using operations management methodologies in various fields. The bio-pharmaceutical industry is starting to acknowledge the gains that can result from using these methodologies. Operations management approaches applied in this industry vary from initiatives to improve output and decrease process variation to the development of predictive models for output estimation.

Many existing work focuses on improvement initiatives to decrease yield variation or increase yield/ process output. For example, the research of Shakoor, Rahman, Rayta, & Chakrabarty (2017) and Ji, Sun, Yang, & Wan (2007) use a static set of data including information on previous cultivation years as input for a supervised machine learning analysis such as artificial neural networks, that can predict the most suitable timing for cultivation processes that maximize the yield. The work of Shah, Dubey, Hemnani, Gala, & Kalbande (2018) also focuses on the optimization of crop yield by using a multivariate prediction model for yield to determine for which temperature and humidity the crop output is optimal. In the context of pharmaceuticals, Y. Yang & Tjia (2010) propose various process improvements based on changing temperature settings, optimizing vacuum systems and increasing equilibrium that resulted in substantial yield increase without capital investments. Another approach of improving yield is to investigate process variations and mitigate or reduce these variation. This approach is used by Nieuwoudt, Ragheb, Nejati, & Massoud (2007); based on the insight of a sensitivity analysis various techniques have been proposed with the purpose of decreasing the impact of process variations on the low noise amplifiers resulting in an increase in yield.

A body of work related to the stream of yield improvement discusses the use of control limits to monitor yield over time and detect process deterioration. Control limits can for example be used to observe changes in the process variance that can potentially lead to problems in yield. T. Chang & Gan (1995) describe the application of cumulative sum control charts, both one-sided and two-sided, for monitoring changes in process variance based on the logistic transformation of the sample variance. In addition, control charts can be used to control the process and quantify process improvements. For example, the study of Gupta, Jain, Meena, & Dangayach (2018) proposes the DMAIC (define-measure-analyze-improve-control) method to improve a tire manufacturing process and explains the use of process control charts to control the process, observe process changes in terms of output. They also quantify the performance capability improvement after improvement initiatives have been implemented.

Another stream of relevance to this research focuses on the prediction of process output quantity. For instance, the study of Aghighi, Azadbakht, Ashourloo, Shahrabi, & Radiom (2018) employs machine learning techniques to estimate the yield of crops by using time series data of remotely sensed data. The

comparison of the state-of-the art machine learning techniques such as Gaussian process regression and random forest regression to conventional regression techniques led to conclude that the machine learning techniques outperform when predicting yield based on time series data. Similarly, Shah et al. (2018) estimate the yield of crops based on both environment variables such as the temperature and humidity and process variables using a combination of multivariate polynomial regression and support vector machine regression and random forest models. Another example of this stream of research is the study of Goyal (2014) where the goal was to predict sediment yield generated within a watershed. The method proposed uses a combination of a decision tree model together with wavelet regression model for the estimation of yield based on measured parameter. Although, numerous studies discuss the prediction of yield, the model proposed in this research differs from these studies based on the generalisability and ease of interpretation. The existing models are most often purely based on relations build using advanced machine learning techniques as a result these models are hard to interpret and also unsuitable for broader application. The model presented in this study is build using the biological relations between yield and parameters used for predicting yield and substantiated with data therefore allowing for both easy interpretation and broader applicability.

Summarizing, this study has contributions to both the life sciences and operations management communities. In this study, a model is developed that builds upon our understanding of the biological relations between yield and process parameters and quantifies these relations using extensive data analysis. Combining a theory-driven with a data-driven modelling approach meets the need for a model that is more generalized and can be used for various bacteriological processes. Additionally, numerical analysis based on industry data is used to validate the model and illustrate its use in a commercial manufacturing environment.

5.3 Mathematical Model

For the application of RTRT for yield, a mathematical model is developed that predicts yield, the biomass at the end of the main fermentation process, based on process parameters measured during the fermentation process. In the requirements and guidelines on RTRT provided by EMA (2012), it is mentioned that a firm should be able to demonstrate its understanding of the production process and of the relationships between the process parameters, material attributes and quality attributes present in the process. In addition, a firm should be able to justify and verify these relationships by sound scientific data. Therefore, the aim is to build a mathematical model that is based on intuitive biological relations between yield and process parameters this way demonstrating enhanced product and process understanding. The justification and verification of this model using scientific data is presented in Section 5.4 where a numerical case study is conducted on industry data.

Moreover, RTRT is a relatively novel strategy that shows great potential in improving productivity by faster product release and decreasing costs however hardly any companies have succeeded in implementing this strategy. For this study to serve a bigger role in the introduction of RTRT within the bio-pharmaceutical industry it is of essence that the model is generic and can be applied for various processes. Formulating a generic model for yield prediction contributes towards more products being released using a RTRT control

strategy.

Various methods have been considered for predicting yield keeping in mind the desire for interpretability, nowadays also referred to as explanable artifical intellegence, and generalisation. First, the well-known machine learning technique named (artificial) neural networks is considered. This method offers numerous advantages such as its ability to model complex non-linear relations between dependent and independent variables and detect interactions between independent variables. However, this methods requires much more data in comparison to other machine learning techniques and has a black box nature. In the biopharmaceutical industry the availability of data is often scarce making it difficult to use neural networks. Moreover, the black box nature of this method is conflicting with the desire to demonstrate enhanced process and product understanding. The next method to be considered is the principal component analysis (PCA); this method reduces the dimensionality of a data set by creating new uncorrelated variables. The methods reduces overfitting and removes correlated features. On the other hand, the independent uncorrelated variables can be difficult to interpret and some information may be lost compared to the original set of variables. The disadvantage of independent variables becoming less interpretable is not in line with the aim for the mathematical model constructed in this study. Also, data analysis has shown that the number of process parameters measured during the fermentation process is too limited to create insightful principal component variables. Furthermore, regression analysis is taken into consideration. This method is well-known for its simplicity and ease of implementation. Though, some researchers comment on the lack of practicality of this method. This especially applies to the linear regression analysis where the relationship between the independent and dependent variables must be linear. Another disadvantage of this method is its assumption that the relationship between two variables remains unchanged. Data analysis showed that the nature of the biological relationships between yield and the measured process parameters is difficult to model with a regression model as relationships tend to change over time. In conclusion, neural networks, principal component analysis and regression analysis are not suitable as modelling approach for predicting yield. Herefore, its chosen to use an existing model for biomass in the context of bacterial growth and modify this model to fit the model characteristics required for the prediction of yield for RTRT. The main advantage of this approach is that the model will be build upon proven relations between biological parameters and biomass that are applicable to almost all bacteriological processes and are easy to interpret.

The process of bacterial growth is a subject discussed often in literature and many models on the growth of the bacterial population have been defined. The growth of a bacteria culture is often modelled by looking at the population size or the biomass. According to Juška, Gedminienė, & Ivanec (2007), Kostov, Popova, Gochev, & Koprinkova-Hristova (2012) and Panesar, Kennedy, Knill, & Kosseva (2007), the rate of growth of the bacteria population or biomass is proportional to the size of the population (or biomass) at that specific moment of time. These researchers have also shown that the relative growth rate, which is defined as the rate of growth proportional to population size, is constant for bacterial growth processes. These characteristics of bacterial growth allow us to model the biomass (or population size) of the bacterial culture using the differential equation presented in Equation 5.1; biomass is modelled with

m(t) and the relative growth rate is represented by μ. This differential equation can be rewritten into the mathematical expression illustrated in Equation 8.2. The derivation of this mathematical expression explained step by step can be found in Appendix 8.

$$\frac{dm(t)}{dt} = \mu * m(t) \tag{5.1}$$

$$m(t) = m(0) * e^{\mu * t}, \qquad t \geq 0 \tag{5.2}$$

The model expressed in Equation 8.2 is used as basis to formulate the yield prediction model for the RTRT approach. This basic model is adapted to fit the model characteristics required for this yield model. First, the constant growth rate assumption is relaxed; in this basic model is assumed that the growth rate (μ) is constant however this is only partially correct. The growth rate differs between different bacterial growth phases but within a phase it can be assumed that the growth rate is constant. This adaption requires for the bacterial growth phases to be modelled separately with for each phase a specific μ and b. As shown in 5.1, there are four bacterial growth phases: the lag phase, the exponential growth phase, the stationary phase and the death phase. The acceleration and deceleration phase are incorporated in the lag and exponential growth phase. The μ_l and μ_e for the lag and exponential growth phase should be positive since biomass is increasing during these phases. In the stationary phase, biomass remains constant resulting in a μ_s of zero therefore this phase is not considered in the model. The μ_d of the death phase is negative as during this phase the biomass decreases.

Modelling the four phases separately requires for b to be determined for each phase. For all phases, except for the lag phase, b_i is equal to the biomass at the end of the previous phase. b_l, the starting amount for the lag phase, is dependent on the amount of biomass injected from the inoculum into the fermentation process. This amount is often estimated by a process parameter that is an indirect measure of the amount of biomass at the start of the fermentation process.

Furthermore, the basic function for m(t) models the amount of biomass as a function of t that way modelling the process of biomass growth and decline over the whole duration of the fermentation process. For the prediction of yield, only the biomass at the end of the fermentation process is of interest. Therefore, instead of predicting the biomass for a specific t, the function can be independent of t changing m(t) to m where m represents the biomass at the end of the fermentation process. As a result of this change, the time dependency is included by multiplying the growth rate of certain phase with the duration of that phase. Some bacteria types do not go through all phases, in that case the duration of the non-occurring phase(s) set to zero. For example, the fermentation process of a bacteria is ended when reaching the stationary phase, in this case the duration of the death phase is set to zero.

The growth rate and duration of a specific phase ((μ_i and b_i with i $\in$ {l,e,d}) can derived from a process parameter that is a direct or indirect measure of biomass (process parameter critical to yield) and resembles the bacterial growth curve. These parameters are measured real-time on-line during the whole duration of the fermentation process and are presented in a curve graph. Examples of such parameters are optical

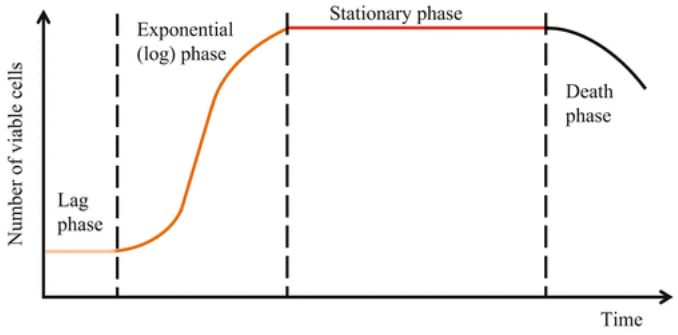


Figure 5.1: Schematic View of a Typical Microbial Growth

density and oxygen pressure (Kiviharju et al. (2008); Hall et al. (1996)). The growth curve models expressed in Chapter 3 can be used to determine the duration of each phase. The maximum growth rate as determined by the growth curve model is not used because it only provides information on the maximum growth rate in the exponential growth rate. The growth rate of each specific growth phase is determined from the curve of the process parameter by taking the average relative growth rate for that specific growth phase.

In addition, the parameter constants a_i and c_i with i $\in$ {l,e,d} are introduced. These constants translate the unit of measurement of the process parameters, used to estimate μ and b_l, to the unit of measurement of yield. For instance, oxygen pressure is used to estimate μ_i with i $\in$ {l,e,d} and the duration of the inoculum process is used to estimate b_l with unit of measurement mg/L and hours, respectively. The process parameters have to be multiplied with a certain factor to express the parameter value in the unit of measurement of yield, number of cells per L. The parameter constants a_i and c_i are determined using data analysis by finding the value for a_i and c_i for which the prediction error for the data set is minimized. This value is fixed but can be updated every certain period as part of model maintenance activities.

The resulting yield model is presented in Equation 5.3 and determines the yield of a particular batch. The growth rate and duration of each phase and the starting amount for the lag phase (b_l) are batch dependent; the parameter constants a_i and c_i with i $\in$ l,e,d are determined by data analysis and are independent of individual batch data.

$$Yield_h = b_d * a_d * e^{c_d * \mu_d * T_d} \tag{5.3}$$

with

$$b_e = b_l * a_l * e^{c_l * \mu_l * T_l}$$

$$b_d = b_e * a_e * e^{c_e * \mu_e * T_e}$$

5.4 Numerical Analysis

The aim of this numerical analysis is to validate the mathematical model described in Section 5.3 and to evaluate the use of this model in practice. This numerical analysis is based on an industry case study at MSD. MSD has provided a data set including historical production data with corresponding yield measurement of batches produced in the last two years for two products. The production processes of the

two products are very similar however the data availability for these products differ significantly. The validation of the model is based on the product with the highest data availability whereafter the obtained results and insights are verified for the product with the lower data availability using the same method. It is chosen to use a time frame of two years as no major process changes have occurred during this period. The historical production data includes all process parameters measured during the inoculum processes and main fermentation process. The process parameters of downstream processes are not included because the scope of this research is limited to yield testing conducted directly after the main fermentation process as is the case for the products included in this analysis.

The RTRT approach for yield as presented in this numerical analysis is developed in close collaboration with MSD through regular meetings with stakeholders. Weekly meetings with the lead of the production department and monthly meetings with operators provided the necessary biological knowledge required for this research. Additionally, during these meetings the intermediate model design and results were discussed and the feedback was used to improve the model and provide required validation of the model. Moreover, monthly meetings with higher management took place to share progress and to discuss managerial insights. This provided valuable information on the business perspective for this study and on the managerial insights that would be interesting for MSD to obtain. Other stakeholders on this project were responsible for quality and regulatory affairs. Through regular meetings with these stakeholders perspectives on quality requirements for RTRT and required dossier changes were shared which helped in the development of the model and the interpretation and reflection of the results.

This project on RTRT for yield started in April 2020 with regular visits and go-see's at the production line with experienced operators to develop an understanding of the production process. The yield model was built during June 2020-September 2020 with various iterations based on feedback meetings with MSD's production and quality department. From September 2020 to February 2021 the model was validated by intensive data analysis and parallel testing on commercial scale for one of the products in scope. Parallel testing means that the process data was used to predict yield based on the prediction model developed however the current yield testing method remained leading for the release of the batch. This project has been shared across the company receiving positive feedback and interest from various people.

Firstly, Section 5.4.1 discusses the construction of the yield prediction model using industry data. Subsequently, Section 5.4.2 explains the validation of this model and verifies the use of this model using the product with lower data availability. In Section 5.4.3 the performance of the prediction model is compared to the performance of the current testing method.

5.4.1 Model Construction

The mathematical model requires the following input parameters: (1) growth rate per growth phase, (2) duration of each growth phase and (3) the starting amount of biomass. Process parameters measured real-time on-line during the fermentation process that are an indirect or, more preferably, a direct measure of biomass can be used to estimate the growth rate and duration for each growth phase. The starting amount of biomass can be estimated by either a process parameter measured during the main fermentation process or during the inoculum processes. This parameter does not have to be measured real-time on-line;

it can also be an off-line measured process parameter measured just once.

Starting with the prediction of growth rate and duration, the number of parameters to be considered for this estimation is limited to those measured on-line real-time. The set of potential process parameters includes both direct and indirect measures of biomass. The preference goes to process parameters that are a direct measurement of biomass because this captures the process behaviour with regards to biomass most accurately. For example, the article of Arnold, Cortada, Gledndinning, & Henderson (n.d.) suggests the use of turbidity for the modelling of biomass because it is a direct measure of biomass and is able to show both increase and decrease in biomass. An example of an indirect measure is lye consumption; bacterial growth causes the pH to drop, to keep pH between certain levels, lye is added. An increase in total lye usage suggests bacterial growth however other reactions within the process could potentially influence total lye usage as well. In the set of potential process parameters for the process at MSD, only one process parameter directly measures biomass. This process parameter, from now on referred to as process parameter critical to yield, will be used to derive the growth rate and the duration of the growth phases as input for the yield model. An additional advantage of this parameter is that, in contrast to the other potential process parameters in the set, the decline in biomass in the death phase can be observed. The other process parameters are either cumulative numbers unable to show decrease or are only affected in the early stages of the process hence the death phase cannot be observed in these curves.

To determine the growth rate and duration of each phase using the curve from the process parameter critical to yield, first the start and end of each growth phase have to be determined. In Chapter 3, growth curve modelling is proposed which is a method that can be used to determine the timing of the different growth curve phase. The process parameter critical to yield is an uncontrolled parameter for which the lag phase, the exponential growth phase and death phase; the stationary phase is negligible and will therefore not be considered. Due to the presence of the death phase, the process parameter can be best modelled with Equation 3.3. The numerical analysis in Chapter 3 already shows the modelling of this same variable; referred to as process parameter 2. The parameters derived from this Equation relevant for determining the start and end of the growth phases are: lambda and t_{death} (see Figure 5.2). In addition, the variable T representing the duration of the main fermentation is input for the model. The start and end time of the lag phase, exponential growth phase and death phase are determined as follows:

- Lag phase:

  $t_{l,start}$=0; the lag phase is the first occurring phase and therefore starts at time is zero.

  $t_{l,end}$=λ; λ models the moment in time at which the process transitions from the lag phase into the exponential growth phase representing the moment at which the lag phase ends.

- Exponential growth phase:

  $t_{e,start}$=λ

  $t_{e,end}$=t_{death}; t_{death} models the moment in time at which the process transitions from the exponential growth phase into the death phase representing the moment at which the exponential growth phase ends.

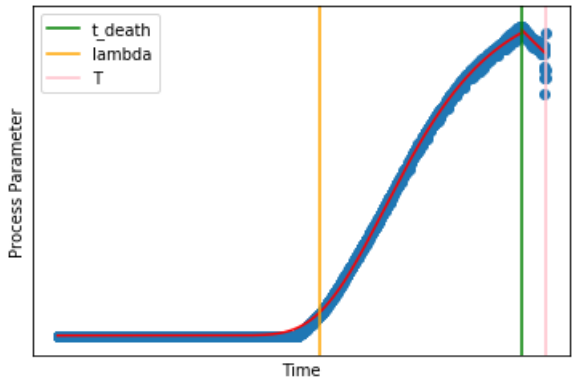


Figure 5.2: The curve of the process parameters with the relevant time points for determining the duration of the growth phases

- Death phase:

  The start and end of the death phase is modelled with an if-statement as not all process runs include a death phase. T represents the duration of the fermentation process and therefore, if a death phase is present, the end of the death phase.

$$t_{d,start} = \begin{cases} t_{death}, & \text{if } t_{death} < T \\ 0, & \text{if } otherwise \end{cases}$$

$$t_{d,end} = \begin{cases} T, & \text{if } t_{death} < T \\ 0, & \text{if } otherwise \end{cases}$$

The growth rate per phase is determined by calculating the average growth rate over the duration of that particular phase considering each data point included in that phase (see Equation 5.4). The duration is calculated by $t_{i,end} - t_{i,start}$ for i in {l,e,d};

$$\mu_i = \frac{\sum_{t=t_{start,i}}^{t_{end,i}} \frac{\frac{dp(t)}{dt}}{p(t)}}{t_{end,i} - t_{start,i}} \qquad , i \in \{l, e, d\} \tag{5.4}$$

where p(t) represents the value for process parameter critical to yield at time t

Moreover, b_l has to be estimated based on a process parameter measured during either the main fermentation process or during the inoculum process. The set of potential process parameter is, in contrast to the estimation of growth rate and duration, not limited to real-time on-line measured process parameters. The potential process parameter are selected based expert knowledge and literature, for instance, Arnold et al. (n.d.) explains that the optical density of the inoculum can be indicative for the starting biomass amount. Extensive data analysis has been conducted where the prediction error is evaluated for different b_l estimators. The process parameter resulting in the most robust prediction with the smallest prediction error is chosen as estimator for b_l.

At last, the values for a_i and c_i with i ∈ {l,e,d} is determined by finding the optimal values for a_i and c_i for which the prediction error is minimized. The resulting model is illustrated in Equation 5.5 (disclaimer: the model coefficients are changed for confidentiality purposes).

$$Yield = b_e * (1.723e + 01) * e^{(7.767e-14)*\mu_d*T_d} \tag{5.5}$$

with

$$b_l = (7.500e - 02) * e^{(2.333e-01)*\mu_l*T_l}$$

$$b_e = b_e * (1.683e + 01) * e^{(3.763e-15)*\mu_e*T_e}$$

The yield is predicted using the model as presented in Equation 5.5. From these results, it was observed that the variance in mu_l was significantly higher compared to the variation in mu_e and mu_d. This is most likely caused by the small amount of data points in the lag phase that have a non-zero relative growth rate therefore the growth rate of one data point has significant effect on mu_l resulting in a high variance. The variation in b_l has a negative effect on the models ability to accurately predict yield. For this reason, it is chosen to change mu_l into a constant value determined using data analysis; the constant is equal to mu_l that minimizes the prediction error. The adapted model is presented in Equation 5.6

$$Yield = b_e * (1.723e + 01) * e^{(7.767e-14)*\mu_d*T_d} \tag{5.6}$$

with

$$b_l = (7.500e - 02) * e^{(2.333e-01)*(2.330e-01)*T_l}$$

$$b_e = b_e * (1.683e + 01) * e^{(3.763e-15)*\mu_e*T_e}$$

5.4.2 Model Validation

The yield model is validated by investigating the following model performance indicators: (1) average prediction error, (2) standard deviation of prediction error, (3) the minimum and maximum prediction error, (4) root mean square error prediction (RMSEP) and (5) distribution of the prediction error. Due to confidentially, prediction errors are expressed as the percentage deviation from the measured yield: $\frac{|yield_{measured} - yield_{predicted}|}{yield_{measured}}$ *100 %. Table 5.1 summarizes the results of the performance indicators of the model for the product (with high data availability) for which the model is constructed in the previous section. In addition, the results for a similar product that has a low data availability and for the current yield test are presented in this table.

First, the results for the product with high data availability for which the model is presented in Section 5.4.1 are discussed. The average prediction error is 4.71 % meaning, on average, the absolute difference between the predicted and measured yield divided by the measured yield is equal to 0.0471. The distribution of this prediction error is not normally distributed whereas the prediction error expressed as $yield_{measured} - yield_{predicted}$ is normally distributed with a mean equal to zero. The latter shows that the prediction is unbaised. The minimum and maximum prediction error ($\frac{|yield_{measured} - yield_{predicted}|}{yield_{measured}}$ *100 %) observed is 0.0043 % and 26.7205 %, respectively. This indicates that the range of prediction

	Prediction		**Measurement**
	High Data Availability	Low Data Availability	Current Yield Test
Average Error	4.71%	6.26%	3.50%
Standard Deviation Error	6.40%	3.49%	2.77%
Minimum Error	0.00%	1.06%	0.18%
Maximum Error	26.72%	13.15%	10.42%
RMSE	0.64	1.01	0.75

Table 5.1: Summary of performance comparison between the yield prediction model and current testing method for a product with high data availability and low data availability

error values is wide. Confidence intervals are constructed to gain more insight in the spread of the prediction error. The confidence intervals are determined based on the prediction error expressed as $yield_measured - yield_predicted$ as only this form of the prediction error is normally distributed. The 95% confidence interval is (-1.51, 0.78) (disclaimer: this interval is multiplied with a certain factor for confidentiality purposes).

For verification, the prediction model is constructed for another product which has, in comparison to the previous product, low data availability. The same method for building the model is used and also for this model it is assumed that μ_l is constant. The average prediction error for this product is 1.55 percentage point higher compared to the product with high data availability. In contrast, the standard deviation in the prediction error is significantly less, however, this could also be the result of the difference in data availability. Moreover, in comparison to the product with high availability, the maximum prediction error of the product with low data availability is substantially less although this could again be a result of the significant difference in the number of included batches. The distribution of the prediction error is evaluated for both $\frac{|yield_{measured} - yield_{predicted}|}{yield_{measured}}$ *100 % and $yield_{measured} - yield_{predicted}$ and showed that neither are normally distributed hence no valuable comparison can be done based on a 95% confidence interval.

In addition, RMSEP, root mean square error prediction, is calculated which incorporates the difference in data availability. Equation 5.7 presents the formula for calculating RMSEP where n represents the number of batches. As can be seen from the formula, the number of batches is included in this performance indicator hereby taking into account differences in data availability. The RMSEP for the product with high data availability is equal to 0.64 whereas the RMSEP for the product with low data availability is equal to 1.01.

$$RMSEP = \sqrt{(\frac{\sum_{i=1}^{n}(yield_{measured_i} - yield_{predicted_i})^2}{n-1}}) \tag{5.7}$$

5.4.3 Comparison to Current Yield Test

A requirement for RTRT is "the RTRT control strategy should be equivalent or superior compared to the existing method". In order to know whether or not this requirement is met, the accuracy of the current testing method should be determined. This is done by conducting a measurement system analysis (MSA). Five samples have been tested by 3 operators with 2 replicates per sample per operator. The rule-of-thumb for a MSA is 10 samples tested by 3 operators with 2 replicates however this was not possible due to the limited availability of operators. The analysis of the yield test has been done in the form of a Gage RR study using the ANOVA method. From this analysis could be concluded that 12.9% of the variability in yield results come from variability in the measurement system.

From this measurement system analysis, the average measurement error for the current yield test could be determined and compared to the average prediction error of both products. This measurement error is expressed in a similar form as that of the prediction error: $\frac{|yield_{measured_i} - yield_{average_i}|}{yield_{measured_i}}$ *100 % where $yield_{measured_i}$ represent the yield measurement of a particular sample from batch i and $yield_average_i$ represents the average yield measurement from batch i. The average measurement error is equal to 3.50 % which is lower compared to average prediction error for both products. The distribution of the measurement error is not normally distributed for both $\frac{|yield_{measured_i} - yield_{average_i}|}{yield_{measured_i}}$ *100 % and $yield_{measured_i} - yield_{average_i}$. The maximum measurement error is lower in comparison to the maximum of the prediction errors although this difference is not substantial when comparing to the product with low data availability. The standard deviation of the measurement error is 2.77 which is less than that of the prediction errors of the model. Additionally, RMSEP is evaluated for the measurement error in order to compare to the prediction model taking into account the difference in data availability. The RMSEP of the measurement error is 0.75 which is higher compared to the product with high data availability and lower compared to the product with low data availability. This shows the impact of considering data availability when evaluating the performance of a prediction or measurement.

Also, a test is conducted were the end concentration that is influenced by the measured or predicted yield is analyzed. The end concentration when using the measured yield is comparable to that of the predicted yield were in some cases the prediction yield is even superior in terms of end concentration compared to the measured yield.

5.5 Conclusion and Discussion

Various quality attributes are measured for a bio-pharmaceutical product to ensure safety and efficacy. Yield is a quality attribute that measures the output of the production process and can be determined either after upstream processing or after downstream processing depending on the product. This work focuses on yield measured directly after the completion of the main fermentation process. To date, yield is determined by conducting a quality test however these tests are costly as specialized labs and trained employees are required for executing these tests and it takes takes time before the results are available. An alternative approach for testing yield is RTRT. The RTRT strategy for yield predicts yield based on process parameters measured during the production process. Some challenges can arise when formulation

this RTRT approach, for instance the process parameters critical for predicting yield have to be identified. In addition, understanding the relation between process parameters and yield can be difficult as it is likely that this relation is non-linear and can change over time. Moreover, the prediction model for yield should be interpretable and generalisable for other bacteria cultures. This will facilitate RTRT approval and can enable wider application of the RTRT approach for yield.

This study addresses these challenges by providing a prediction model and explaining the model input in the form of process parameters critical to yield prediction. The mathematical model is based on existing models that aim to model biomass with respect to time. These existing models are expanded and adapted to predict yield. The resulting prediction model for yield models the biomass formation of each growth phase separately using the biological relation between growth rate, duration and biomass. The starting amount of biomass and the growth rate and duration of each phase are derived from process parameters critical to yield. Building the yield prediction model upon biological relations and separating the modelling of growth phases facilitates the interpretability and generalisability of the model. The prediction model for yield has been validated using an industry case study at MSD with as goal evaluating the applicability and accuracy of the model. In addition, this case study contributes to the validation required for RTRT application. In Chapter 2 it was mentioned "the validation of the RTRT method should be supported by substantial comparative data at commercial scale to demonstrate the relation between the RTRT methods and end product testing method". The results of this case study are obtained from batches produced at commercial scale and provide the first results for the comparative data analysis of the two methods. The results of the case study shows that the model was able to predict the yield accurately using a process parameter measured real-time that is a direct measure of biomass. The prediction error, the difference between the measured and the predicted yield, was in most cases found acceptable based on the insights obtained from the measurement system analysis of the traditional yield test. In addition, it was found that the amount of data available for constructing the model significantly influences the performance of the model. Low data availability for constructing the model results in a higher average prediction error and a higher root of mean squared error corrected for the number of batches but results in a lower standard deviation of the prediction error. Moreover, the comparison to the current yield test was found to be inconclusive. For most performance measures traditional yield testing method is superior to RTRT prediction models although in some cases the difference between RTRT and traditional testing is very small. In contrast, the measure that corrects for data availability indicates that RTRT with high data availability provides a higher accuracy compared to the traditional test. This illustrates the importance of a measurement analysis with a substantial number of batches for a reliable assessment of measurement accuracy.

In this study, the insight from operations management and biology are combined to build a model that is generalisable and interpretable. The model developed in this work could be the first step towards wide introduction of RTRT for yield. RTRT allows for the elimination of quality testing hereby decreasing both product manufacturing cost and product leadtime. As a result, this allows for a competitive position in the bio-manufacturing product market and for offering an affordable high quality product to customers.

Future work could expand the yield model to tests that measure yield after downstream processes. A substantial amount of yield tests are conducted after downstream processing therefore incorporating the influence of the downstream process on yield would be of high value. However, this may require the introduction of new process parameter measurements in the downstream production process. Moreover, building a model that could provide performance guarantees for the model and could give insight in the number of batches required for the validation of the model would be interesting. For example, the product with low data availability needs more batches for the validation of the prediction model however the amount of extra validation batches is unknown.

Chapter 6

Decision-support and Process Monitoring Tool

In this chapter, a dashboard is presented that functions both as a process monitoring tool and as a decision-support tool. RTRT opens many opportunities for process monitoring such as trending analysis. This dashboard is build to process and analyze quality attribute data, obtained from either a RTRT strategy or quality testing, so process issues can be detected early on and preventive actions can be undertaken to minimize further impact on supply. In addition, the dashboard can function as decision-support tool for identifying product candidates for RTRT application and other improvement initiatives. Insight is gained on which products consistently deliver the desired quality and are therefore suitable for RTRT and which products are instable in terms of quality and could be a good candidate for improvement initiatives. First, Section 6.1 describes the goals of the dashboard with relation to RTRT and the wider applicability of the dashboard. Subsequently, the approach used to design the dashboard, collect the data and formulate the rules for the trending analysis is discussed in Section 6.2. In Section 6.3 the implementation of the dashboard at MSD Animal Health is described and corresponding implementation results are presented. Lastly, the conclusion and managerial insights are presented.

6.1 Introduction

This study focuses on various aspects of RTRT; what are the requirements and guidelines for this strategy, which methods are used, how can the critical process parameters be identified and how can models be used to predict quality. These questions are related to either finding a method for defining a RTRT strategy or formulating a model that can predict quality using the RTRT strategy. However, looking at RTRT strategy in a broader perspective considering the future of RTRT, relevant questions are "which products are suitable for RTRT?" and "how do we gain from RTRT implementation".

EMA (2012) emphasizes that before considering RTRT, the production process should be in control and should consistently deliver the specified product quality. Hence, to determine which products are suitable for control, one should analyze which production processes consistently deliver the desired product quality

and can therefore considered to be in control. In addition, many opportunities in regards to process monitoring arise when RTRT is implemented. Currently, quality is tested using quality tests that have a significant leadtime before the measurement results come in. RTRT enables for quality results to be available in real-time, meaning directly after the process has been completed or is some cases even during the process. As a result, the leadtime for receiving the quality results becomes zero. This opens an opportunity for real-time process monitoring; the results from a batch can show the performance of the process and can detect certain process issues. This information can be used to prevent quality problems for future batches. For example, if the yield result of a batch is low it can be an indication of a problem with the raw materials, an action plan can be followed that identifies the problem and solves the problem before producing the next batch. In addition, eliminating the leadtime for measurement results can improve the efficiency of planning; batches can be added or removed based on the real-time quality results. Using the example of yield, if results show that the yield of a batch is low, an extra batch can be added to the planning schedule which prevents supply problems in the future. If the yield of a batch is high, a batch can be removed from planning assuring no excess stock.

To identify RTRT candidate products and seize opportunities in the field of process monitoring, a dashboard was developed that functions both as a decision support tool and as a process monitoring tool. The product candidates for RTRT are products that consistently deliver the desired quality; quality consistency can be identified from the dashboard by conducting various analyses. Additionally, the process is monitored (1) by identifying results that are outside specification limits and may require planning changes and (2) by checking for trends that can be indicative for potential process issues.

RTRT is not yet broadly implemented in the bio-pharmaceutical industry due to strict regulations and the novelty of this control strategy. Nevertheless, using this dashboard opens opportunities for process monitoring and process improvement also with the current testing strategy. In the case of MSD, the dashboard would significantly reduce the information processing leadtime and would provide insights in trending and observations outside the specification limits. The current process flow for releasing information within MSD results in significant information leadtime since information is not directly shared when results come available. This negatively impacts production performance because production representatives cannot react earlier to quality observations. Moreover, trend analysis or relating analyses on quality results are rarely applied.

This dashboard processes the measurement results as soon as they come available and automatically conducts further analysis on the quality results. As a result, required changes in planning can be processed sooner hereby improving the planning efficiency and departments can react (faster) to trends which can prevent potential problems for future batches. Furthermore, this dashboard can be used to identify which products would be suitable for a RTRT strategy as discussed in previous paragraphs.

The scope of this study is limited to the design, construction and implementation of dashboard used for processing yield data at MSD Animal Health. In the future, the dashboard could be expanded to other numeric quality attributes.

6.2 Approach

The approach includes three phases: (1) the design phase, (2) the data collection phase and (2) the phase where trending rules are defined. These phases are discussed in Sections 6.2.1 to 6.2.3. The approach is first discussed in a general form whereafter the specific approach for MSD's yield dashboard is presented.

6.2.1 Dashboard Design

The design phase is the first step in building the dashboard. In this phase, it is defined who are the stakeholders and what decisions they need to make, what insights they want to obtain and which metrics would support those decisions and insights. In addition, the decision on which application will be used to build the dashboard is made in this phase. After defining the needs and requirements for the dashboard the first step is prototyping for example, sketching the design with pen and paper. Prototyping is an iterative process; make a prototype, get feedback from stakeholders, and iterate.

In general, the stakeholders of this dashboard are operators, process analysts and planners. The purpose of this dashboard is two-fold: identifying RTRT product candidates and process monitoring using quality data. The performance of a product's production process determines whether or not that product is an interesting candidate for RTRT implementation. For RTRT a process should be able to consistently deliver the desired quality which can be evaluated by checking for outside-specification-limit observations and trends. An outside-specification-limit observation represents irregular variance in the process and the inability to deliver the desired quality (Ziegel, 1992). The presence of trends illustrates irregular variance and patterns in the production process meaning the production process is not in control (Cadinoska et al., 2019). Gaining insight in the occurrence of trends and of outside-specification-limit (out-of-spec) observations can show whether or not a product would be suitable for RTRT.

The monitoring of the production process with respect to the quality attribute of interest can include various evaluations, for example outliers or increase over time. Process monitoring is interesting from a business perspective because quality results can have influence on supply which is of interest for planning and monitoring the process can help to identify process issues which is of interest for production. Starting with planning, batches that have a quality result that deviates significantly from the desired quality can impact the supply and may therefore require changes in the planning schedule. The definition of a significant deviation can depend on the quality attribute but may also vary between departments. Therefore, when designing the dashboard, it should be clear what is defined as a significant deviation that could potentially require planning changes. In production, the main interest is trending because a trend can be an indication of issues with the process or with the input materials. Trending analysis studies the presence of underlying trends or irregular variation in process or quality data. A specific action plan can be performed when a certain trend is observed that can potentially identify the cause of the underlying pattern or irregular variation. Addressing this cause can prevent problems in the future or, when trends show a desired process change, leverage this insight to improve the performance of the process.

The application in which the dashboard is build depends on the frequency of use, the users and the required capabilities. For users it is important that the application is easy to use and does not require

to much instructions. Therefore, in most cases the preference goes to an application where users are experienced with, for example Excel. Applications such as PowerBI and Aera have more capabilities in terms of data analytics and are faster in processing large amounts of data however most users are not familiar with these programs.

In summary, the dashboard is designed for operators, process analysts and planners and should include trending analysis and provide an overview of outside-specification-limit observations. The application in which the dashboard is build should be user friendly and have the required capabilities.

At MSD, the design phase took place in close collaboration with the stakeholders and future users. From a plannings perspective the desire is to have an overview of the batches that have a yield that falls outside the specification limits. In addition, the following information of an out-of-spec batch besides the batch number is of interest for reviewing the planning schedule: (1) material number, (2) date of manufacturing, (3) magnitude and direction of deviation and (4) status of release. Planners are especially interested in the latest out-of-spec observations hence the overview should be sorted by date starting with the batches with the most recent date of manufacturing. From a production perspective it is of interest to see an overview of process deviations per product (material number). A deviation is referred to as the presence of an underlying pattern or the observation of a result that is significantly lower or higher compared to the expected result. Trending analysis can identify these deviations using statistical rules that evaluate a result by taking into account the mean and standard deviation and by comparing to previous observations. Both proven process deviations and pre-signals are relevant for process monitoring therefore the traffic light principle is introduced where the colors red, orange and green have the following meaning:

- Red:
  A proven process deviation is detected, for example 6 observations in a row, all increasing or all decreasing.
- Orange:
  A pre-signal of a process deviation is detected meaning a process deviation as described for Red is partly observed. For instance, 4 observations in a row, all increasing or all decreasing. This signal can be an indication of a possible process deviation in the future when no correcting actions are performed.
- Green:
  No process deviations or pre-signals for process deviations are observed. Green classifications are not included in the dashboard as these observations do not require any attention or actions.

The information to be mentioned by a red or orange observation is (1) material number, (2) product description, (2) date and (4) trending type. Trending analysis is done on product level instead of batch level; most trending rules evaluate the process of a product by analyzing multiple batches. There exist some exceptions where a trending rule evaluates the result of just one batch. The observation of a trend that evaluates the result of multiple batches is triggered by the batch that completes the trend observations; the last batch in the streak of batches relevant to the trend that changes the trend statement from being

false to being *true*. The date included in the overview is equal to the date of manufacturing of the last batch. Moreover, as trending is done on product level there can be multiple trends observed for a specific product. Similar to planning, the recent observations are most relevant therefore it is chosen to only include the most recent observed trend for a product when multiple trends are included.

The dashboard is also used for identifying product candidates for the RTRT of the quality attribute yield. The insights relevant for planning and production are also of interest when deciding on product candidates for RTRT. An overview is provided that shows which products have out-of-spec observations or process deviations (trends). If a product appears often in this overview this means that the process cannot consistently deliver desired quality quality or the process is not in control. Hence, these products are not suitable for RTRT. Products that are barely observed in this overview are possible candidates for RTRT of yield. This can also be used to identify product candidates interesting for improvement initiatives. Products that have inconsistent quality delivery would benefit from process improvement. Although, the demand and profit margin for that product should also be considered in this decision.

Aside from the overview on out-of-spec observations and trends, the stakeholders of MSD are interested in a graphical analysis that allows for observing a certain trend or out-of-spec observation in connection to other observations of this product.

Based on conversations with the users of the dashboard it was decided to use Excel as application. Alternative applications that have been considered however most users indicated that they are not familiar with these applications and they feel more comfortable with a familiar application. The capabilities of Excel allowed for the dashboard to be build in this application.

6.2.2 Data Collection

The dashboard can provide insights in process stability, process deviations and out-of-spec observations by analyzing process and quality data. Process deviations are observed by using rules that detect trends based on the comparison with previous observations and reference standards. This section explains the data requirements and the process of data collection for the dashboard. First, the data collection process is discussed from a general perspective. Than, the data collection for the yield dashboard of MSD is discussed in more detail.

Production and Quality Data

In the design phase it is discussed that the dashboard includes both out-of-spec analysis and trending analysis. For these analyses data has to be collected that serves as input for these analysis. In most companies an Enterprise Resource Planning (ERP) system is used to store and couple data of various subsystems and can therefore play an essential role in the data collection process. In general, the analysis in the dashboard requires both production data and quality data though this is very company specific. For example, the results of the quality attribute analyzed and the batch and material numbers should be provided as input.

At MSD, data from various data sources are combined and stored in a system. This system also enables the construction of customized data sets that can be exported to Excel. This feature is used to generate a

Excel file that can serve as input for the dashboard. The next step is to determine the data that needs to be included in this input file. For the out-of-spec analysis, the following batch data and quality data is required:

- Batch number
- Material number
- Material description
- Date of manufacturing
- Status of release
- Inspection characteristic
- Yield result
- Upper and lower specification limit for yield (product specific)
- Department

The batch number, material number, date of manufacturing and status of release are visualized in the overview of the dashboard. Inspection characteristic describes which type of test is conducted or what quality attribute is measured. The yield result and the upper and lower specification limits are used to determine if an observation should be classified as out-of-spec; the observation should be shown in the overview if the result is outside these limits. For the yield dashboard only the inspection characteristic for the measurement of yield should be included therefore this variable is used for filtering. The variable 'department' is also included for filtering purposes. For the out-of-spec analysis planners are interested in the magnitude and direction of the deviation; these are determined in Excel.

The data required for the trend analysis, besides the reference standards (mean and standard deviation), is the following:

- Batch number
- Material number
- Material description
- Date of manufacturing
- Inspection characteristic
- Yield result
- Department
- Production line

The material number, material description and date of manufacturing are visualized in the trending analysis overview. The yield result is required as input for the trending analysis. The inspection characteristic

and department are included for filtering purposes. The production line is included for the trend analysis as the mean and standard deviation for yield can differ between production lines. Therefore, trending analysis is done separately for each production line.

Trending Standards

Trending analysis evaluates if process deviations are observed by comparing to previous observations and reference standards. References standards are product and production line specific and can be modelled with mean and standard deviation. These standards illustrate the expected observation for that product under normal process conditions. The mean and standard deviation can be determined from a period in which no process changes or major process issues have occurred. Process changes and major process issues can result in observations that deviate from normal process behaviour and can therefore produce an unrepresentative mean and standard deviation.

The reference standard for the yield dashboard are determined in one of two ways.

- Existing trending standards are used; the mean and standard deviation used for trending are known for production departments that already conduct trend analysis manually.
- A production period is determined in which no process changes or major issues occurred. The mean and standard deviation for yield is calculated for this period and used as reference standard.

6.2.3 Trending Rules

The last phase focuses on defining the rules for the trending analysis. A body of work on trending rules is available in the field of process quality and control. According to Cadinoska et al. (2019), two well-known sets of trending rules in the pharmaceutical industry are the WECO Western Electric rules and the Nelson rules. The WECO Western Electric rules are set up by a specialized manufacturing committee at the Western Electric Company and became a standard for the manufacturing field. The set includes four rules that focus on detecting irregular variation. The deviation of an observation from the center line expressed in standard deviations is evaluated; consistent abnormal variation or significantly high variation is considered as trend. The set of Nelson rules checks for both irregular variation and underlying patterns using eight trending rules. The rules either evaluate the deviation from the center line expressed in standard deviations or check for patterns between consecutive observations, for example an alternating pattern of observations. The Nelson rules provide a more complete analysis of process deviations because underlying patterns are considered therefore this set of rules is most suitable for the trending analysis of quality attributes.

For the trending analysis at MSD both proven trends and pre-signals of trends are evaluated. The set of rules for the proven trends includes all eight Nelson trending rules. The set of rules for the pre-signals of trends is based on the Nelson rules however includes only six rules as two Nelson rules are not suitable for early detection signaling. Table 6.1 presents the rules for proven trends and pre-signals for trends as used in the yield dashboard. Production departments may experience that some trends are not relevant for their specific production process. Therefore, a feature is included in the dashboard that allows to turn-off certain trending rules meaning these trending rules are not considered in the trending analysis.

	Proven Trend (Red)	Pre-Signal for Trend (Orange)
1	1 point $> 3\ \sigma$ from mean (outlier)	-
2	9 (or more) points in a row are on same side of mean	6 (or more) points in a row are on same side of mean
3	6 (or more) points in a row are continually increasing (or decreasing)	4 points in a row are continually increasing (or decreasing)
4	14 (or more) points in a row alternate in direction	9 points (or more) in a row alternate in direction
5	2 (or 3) out of 3 points in a row are $> 2\ \sigma$ from mean (same direction)	-
6	4 (or 5) out of 5 points in a row are $> 1\ \sigma$ from mean in same direction	3 (or more) out of 5 points in a row are $> 1\ \sigma$ from mean in same direction
7	15 points in a row are all within $1\ \sigma$ of mean (both sides)	10 points in a row are all within $1\ \sigma$ of mean (both sides)
8	8 points in a row exist with none within $1\ \sigma$ of mean (both sides)	5 points in a row exist with none within $1\ \sigma$ of mean (both sides)

Note: σ is defined as the standard deviation

Table 6.1: Overview of trending rules for yield dashboard

6.3 Implementation at MSD

This section elaborates on the implementation of the dashboard at MSD Animal Health. First, Section 6.3.1 discusses the timeline of this project from the kick-off to implementation. Than, Section 6.3.2 presents the improvements in relation to costs and decision making resulting from this implementation. At last, the feedback of other departments and sites are presented and the implementation challenges are explained.

6.3.1 Timeline

The yield dashboard has been designed, constructed, validated and implemented over the duration of this thesis in close collaboration with the departments at MSD. In the first few months, the main focus was on gaining insight in the RTRT strategy and the prediction of quality attributes yield and purity. The kick-off of the yield dashboard project, a sub project of the RTRT project, took place after some information was gathered on RTRT. After this kick-off, the first months were used to define user needs and translate these needs into the design of the dashboard. Defining the user needs and formalizing the design of the dashboard has been done in close collaboration with the stakeholders. The following months were allocated to data collection. Than, the yield dashboard was constructed with regular stakeholder meetings for user feedback. Prior to the implementation of the dashboard, a training session and follow-up meeting were held for the users of the dashboard. The dashboard was implemented at the collaborating departments and presented at a side-wide meeting with positive responses.

6.3.2 Implementation Results

The implemented yield dashboard is presented in Figures 6.1, 6.2 and 6.3. Figure 6.1 illustrates the graphical analysis that is included in the dashboard for providing more insight in trends and out-of-spec observations in relation to previous observations. In addition, a feature is included that can be used to select the desired time period for which the analyses are conducted. In Figure 6.2, the overview for planning of out-of-spec observations is presented. The percentage deviation is calculated by dividing the absolute difference between the observation and the limit that is violated (either upper or lower limit) with the limit that is violated and multiplying this with 100 %. Figure 6.3 shows the overview of the trending analysis for both proven trends, in the red column, and pre-signals of trends, in the orange column. Additionally, the dashboard includes description of the trend types and a feature of that can

turn a trend on or off where off means that the trend is not considered in the analysis.

Eventhough the dashboard was implemented just recently, production departments have used it to conduct trend analysis and gain insight. The use of this dashboard has led to significant information leadtime reduction and can lead to substantial cost reductions. By processing the results as soon as they are available instead of waiting till the batch is released an information leadtime reduction of 33 to 50% is realized. The cost reductions are hard to quantify since the dashboard is just implemented and are dependent on the process issues detected. The general improvements that result from using this dashboard are the following:

- *Reduction in inventory cost and lost sales.* The dashboard allows for earlier observation of batches that have a yield result that falls outside specification limits. This enables planners to react earlier and take appropriate measures. As a result, supply and demand are better aligned which reduces excessive inventory and supply shortages (Mohebbi, 2003). Adapting production planning based on supply changes due to out-of-spec observations is especially important for products, such as produced by MSD, that have limited shelf life. Having a lot of inventory can cause products to exceed their shelf life which would mean that these batches have to be disposed as waste. Out-of-specs can also lead to lost sales when it negatively impacts supply. The low ratio between safety stock expressed in time and the leadtime from raw materials to final product makes it difficult to react to changes in supply because the production process takes significantly longer than can be covered with the available safety stock. Detecting process issues more quickly or adapting the production planning can prevent problems in supply and reduce the amount of lost sales. Reducing inventory and lost sales can lead to significant cost reductions and higher profits. The impact of this improvement is determined using the number of out-of-spec observations and the estimated times in which a planning adjustment would be relevant. Approximately 5 % of the batches produced yearly require planning changes to limit supply impact (inventory or lost sales).
- *Timely analysis of process issues.* The trend analysis in the dashboard includes the detection of both proven trends and pre-signals for trends. A trend shows underlying patterns and/or irregular variation which indicate the occurrence of a process issue or provide a signal for a potential process issue in the nearby future (Cadinoska et al., 2019). The insight in these trends can help prevent process issues from happening by taking the right actions and preventive measures or can help detect and solve these issues in a timely matter to prevent supply impact. With this dashboard, the information leadtime is reduced with 33 to 50 % which allows for faster detection and solving of (potential) process issues as a result output performance can be increased and supply problems resulting in lost sales can prevented. The impact of this improvement is estimated by evaluating the number of process issues concerning yield. Approximately 15% of the batches produced per year experience issues concerning yield from which 20% have significant impact on supply. The dashboard enables for earlier problem solving as the issue is observed 33 to 50 % faster and in some cases the process issue can even be prevented because it is detected by the pre-signals for trends. This can significantly reduce or prevent supply impact for approximately numerous batches which

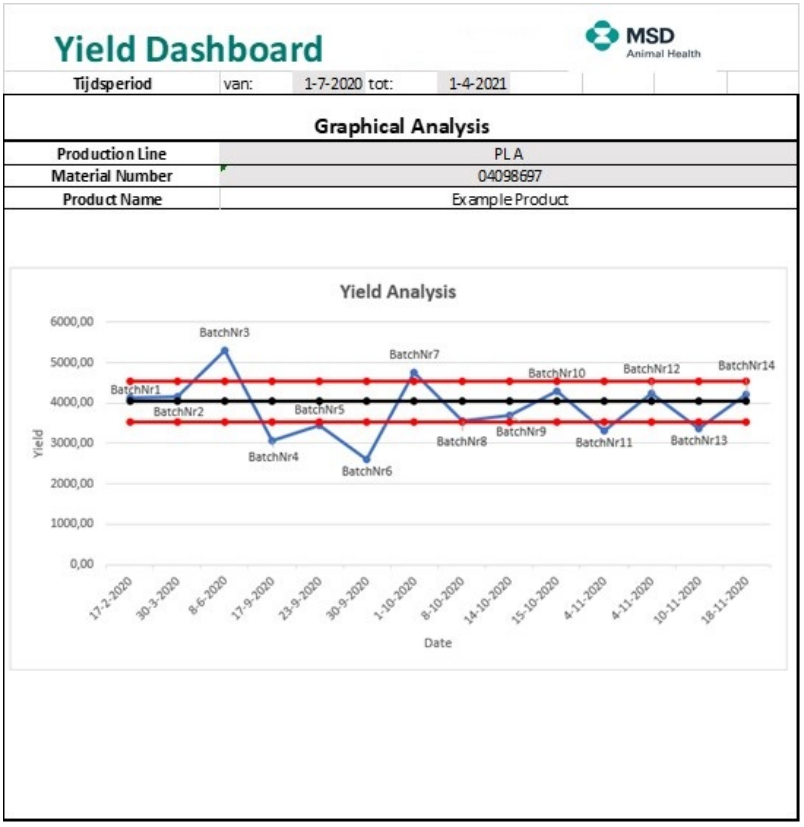


Figure 6.1: Graphical analysis and definition time period feature in the yield dashboard (Note: the dates are fictitious)

leads to substantial financial gains.

- *Data-based decision making for production selection for improvement initiatives and RTRT.* Prior to the implementation of the yield dashboard, most decision making was based on a combination of expert knowledge and limited data analysis. This made gaining insight in the instability of yield rather difficult. The dashboard allows for data-driven decision making for RTRT implementation or for other improvement initiatives. With the introduction of the yield dashboard, product with the most instable yield become visible which enables data supported decisions on which products are most suitable for RTRT application. The yield of these products can be predicted using a RTRT approach as presented in Chapter 5. Data supported decisions for improvement initiatives and RTRT introduction can improve return on investment and result in higher gains resulted from the investment (Brynjolfsson et al., 2011). Specifically for RTRT, it is important to show to authorities that the process is in control and desired quality is delivered consistently which can be proven using the data from this dashboard (EMA, 2012).

6.3.3 Feedback from MSD Community and Challenges

The yield dashboard is build for the departments located at the production site in Boxmeer. The goal, approach and the dashboard is presented to a broader audience at MSD Boxmeer and other production sites. The response has been enthusiastic and efforts are made to implement this dashboard at another production site.

Planning

Material Number	Batch Number	Date	↑/↓	Status
3698645	Example Product 1	29-3-2021	↓1.85%	Wait for release
1452611	Example Product 14	14-1-2021	↑2.44%	Wait for release
9887442	Example Product 10	4-11-2020	↓8.01%	Rejected
2265422	Example Product 8	9-10-2020	↑15.23%	Released
1432569	Example Product 3	28-9-2020	↑0.09%	Rejected
5998677	Example Product 7	1-9-2020	↑5.21%	Rejected
6332199	Example Product 17	23-7-2020	↓0.98%	Released

Figure 6.2: Overview for planning of batches with observations outside specification limits in the yield dashboard (Note: the dates are fictitious)

Productie

Material Number	Product Name	Date	Trending Type
4098697	Example Product 2	27-3-2021	4
5998677	Example Product 7	13-3-2021	1
3698645	Example Product 1	27-12-2020	8
9984720	Example Product 11	19-9-2020	3
1978851	Example Product 19	2-8-2020	6
2265422	Example Product 8	11-7-2020	3

(a) Overview for production of products that have proven trends

Productie

Material Number	Product Name	Date	Trending Type
3698645	Example Product 1	28-3-2021	5
1978851	Example Product 19	19-2-2021	4
4894897	Example Product 22	1-1-2021	6
2265422	Example Product 8	11-12-2020	1
9984720	Example Product 11	4-10-2020	1
5998677	Example Product 7	24-9-2020	4
1452611	Example Product 14	12-9-2020	6
1432569	Example Product 3	2-7-2020	4

(b) Overview for production of products that have pre-signals for trends

Figure 6.3: Illustration of trending analysis in the yield dashboard (Note: the dates are fictitious)

During the implementation of this dashboard it was recognized that finding a general applicable definition for yield could be challenging. In bio-pharmaceutical industry yield is defined in various ways, for example yield can be referred to as antigen content or kilograms vaccine. The definition of yield is often dependent on the production department and can also differ for a planning perspective compared to a production perspective. To facilitate different definitions of yield the Aera data input file is made department specific. When the correct yield measure is included in the Aera data input file no further changes have to be made to the dashboard. The input files for the production department only include information on the products produced by that department. It is assumed that within a department the same definition of yield is used; this assumption is verified by checking with various production department leads. The assumption is made that planning uses the same definition of yield for all production departments; in some cases, this definition differs from that of the production department. This assumption is supported by the stakeholders of this project. If in the future this assumption would be relaxed simple changes can be made to the dashboard in order to integrate the use of different yield definitions within one dashboard.

Another implementation challenge is related to collection of the correct production and quality data. For this dashboard, or for data analytics in general, standardized data formatting is required meaning a standard format or method for storing and naming variables should be in place. Standard formatting enables the grouping of variables based on a certain common characteristic and the comparison of variable values. During this project it became clear that not for all variables included in the input file a standard formatting method was in place. For instance, the variable production line is formatted as an open info field with no standard formatting method. One employee writes in the open info field for production line A 'PLA' whereas another employee names this same line as 'ProLiA'. The difference in naming causes difficulties when grouping and comparing variables in the dashboard. The solution for this challenges included both a corrective and a preventive action. The corrective action involved finding the most used data format and only include batches that used this data format. With this action, approximately 80 percent of the batches are included in the dashboard. For the preventive solution, a side-project is started that focuses on the standardization of data formatting. Using the example of the naming of production lines, the open info field is changed to a field that only allows you to select from a limited number of options where each option represents one of the production lines.

6.4 Conclusion and Discussion

RTRT is a novel strategy that caught the interest of various bio-pharmaceutical companies however many struggle with the implementation of this strategy. This study has focused on defining a method for formulating a RTRT control strategy and defining a RTRT method for purity and yield. With the insights gained from this research doors open to the application of RTRT for a larger product portfolio and companies come closer to implementing and using RTRT approach for the release of products. With the desire for wider application and the perspective of implementing this strategy, new questions arise "what products are suitable for RTRT and how can we make optimal use of the benefits arising from RTRT?". By EMA (2006a) it is stated that for RTRT to be considered the process should be able to consistently deliver the desired quality. However, determining which products meet this requirement remains a challenge due to complex data and large product portfolios. In addition, RTRT implementation enables faster data processing and real-time process monitoring resulting in early detection of process issues and out-of-spec observations. As a result, issues can be analyzed and solved in a timely matter minimizing the impact on supply. The challenge is to formulate a method for processing and analyzing the data and visualize this data in such a way that companies seize these opportunities arising from RTRT.

To address these challenges, a dashboard is developed that includes an analysis that detects observations that are outside specification limit indicating inconsistency in quality and potential supply impact. In addition, a trend analysis is included and expanded with the use of the traffic-light principle; orange represent pre-signals of trends and red illustrates a detected trend. Trends illustrate the presence of irregular variation and patterns which are indicative for process issues. This dashboard functions both as a decision-making tool and as a process monitoring tool. The decision-making tool can be used for building a product portfolio for RTRT implementation but is also suitable for identifying products that would

benefit most from improvement initiatives. The insights of both the trend analysis and out-of-spec analysis are input for the decision-making support tool. Products for which trend and out-of-spec observations are observed regularly are not suitable for RTRT but could be good candidates for improvement initiatives. The process monitoring tool is beneficial from both a planning and a production perspective. The input for this tool depends on the perspective of the user. For planning, out-of-spec observations are of interest because such observations may require adaptions to the planning schedules to ensure alignment between supply and demand. Planning focuses mostly on the trend analysis; process issues can be prevented or solved in time to avoid future batches to be affected. Action plans can be formulated per trend that describe the actions that have to be conduct to analyze the cause of the trend and identify possible solution directions.

This dashboard has been developed and implemented at MSD for the quality attribute yield obtained from end testing, but in the future this can be obtained from a RTRT model. Before the introduction of this dashboard, yield results were made available after the release of a batch. Now, yield results are processed and analyzed as soon as quality testing is completed resulting in a information leadtime reduction of 33 to 50%. Implementation insights indicate that this dashboard leads to cost reduction and improved decision making. The dashboard increases planning efficiency resulting in lower inventory cost and less lost sales. In addition, the process issues can be detected earlier due to the information leadtime reduction and the dashboard gives warnings for potential future trends. As a result, process issues can be solved quicker or even be prevented which limits supply impact and decreases lost sales (Mohebbi, 2003). Also, the decision-making process for improvement initiatives and candidate selection for RTRT is improved resulting in higher returns on investments. As this dashboard has been implemented just recently it is hard to quantify the benefits. Furthermore, this dashboard and the approach used for building the dashboard and collecting the data has been shared and other production sites have shown interest in implementing this dashboard.

Future research could focus on problem diagnostics for the trend analysis. Automatic problem diagnostics can provide information on potential causes for the trend based on the process parameters measured for the batch(es) for which this trend is observed. For example, a trend is observed for product A, a pop-up screen should appear with "trend x is observed for product A with probable causes c1 and c2". In addition, it would interesting to expand the dashboard to other numeric quality attributes.

Chapter 7

Conclusion and Managerial Insights

The goal of this research is to make the first steps towards the introduction of RTRT at MSD Animal Health. This includes formulating a general method that can be used for RTRT application, developing a RTRT approach for the numeric quality attribute yield and dichotomous quality attribute purity and addressing future steps for a wider introduction of RTRT and for seizing benefits from RTRT implementation.

MSD has recognized the potential of the RTRT control strategy but experiences difficulties with developing a RTRT approach and implementing this strategy mainly as a result of the limited information on requirements and guidelines on RTRT. Using the available documentation provided by authorities and the insights gained from case studies a 6-phase framework is developed which is inspired by the Quality by Design approach and explains the steps required for replacing the traditional quality testing method for a RTRT approach. This generic iterative framework supports authorities in understanding the applied methods and can facilitate in the approval process. This research applies this framework for the formulation of a RTRT approach for the quality attributes yield and purity; skipping the first two steps as the quality attributes are already defined.

The method towards RTRT for purity and yield starts with the risk assessment which identifies material attributes and process parameters that are critical for the quality attribute. Casestudies and experts emphasize on the importance of using process parameters measured real-time to model and analyze process behaviour. Growth curve modelling is a method for quantifying longitudinal process parameter data into curve parameters that can be used for further analysis and can serve as input for RTRT models. The existing literature on growth curve modelling is limited in its application; only process parameters that resemble the standard bacterial growth curve can be modelled. This research expands the work on growth curve modelling by developing two growth models that can be applied to indirectly controlled process parameters or process parameters that include a death phase. This enables a wider application of growth curve modelling allowing for more process parameters to be analyzed with this method and serve

	Current strategy	RTRT strategy	Future Benefits RTRT
Purity	Incubation on agar plate	Control limits for critical process parameters	Leadtime reduced by 2 days Information on purity during process Cost reduced by 10 %
Yield	Off-line yield test	Prediction based on critical process parameter	Leadtime reduced by 30 min Cost reduced by 10 %

Table 7.1: Summary of current strategy and RTRT strategy for purity and yield with an estimation of future benefits resulting from RTRT when eliminating the current strategy after approval from the EMA

as potential critical process parameter for RTRT models or other prediction models. The growth curve modelling is applied for the RTRT approaches for yield and purity to quantify potential critical process parameters.

For the research on the RTRT approach for yield and purity, Table 7.1 summarizes the differences between the current strategy and the suggested RTRT strategy with the corresponding estimated benefits. The estimated benefits are based on replacing the current testing method by the RTRT strategy however this requires an extensive approval process which is yet to start. This research provides the strategy for RTRT implementation though actual implementation of RTRT and realizing these promising benefits awaits EMA approval. The benefits are in the area of cost and leadtime; the costs of testing are eliminated by implementing RTRT because testing is replaced by a data-driven approach and the leadtime for obtaining the test results becomes zero as RTRT provides immediate results. The reduction in leadtime is therefore equal to the duration of the test currently used for quality end testing. Depending on which critical process parameters are found and the meaning of these process parameters, the RTRT approach for purity can even provide information on the purity of the batch during the production process in contrast to a purity test which can only provide information afterwards. Though the real benefit of this research is the insight gained on RTRT application and the RTRT models that can be used for future RTRT application of the quality attributes tested by quality control (QC) after downstream processing. The impact when replacing QC testing with RTRT is higher compared to the replacement of in-house tests such as yield and purity as the costs and leadtime for conducting these tests is significantly higher. Replacing QC testing with a RTRT approach is expected to reduce overall manufacturing cost with 30% and decrease product leadtime with 40 days equal to a 80 % leadtime reduction. However, these tests are more difficult to replace with a RTRT approach as the impact on quality assurance is higher compared to the tests only related to the upstream process. This research is the first step towards RTRT for QC testing and can help obtain even more benefits in the future.

The RTRT strategy for purity, a dichotomous quality attribute, is based on the formulation of acceptance and rejection criteria, also referred to as control limits, for the process parameters found critical to purity classification. Defining these control limits is a trade-off between the risk of having a false positive misclassification resulting in unnecessary waste and the risk of having a false negative misclassification resulting in a quality risk and unnecessary value adding activities. An optimization model is developed that captures this trade-off while considering the consequences of a certain type of misclassification, the

relation between the process parameters and purity, and taking into account the operational ranges of the process parameters. This model is validated by conducting an industry case study at MSD using production data and experimental data of one product for both one and two critical process parameters. The second process parameter used for the two parameter model appeared to be non-critical to purity for the product in scope therefore the data of another product, where this parameter was critical, is used for data manipulation. The analysis shows significant improvements compared to existing methods especially when the weight on false positives compared to false negatives increases. In particular, it was found that most benefits arise when multiple process parameters are used for classifying purity because, in contrast to existing models, this model allows for joint optimization of the control limits of the included process parameters. In addition, the principle of the Swiss Cheese model demonstrates the importance of using multiple process parameters for defining the purity of a batch to ensure product quality. Both address the need for additional process parameters that can be used for purity classification. This is supported by the need for data manipulation when expanding to a two parameter model because no second critical process parameter could be identified. Analysis found that process parameters with a strong positive relation with purity would benefit most from this joint optimization model and would therefore be most interesting as additional process parameter. Moreover, the comparison to the traditional purity test shows that quality testing is still superior to the RTRT method in detecting contaminations introduced late in the process. From a quality perspective not detecting a contamination late in the process could potentially be an acceptable risk for two reasons: (1) the traditional test is in some cases also not able to detect these contaminations and (2) the probability of occurrence is minimal and when occurring it is mostly a result of an observable issue that can be included in the contingency plan for RTRT.

The RTRT strategy for yield, a numeric quality attribute, is based on a prediction model that captures the relation between critical process parameters and yield where yield is defined as the biomass at the end of the main fermentation process. In this study, a prediction model is build that models the formation of or reduction in biomass per bacterial growth phase using the starting amount of biomass, the growth rate and the duration of the phase to estimate yield. An industry case study at MSD is used to validate the model and compare its performance to the existing testing method. The results show that the model was able to predict yield accurately with a prediction error that is acceptable compared to the existing measurement error. In addition, it is found that data availability has substantial effect on the performance of the prediction model. Higher data availability results in a lower average prediction error but more variation in the prediction error. The influence of data availability also became visible when comparing the RTRT strategy to the traditional testing method. The comparison between the traditional yield test and RTRT prediction model is inconclusive. The comparison on average value and variation in error finds that the yield test is superior to the RTRT model however a performance measure correcting for data quantity and a test on end concentration indicate the opposite. This shows a need for additional data for a more balanced comparison between RTRT and quality testing especially on the measurement system analysis of the yield test.

After formulating a method for RTRT application and defining RTRT approaches for yield and purity,

this research continues by focusing on a wider application of RTRT and seizing the opportunities that arise from RTRT application. A dashboard is designed, constructed and implemented for MSD that functions both as a decision-support tool and as a process monitoring tool by integrating analyses on observations outside specifications and on observed trends and pre-signals for trends. Although RTRT is not yet implemented at MSD, significant improvements can be realized with the introduction of this dashboard. Implementation insights indicate that with the use of the dashboard information leadtime is reduced by 33 to 50% . Moreover, data-driven decision making enabled by the dashboard can result in higher returns on investments for RTRT and other improvement initiatives (Brynjolfsson et al., 2011). Also, motivating the suitability of RTRT for a product using data can support the validation process for RTRT approval. In addition, the process monitoring tool enables the detection of out-of-spec observation and a timely analysis of process issues that can prevent or limit supply impact. According to Mohebbi (2003), this can result in a reduction in inventory cost and lost sales.

To date, RTRT implementation in the bio-pharmaceutical industry is limited to none. This work could be the first step towards RTRT and could open a discussion with authorities to define the next steps for RTRT approval. MSD and other bio-pharmaceutical companies can benefit from RTRT as it would lead to cost reduction, can increase quality assurance and opens opportunities for real-time process control with as main vision to improve animal welfare, ensure food safety and support the bond between people and animals.

This study is limited to the RTRT for one product produced on a particular production line however the suggested methods and developed models can be applied to other products and production lines as well. Another limitation that should be addressed is that the RTRT approaches only focus on the upstream process and do not consider the influences of downstream process processing. For some products yield is measured after downstream processing, including the effects of downstream processing could be challenging as the number of process parameters measured in this phase of the production process is limited. This limitation is discussed with MSD where it was mentioned that the effect of downstream processing on yield could potentially be modelled as a constant factor that is multiplied with the result obtained from the prediction model of yield. However, this should be further researched when considering downstream processing for the RTRT approach of yield. In addition, only one contaminating bacteria is analyzed eventhough purity refers to the absence of any contaminating bacteria. Replacing the traditional testing method with the RTRT approach requires analysis on more contaminating bacteria. Lastly, only the quality attributes purity and yield are analyzed eventhough many more quality attributes exist. These are mostly dichotomous quality attributes that can be modelled in a similar way as proposed for purity.

Future work could expand this research to other production processes and quality attributes. Especially for sterility and inactivation tests RTRT would result in substantial cost and leadtime reduction. However, this may be challenging due to the limited amount of process parameters that are available for modelling downstream process behavior. In addition, expanding the joint optimization model to the field of condition based maintenance could be interesting. In particular, maintenance policies that use multiple parameters to determine when preventive maintenance should be conducted could benefit from the use of this model.

Moreover, future research could focuses on expanding growth curve modelling to other applications such as process monitoring or process improvement and other end-users, for example, the food industry.

Chapter 8

Recommendations

The conclusions and managerial insight presented in Chapter 8 are used to formulate recommendations that support MSD in the process to RTRT approval. These recommendations are based on the research questions defined at the beginning of this research.

MSD, and the bio-pharmaceutical industry in general, expressed a need for a method that describes the steps that need to be conducted for RTRT to be applied and approved by authorities. This research has developed a iterative framework that describes the 6 phases of introducing RTRT as control strategy. It is recommended to use this framework for future RTRT initiatives and for discussion with the authorities. It is mentioned in authority documents that using a scientific approach to RTRT application can support assessors and inspectors in understanding the reasoning behind the suggested approach and steps conducted (EMA, 2012). Therefore, using this framework will facilitate the approval for RTRT application.

An important element of the formulation of a RTRT approach is the risk assessment; defining the critical process parameters for a certain quality attribute (EMA, 2012). This research and other studies (Torres, 2017) found that critical process parameters are often real-time measured process parameter or are derived from real-time measured process parameters. To date, process parameters measured real-time during the main fermentation process are not analyzed with growth curve modelling. Not only does this limit analysis to only detecting observable curve deviations, the RTRT approaches require quantitative analysis to determine curve parameter that can serve as input for the RTRT models. Therefore, it is recommended that for the process parameters that provide input for the RTRT models the corresponding growth curve models are implemented. In addition, it is recommended that for all real-time process parameters growth curve modelling is implemented also for those which do not serve as input for the RTRT models. The insights gained from implementation growth curve modelling can be valuable for quantifying growth behaviour in case of process issues or for process improvement projects. The implementation of growth curve modelling would be a two-phase process as illustrated in Figure 8.1. The first phase would be introducing growth curve modelling for the real-time parameters used for RTRT for the product and production line in scope by manual data processing. The second phase starts after systems are installed that allow for automatic data processing. During this phase, growth curve modelling is applied for all

products and production lines for each real-time measured process parameter.

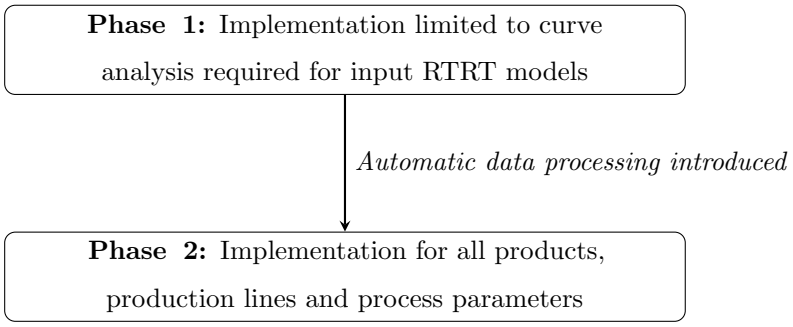


Figure 8.1: Two phased implementation process growth curve modelling

After the implementation of the growth curve models to obtain the critical process parameters from the real-time measured process parameters, the RTRT approaches for purity and yield can be implemented in parallel to the current testing method. Starting with purity, it is recommended to conduct parallel testing for purity using the critical process parameter and control limit from the one parameter model. The critical process parameters and corresponding control limits obtained from the two parameter model are not recommended as data manipulation is used for the second critical process parameter. Additionally, the validation of the models presented in the numerical analysis is based on contamination data at experimental scale. In the EMA documents it is mentioned that the RTRT model validation should be supported by substantial comparative data at commercial scale therefore it is advised to conduct parallel testing which supports further model validation at commercial scale (EMA, 2012). The advice is to use the information on misclassifications during parallel testing to update the model and improve model accuracy in particular especially since the contamination data used for the model construction is on experimental scale. Another recommendation is to quantify lost profit as a result of a false positive misclassification. This would provide a more representative solution in terms of cost impact for false positive misclassifications.

Moreover, it is advised to conduct research on additional process parameters critical to purity with the focus on process parameters that have a strong positive relation with purity. The accuracy of the purity model could be improved by finding additional process parameters that are indicative for the presence of late contaminations and can be used as check for purity. A two-fold recommendation as illustrated in Figure 8.2 is proposed: (1) expand the RTRT approach to products that have more contamination data on commercial scale and (2) conduct research on potential additional process parameters that can be used to predict purity. On commercial scale additional process parameters are measured that have not been included in this research due to limitations in experimental scale but may be critical to purity.

In addition, the recommendation is to conduct further research on various types of contaminating bacteria by either conducting experiments or evaluating contamination data for products that have experienced different types of contaminations. For this research only one contaminating bacteria is considered although purity refers the absence of any contaminating bacteria. Contaminating bacteria can differ in growth characteristics which can influence the detectability of a certain bacteria. Especially contaminating bacteria with a similar growth pattern or lower growth rate in comparison to the cultivating bacteria

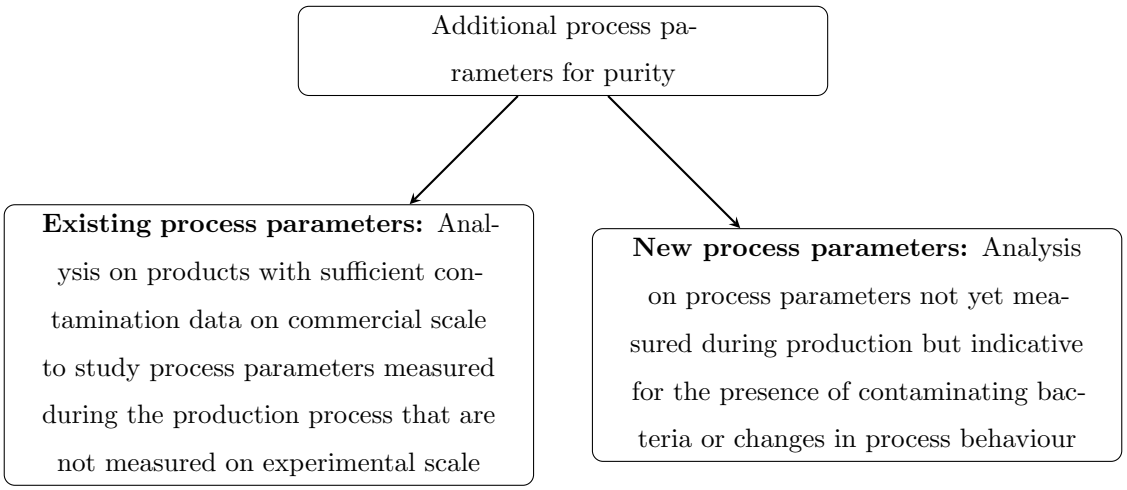


Figure 8.2: Two-fold recommendations for introducing additional process parameter to the RTRT approach for purity

may be harder to detect. Replacing the traditional purity test with a RTRT approach requires extensive analysis on all bacteria that can possibly cause a contamination in the production environment of MSD. Defining the control limits for multiple contaminating bacteria can be done in two ways: (1) formulate a set of control limits for each contaminating bacteria where all control limits need to be met for a batch to be classified as pure or (2) the contamination data including the various contaminating bacteria is combined and one set of control limits is determined. The former shows the difference between contaminating bacteria and provides insight in detection sensitivity. The latter is easier in use and includes the trade-off on detecting various types of contaminations.

Similar to the RTRT approach for purity, it is recommend to conduct further parallel testing for the RTRT approach for yield to increase data availability for validation. In addition, it is advised to increase the data availability of the measurement system analysis for the traditional testing method with at least five batches. For this study, the measurement system analysis is based on five batches which is below the advised number of 10 batches for these type of tests. Results have shown the importance of data availability when evaluating the performance of a testing method or RTRT model. Expanding the existing measurement system analysis would enable a better comparison of the performance of the two methods and would provide the required information on test performance for the discussion with authorities.

Although the focus is on further expanding and validating the RTRT approach, MSD should keep in mind that for RTRT approval additional documentation is required. EMA (2012) has mentioned that for RTRT procedures for reporting, model maintenance plans and contingency plans should be in place. It is recommended that during validation procedures possible risks are defined and included in the contingency plan. This contingency plan should be based on existing standard operation procedures on quality risk management and expanded with requirements on RTRT mentioned by EMA (2012). For purity, the risk of not detecting late contamination in the fermentation process should be further investigated and discussed with quality representatives. The yield prediction model also knows some risks when the over- or underestimation is significant and impacts downstream processes.

Lastly, the advice is to use the yield dashboard that is implemented to identify other products for which a RTRT approach can be used for to predict yield. The yield model developed in this study is applicable for various products allowing for further expanding RTRT for yield and collecting validation data for parallel testing. In addition, it is recommended to formulate actions plans for trends in order to speed up the process of analyzing trends and solving process issues.

References

Abbasian, H., Zenouz, R., Abdollahiasl, A., Toroski, M., Nikfar, S., Shadabad, M., & Kebriaeezadeh, A. (2021, 01). Risk factors of supply chain in biopharmaceutical companies in iran. doi: 10.34172/PS.2020.93

Aebtarm, S., & Bouguila, N. (2011a). An empirical evaluation of attribute control charts for monitoring defects. *Expert Systems with Applications*, *38*(6), 7869 - 7880. Retrieved from http://www.sciencedirect.com/science/article/pii/S0957417410014569 doi: https://doi.org/10.1016/j.eswa.2010.12.093

Aebtarm, S., & Bouguila, N. (2011b). An empirical evaluation of attribute control charts for monitoring defects. *Expert Syst. Appl.*, *38*(6), 7869–7880. Retrieved from https://doi.org/10.1016/j.eswa.2010.12.093 doi: 10.1016/j.eswa.2010.12.093

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H., & Radiom, S. (2018). Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(12), 4563-4577. doi: 10.1109/JSTARS.2018.2823361

Ahangar, N., & Chimka, J. (2015). Attribute control charts with optimal limits. *Quality and Reliability Engineering International*, *32*(4). Retrieved from https://doi.org/10.1002/qre.1839

Ahmad, R., & Kamaruddin, S. (2012). An overview of time-based and condition-based maintenance in industrial application. *Computers Industrial Engineering*, *63*(1), 135 - 149. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360835212000484 doi: https://doi.org/10.1016/j.cie.2012.02.002

Allman, T. (2020, June). *How to troubleshoot contamination of bioreactors.* Infors HT. Retrieved from https://www.infors-ht.com/en/blog/how-to-troubleshoot-contamination-of-bioreactors/

Annadurai, G., Babu, S., & Srinivasamoorthy, V. (2000). Development of mathematical models (logistic, gompertz and richards models) describing the growth pattern of pseudomonas putida (nicm 2174). *Bioprocess Engineering*, *23*, 607- 612. doi: https://doi.org/10.1007/s004490000209

Arnold, A., Cortada, J., Gledndinning, K., & Henderson, L. (n.d.). *Real-time total cell density measurement of yeast fermentations.* Retrieved from https://craft-sensors.s3.amazonaws.com/File-Uploads/CD_Yeast-Fermentations_AppNote.pdf?mtime=20180625105216&focal=none

Bailey, R. (2018, September). *Phases of the bacterial growth curve.* Retrieved from https://www.thoughtco.com/bacterial-growth-curve-phases-4172692

Banjevic, D., Jardine, A., Makis, V., & Ennis, M. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR: Information Systems and Operational Research*, *39*(1), 32-50. doi: 10.1080/03155986.2001.11732424

Basílio, A., De Araújo, P., De Morais, J., Da Silva Filho, E., de Morais, M., & Simões, D. (2008). Detection and identification of wild yeast contaminants of the industrial fuel ethanol fermentation process. *Current Microbiology*, *56*, 322-326. doi: https://doi.org/10.1007/s00284-007-9085-5

Blackwell, J. (2017, January). *Troubleshooting bacterial contamination in bioreactors.* Bioprocess Online. Retrieved from https://www.bioprocessonline.com/doc/troubleshooting-bacterial-contamination-in-bioreactors-0001

Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486 doi: https://doi.org/10.1016/S0305-0548(01)00108-3

Cadinoska, M., Popstefanova, N., Ilievska, M., Karadzinska, E., Jovanoska, M., & Dodov, M. (2019). Trending and out-of-trend results in pharmaceutical industry. *Macedonian Pharmaceutical Bulletin*, *65*(1).

Chang, M., Liou, J., & Lo, H. (2019). A hybrid mcdm model for evaluating strategic alliance partners in the green biopharmaceutical industry. *Sustainability*, *11*(15). Retrieved from https://www.mdpi.com/2071-1050/11/15/4065 doi: 10.3390/su11154065

Chang, T., & Gan, F. (1995). A cumulative sum control chart for monitoring process variance. *Journal of Quality Technology*, *27*(2), 109-119. doi: 10.1080/00224065.1995.11979574

Chatterjee, T., Chatterjee, B., Majumdar, D., & Chakrabarti, P. (2015). Antibacterial effect of silver nanoparticles and the modeling of bacterial growth kinetics using a modified gompertz model. *Biochimica et Biophysica Acta*, *1850*(2), 299-306. doi: https://doi.org/10.1016/j.bbagen.2014.10.022

Chen, D., & Trivedi, K. (2005). Optimization for condition-based maintenance with semi-markov decision process. *Reliability Engineering System Safety*, *90*(1), 25 - 29. Retrieved from http://www.sciencedirect.com/science/article/pii/S0951832004002601 doi: https://doi.org/10.1016/j.ress.2004.11.001

Dalgaard, P., & Koutsoumanis, K. (2001). Comparison of maximum specific growth rates and lag times estimated from absorbance and viable count data by different mathematical models. *Journal of Microbiological Methods*, *43*(3), 183-196. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167701200002190 doi: https://doi.org/10.1016/S0167-7012(00)00219-0

Dasgupta, R. (2017). *Growth curve models and applications.*

de Souza Liberal , A., da Silva Filho, E., de Morais, J., Simoes, D., & de Morais, M. (2005). Contaminant

yeast detection in industrial ethanol fermentation must by rdna-pcr. *Letters in Applied Microbiology*, *40*, 19-23. doi: https://doi.org/10.1111/j.1472-765X.2004.01618.x

Elmroth, I., Valeur, A., Odham, G., & Larsson, L. (1990). Detection of microbial contamination in fermentation processes: Mass spectrometric determination of gram-negative bacteria inleuconostoc mesenteroides cultures. *Biotechnology and Bioengineering*, *35*, 787-792. doi: https://doi.org/10.1002/bit.260350806

EMA. (2004). *Ich guideline q8 (r2) on pharmaceutical development* (Tech. Rep.). Retrieved from `https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-11.pdf"`

EMA. (2006a). *Guideline on parametric release* (Tech. Rep.). Retrieved from `https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-parametric-release_en.pdf`

EMA. (2006b). *Ich guideline q9 on quality risk management* (Tech. Rep.). Retrieved from `https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-3.pdf`

EMA. (2012). *Guideline on real time release testing (formerly guideline on parametric release)* (Tech. Rep.). Retrieved from `https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-real-time-release-testing-formerly-guideline-parametric-release-revision-1_en.pdf`

Factors of safety and reliability in geotechnical engineering. (n.d.). *Journal of Geotechnical and Geoenvironmental Engineering, volume = 126, year = 2000, url = https://ascelibrary.org/doi/abs/10.1061/(ASCE)1090-0241(2000)126:4(307), author =J. Duncan.*

Farid, S., Washbrook, J., & Titchener-Hooker, N. (2008). Decision-support tool for assessing biomanufacturing strategiesunder uncertainty: Stainless steel versus disposable equipment forclinical trial material preparation. *Biotechnology Progress*, *21*(2), 486-497. doi: 10.1021/bp049692b

FDA. (2010, February). *Guidance for industry - submission of documentation in applications for parametric release of human and veterinary drug products terminally sterilized by moist heat processes.* Retrieved from `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submission-documentation-applications-parametric-release-human-and-veterinary-drug-products`

Frey, B. (2018). *Growth curve modelling.* doi: https://dx.doi.org/10.4135/9781506326139.n296

Fujikawa, H., Kai, A., & Morozumi, S. (2004). A new logistic model for escherichia coli growth at constant and dynamic temperatures. *Food Microbiology*, *21*(5), 501-509. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0740002004000073` doi: https://doi.org/10.1016/j.fm.2004.01.007

Gandhi, A., & Roy, C. (2016). Quality by design (qbd) in pharmaceutical industry: Tools, perspectives and challenges. *PharmaTutor*, *4*(11), 12-20. Retrieved from https://www.pharmatutor.org/articles/quality-by-design-qbd-in-pharmaceutical-industry-tools-perspectives-and-challenges

Glushkovsky, E. (1994). 'on-line' g-control chart for attribute data. *Quality and Reliability Engineering International*, *10*(3). doi: https://doi.org/10.1002/qre.4680100312

Golmakani, H., & Fattahipour, F. (2011). Age-based inspection scheme for condition-based maintenance. *Journal of Quality in Maintenance Engineering*, *17*(1), 93 - 110. doi: https://doi.org/10.1108/13552511111116277

Goyal, M. (2014). Modeling of sediment yield prediction using m5 model tree algorithm and wavelet regression. *Water Resources Management*, *28*, 1991-2003. doi: https://doi.org/10.1007/s11269-014-0590-6

Grand View Research. (2020, February). *Animal health market size, share trends analysis report by animal type (production, companion), by end use, by product (pharmaceuticals, feed additives, vaccines), by distribution channel, and segment forecasts, 2020 - 2027.* Retrieved from https://www.grandviewresearch.com/industry-analysis/animal-health-market

Gupta, V., Jain, R., Meena, M., & Dangayach, G. (2018). Six-sigma application in tire-manufacturing company: a case study. *Journal of Industrial Engineering International*, *14*, 511-520. doi: https://doi.org/10.1007/s40092-017-0234-6

Gurău, C. (2004). Positioning strategies in the value-added chain of the biopharmaceutical sector: the case of uk smes. *Journal of Consumer Marketing*, *21*(7), 476 - 485. doi: https://doi.org/10.1108/07363760410568699

Hall, J., McNeil, B., Rollins, M., Draper, I., Thompson, B., & Macaloney, G. (1996). Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial escherichia coli fermentation. *Applied Spectroscopy*, *50*(1), 102–108. Retrieved from http://as.osa.org/abstract.cfm?URI=as-50-1-102

Harrell, F. (2015). *Binary logistic regression.* Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-19425-7_10 doi: 10.1007/978-3-319-19425-7_10

Heng, A., Tan, A., Mathew, J., Montgomery, N., Banjevic, D., & Jardine, A. (2009). Intelligent condition-based prediction of machinery reliability. *Mechanical Systems and Signal Processing*, *23*(5), 1600 - 1614. Retrieved from http://www.sciencedirect.com/science/article/pii/S088832700900003X doi: https://doi.org/10.1016/j.ymssp.2008.12.006

Holder, M., Devpura, A., Lee, A., & Chandran, S. (2018). Aligning data analytics and supply chain strategy in the biopharmaceutical industry. *Aligning Business Strategies and Analytics*, 67 - 78. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-93299-6_5 doi: https://doi.org/10.1007/978-3-319-93299-6_5

Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., & Lübbert, A. (2006). Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioprocess and Biosystems Engineering*, *29*, 19-27. doi: https://doi.org/10.1007/s00449-006-0051-6

Ji, B., Sun, Y., Yang, S., & Wan, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions. *The Journal of Agricultural Science*, *145*(3), 249–261. doi: 10.1017/S0021859606006691

Joghee, R. (2017). Control chart for high-quality processes based on six sigma quality. *International Journal of Quality Reliability Management*, *34*(1), 2-17. doi: https://doi.org/10.1108/IJQRM-05-2015-0080

Juška, A., Gedminienė, G., & Ivanec, R. (2007). Growth of microbial populations. mathematical modeling, laboratory exercises, and model-based data analysis. *Biochemistry and Molecular Biology Education*, *34*(6). doi: https://doi.org/10.1002/bmb.2006.494034062669

Kaminsky, P., & Wang, Y. (2015). Analytical models for biopharmaceutical operations and supply chain management: a survey of research literature. *Pharmaceutical Bioprocessing*, *3*, 61-73.

Kiviharju, K., Salonen, K., Moilanen, U., & Eerikäinen, T. (2008). Biomass measurement online: the performance of in situ measurements and software sensors. *Journal of Industrial Microbiology and Biotechnology*, *35*(7), 657-665. Retrieved from `https://doi.org/10.1007/s10295-008-0346-5` doi: 10.1007/s10295-008-0346-5

Koca, Y., Martagan, T., Adan, I., Maillart, L., & van Ravenstein, B. (2020). Increasing biomanufacturing yield with bleed-feed: Optimal policies and insights.
doi: http://dx.doi.org/10.2139/ssrn.3659907

Konishi, S., & Kitagawa, G. (2008). *Information criteria and statistical modeling.* New York: Springer. doi: https://doi.org/10.1007/978-0-387-71887-3

Kostov, G., Popova, S., Gochev, V., & Koprinkova-Hristova, P. (2012). Modeling of batch alcohol fermentation with free and immobilized yeasts saccharomyces cerevisiae 46 evd. *Biotechnology Biotechnological Equipment*, *26*(3), 3021-3030. doi: https://doi.org/10.5504/BBEQ.2012.0025

La Rosa, F., & Liberatore, G. (2014). Biopharmaceutical and chemical firms' rd disclosure, and cost of equity: The impact of the regulatory regime. *European Management Journal*, *32*(5), 806 - 820. Retrieved from `http://www.sciencedirect.com/science/article/pii/S026323731400019X` doi: https://doi.org/10.1016/j.emj.2014.01.003

Larouzee, J., & Le Coze, J. (2020). Good and bad reasons: The swiss cheese model and its critics. *Safety Science*, *126*, 104660. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0925753520300576` doi: https://doi.org/10.1016/j.ssci.2020.104660

Laursen, K., Frederiksen, S., Leuenhagen, C., & Bro, R. (2010). Chemometric quality control of chromatographic purity. *Journal of Chromatography*, *1217*, 6503-6510. doi: https://doi.org/10.1016/j.chroma.2010.08.040

Leachman, R., Johnston, L., Li, S., & Shen, Z. (2014). An automated planning engine for biopharmaceutical

production. *European Journal of Operational Research*, *238*(1), 327 - 338. Retrieved from http://www.sciencedirect.com/science/article/pii/S0377221714002161 doi: https://doi.org/10.1016/j.ejor.2014.03.002

Lim, A., Zhou, Y., Washbrook, J., Titchener-Hooker, N., & Farid, S. (2004). A decisional-support tool to model the impact of regulatory compliance activities in the biomanufacturing industry. *Computers Chemical Engineering*, *28*(5), 727 - 735. Retrieved from http://www.sciencedirect.com/science/article/pii/S0098135404000407 (ESCAPE 13) doi: https://doi.org/10.1016/j.compchemeng.2004.02.013

Limon, Y., & Krishnamurthy, A. (2020). Dynamic resource scheduling of biomanufacturing projects. *Computers Industrial Engineering*, *147*, 106527. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360835220302618 doi: https://doi.org/10.1016/j.cie.2020.106527

Loutfi, H., Le Jeune, B., Ltief, R., Kallassay, M., Le Brun, G., & Abbound, M. (2018). Real-time monitoring of bacterial growth kinetics in suspensions using laser speckle imaging. *A Nature Research Journal*, *10*(408). doi: https://doi.org/10.1038/s41598-019-57281-2

Lu, X., Xie, M., Goh, T., & Lai, C. (1998a). Control chart for multivariate attribute processes. *International Journal of Production Research*, *36*, 3477- 3489. doi: 10.1080/002075498192166

Lu, X., Xie, M., Goh, T., & Lai, C. (1998b, 12). Control chart for multivariate attribute processes. *International Journal of Production Research*, *36*, 3477- 3489. doi: 10.1080/002075498192166

Lu, Z., Sebranek, J., Dickson, J., Mendonca, A., & Bailey, T. (2005). Application of predictive models to estimate listeria monocytogenes growth on frankfurters treated with organic acid salts. *Journal of Food Protection*, *68*(11), 2326-2332. doi: https://doi.org/10.4315/0362-028X-68.11.2326

Ma, Y. (2011). Risk management in biopharmaceutical supply chains.. Retrieved from https://escholarship.org/uc/item/70d7s6xw

Martagan, T. (2015). Stochastic models to optimize biomanufacturing operations. *University of Wisconsin-Madison*.

Martagan, T., Koca, Y., Adan, I., van Ravenstein, R., Baaijens, M., & Repping, O. (2020). Operations research improves biomanufacturing efficiency at msd animal health. *INFORMS Journal on Applied Analytics*.

Martagan, T., Krishnamurthy, A., & Maravelias, C. (2016). Optimal condition-based harvesting policies for biomanufacturing operations with failure risks. *IIE Transactions*, *48*(5), 440-461. doi: 10.1080/0740817X.2015.1101523

Mazzola, E., Bruccoleri, M., & Perrone, G. (2015). Supply chain of innovation and new product development. *Journal of Purchasing and Supply Management*, *21*(4), 273 - 284. Retrieved from http://www.sciencedirect.com/science/article/pii/S1478409215000400 doi: https://doi.org/10.1016/j.pursup.2015.04.006

McGillicuddy, N., Floris, P., Albrecht, S., & Bones, J. (2018). Examining the sources of variability in cell culture media used for biopharmaceutical production. *Biotechnology Letters*, *40*(1), 5 - 21. doi: https://doi.org/10.1007/s10529-017-2437-8

Mid, E., & Dua, V. (2018). Fault detection of fermentation processes. *Computer Aided Chemical Engineering*, *43*, 1171-1176. doi: https://doi.org/10.1016/B978-0-444-64235-6.50204-7

Mohebbi, E. (2003). Supply interruptions in a lost-sales inventory system with random lead time. *Computers Operations Research*, *30*(3), 411-426. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0305054801001083` doi: https://doi.org/10.1016/S0305-0548(01)00108-3

Mytilinaios, I., Salih, M., Schofield, H., & Lambert, R. (2012). Growth curve prediction from optical density data. *International Journal of Food Microbiology*, *154*(3), 169-176. Retrieved from `https://www.sciencedirect.com/science/article/pii/S016816051100763X` doi: https://doi.org/10.1016/j.ijfoodmicro.2011.12.035

Nieuwoudt, A., Ragheb, T., Nejati, H., & Massoud, Y. (2007). Increasing manufacturing yield for wideband rf cmos lnas in the presence of process variations. In *8th international symposium on quality electronic design (isqed'07)* (p. 801-806). doi: 10.1109/ISQED.2007.89

Panesar, P., Kennedy, J., Knill, C., & Kosseva, M. (2007). Applicability of pectate-entrapped lactobacillus casei cells for l(+) lactic acid production from whey. *Applied Microbiology and Biotechnology*, *74*, 35-42. Retrieved from `https://link.springer.com/article/10.1007/s00253-006-0633-x`

Papapetridis, I., Goudriaan, M., M. Vitali, N. d., van den Broek, M., van Maris, A., & Pronk, J. (2018). Optimizing anaerobic growth rate and fermentation kinetics in saccharomyces cerevisiae strains expressing calvin-cycle enzymes for improved ethanol yield. *Biotechnology for Biofuels volume*, *11*(17). doi: https://doi.org/10.1186/s13068-017-1001-z

Pawar, P., Wang, Y., Keyvan, G., Callegari, G., Cuitino, A., & Muzzio, F. (2016). Enabling real time release testing by nir prediction of dissolution of tablets made by continuous direct compression (cdc). *International Journal of Pharmaceutics*, *512*(1), 96-107. doi: 10.1016/j.ijpharm.2016.08.033

Peleg, M., & Corradini, M. (2011). Microbial growth curves: What the models tell us and what they cannot. *Critical Reviews in Food Science and Nutrition*, *51*(10), 917-945. doi: 10.1080/10408398.2011.570463

Phillips, J., Cripps, E., Lau, J., & Hodkiewicz, M. (2015). Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing*, *60-61*, 316-325. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0888327014005093` doi: https://doi.org/10.1016/j.ymssp.2014.12.020

Ramirez-Arcos, S., DiFranco, C., McIntyre, T., & Goldman, M. (2017). Residual risk of bacterial contamination of platelets: six years of experience with sterility testing. *Transfusion*, *57*, 2174-2181. doi: https://doi.org/10.1111/trf.14202

Remund, K., Dixon, D., Wright, D., & Holden, L. (2001). Statistical considerations in seed purity testing for transgenic traits. *Seed Science Research*, *11*, 101-119. doi: https://doi.org/10.1079/SSR200166

Rivera-Gómez, H., Gharbi, A., Kenné, J., Montaño-Arango, O., & Corona-Armenta, J. (2020). Joint optimization of production and maintenance strategies considering a dynamic sampling strategy for a deteriorating system. *Computers Industrial Engineering*, *140*, 106-273. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0360835220300073` doi: https://doi.org/10.1016/j.cie.2020.106273

Runje, B., Novak, A., & Razumić, A. (2017). *Measurement system analysis in production process.* XVII International Scientific Conference on Industrial Systems (IS'17). Retrieved from `http://www.iim.ftn.uns.ac.rs/is17`

Sanaeifar, A., ZakiDizaji, H., Jafari, A., & Guardia, M. (2017). Early detection of contamination and defect in foodstuffs by electronic nose: A review. *TrAC - Trends in Analytical Chemistry*, *97*, 257-271. doi: https://doi.org/10.1016/j.trac.2017.09.014

Scarf, P. (1997). On the application of mathematical models in maintenance. *European Journal of Operational Research*, *99*(3), 493 - 506. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0377221796003165` (Eleventh EURO Summer Institute: Operational Research Models in Maintenance) doi: https://doi.org/10.1016/S0377-2217(96)00316-5

Shafiee, M., & Finkelstein, M. (2015). An optimal age-based group maintenance policy for multi-unit degrading systems. *Reliability Engineering System Safety*, *134*, 230 - 238. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0951832014002324` doi: https://doi.org/10.1016/j.ress.2014.09.016

Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. (2018). Smart farming system: Crop yield prediction using regression techniques. *Proceedings of International Conference on Wireless Communication*, 49-56. doi: https://doi.org/10.1007/978-981-10-8339-6_6

Shakeri, R., & R, R. (2017). Antecedents of strategic alliances performance in biopharmaceutical industry: A comprehensive model. *Technological Forecasting and Social Change*, *122*, 289 - 302. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0040162516000044` doi: https://doi.org/10.1016/j.techfore.2016.01.003

Shakoor, M. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017). Agricultural production output prediction using supervised machine learning techniques. In *2017 1st international conference on next generation computing applications (nextcomp)* (p. 182-187). doi: 10.1109/NEXTCOMP.2017.8016196

Shin, K., Kim, S., & Park, G. (2015). How does the partner type in r&d alliances impact technological innovation performance? a study on the korean biotechnology industry. , *33*, 141 - 164. doi: https://doi.org/10.1007/s10490-015-9439-7

Singh, A. (2015, October). *Industry perspective on pre-approval inspection (pai).* North Bethesda,

Maryland: PDA/PQRI Conference on Evolving Product Quality. Retrieved from http://pqri.org/wp-content/uploads/2015/10/01-FDA-PQRI-Oct-2015-A-Singh-Talk-Final-Sept-30.pdf

S.Sivakesava, Irudayaraj, J., & Ali, D. (2001). Simultaneous determination of multiple components in lactic acid fermentation using ft-mir, nir, and ft-raman spectroscopic techniques. *Process Biochemistry*, *37*(4), 371-378. Retrieved from https://www.sciencedirect.com/science/article/pii/S0032959201002230 doi: https://doi.org/10.1016/S0032-9592(01)00223-0

Sue, T., Obolonkin, V., Griffiths, H., & Villas-Bôas, S. (2011). An exometabolomics approach to monitoring microbial contamination in microalgal fermentation processes by using metabolic footprint analysis. *Applied and Environmental Microbiology*, *77*, 7605-7610. doi: https://doi.org/10.1128/AEM.00469-11

Sun, Y., Tang, L., Sun, Q., Wang, M., Han, X., & Chen, X. (2019). Study on the prediction of the contamination symptoms in the fermentation process of chlortetracycline based on soft sensor modeling method. *Technology and health care : official journal of the European Society for Engineering and Medicine*, *27*, 205–215. doi: https://doi.org/10.3233/THC-199020

Suvarna, K., Lolas, A., Hughes, P., & Friedman, R. (2011, January). *Case studies of microbial contamination in biologic product manufacturing.* American Pharmaceutical Review - The Review of American Pharmaceutical Business Technology. Retrieved from https://www.americanpharmaceuticalreview.com/Featured-Articles/36755-Case-Studies-of-Microbial-Contamination-in-Biologic-Product-Manufacturing/

Tagaras, G. (1988). An integrated cost model for the joint optimization of process control and maintenance. *The Journal of the Operational Research Society*, *39*(8), 757–766. Retrieved from http://www.jstor.org/stable/2583771

Tirumalai, R., & Porter, D. (2005). Terminal sterilization and potential for parametric release. *US Pharmacopeia*, 26-31. doi: http://microbiologynetwork.com/content/terminal_sterilization_potential_for_parametric_release.pdf

Todar, K. (n.d.). *The growth of bacterial populations.* Retrieved from http://textbookofbacteriology.net/growth_3.html

Topalidou, E., & Psarakis, S. (2009). Review of multinomial and multiattribute quality control charts. *Quality and Reliability Engineering International*, *25*(7). doi: https://doi.org/10.1002/qre.999

Torres, J. (2017, March). *A case study in real-time release testing.* Parental Drug Association. Retrieved from https://www.bioprocessonline.com/doc/a-case-study-in-real-time-release-testing-0001?vm_tId=2057140&user=0ce0a71d-cc09-42c1-a588-1d7b23f02077&utm_source=et_6212871&utm_medium=email&utm_campaign=BIO_03-21-2018&utm_term=0ce0a71d-cc09-42c1-a588-1d7b23f02077&utm_content=A+Case+Study+In+Real-Time+Release+Testing

Umesh. (2012). *Pharmaceutical "quality by design" (qbd): An introduction, process development and applications.* Retrieved from https://learnaboutgmp.com/good-validation-practices/pharmaceutical-quality-by-design-qbd-an-introduction-process-development-and-applications/

Vieira, M., Pinto-Varela, T., & Barbosa-Póvoa, A. (2019). A model-based decision support framework for the optimisation of production planning in the biopharmaceutical industry. *Computers Industrial Engineering*, *129*, 354 - 367. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360835219300506 doi: https://doi.org/10.1016/j.cie.2019.01.045

Vieira, M., Pinto-Varela, T., Moniz, S., Barbosa-Póvoa, A., & Papageorgiou, L. (2016). Optimal planning and campaign scheduling of biopharmaceutical processes using a continuous-time formulation. *Computers Chemical Engineering*, *91*, 422 - 444. Retrieved from http://www.sciencedirect.com/science/article/pii/S0098135416301053 (12th International Symposium on Process Systems Engineering 25th European Symposium of Computer Aided Process Engineering (PSE-2015/ESCAPE-25), 31 May - 4 June 2015, Copenhagen, Denmark) doi: https://doi.org/10.1016/j.compchemeng.2016.04.009

Walch, N., Scharl, T., Felföldi, E., Sauer, D., Melcher, M., Leisch, F., ... Jungbauer, A. (2019). Prediction of the quantity and purity of an antibody capture process in real time. *Biotechnology Journal*, *14*. doi: https://doi.org/10.1002/biot.201800521

Wang, B., Xie, W., Martagan, T., Akcay, A., & Corlu, C. G. (2019). Stochastic simulation model development for biopharmaceutical production process risk analysis and stability control. In *2019 winter simulation conference (wsc)* (p. 1989-2000). doi: 10.1109/WSC40007.2019.9004778

Wang, H. (2009). Comparison of p control charts for low defective rate. *Computational Statistics Data Analysis*, *53*(12), 4210 - 4220. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167947309001996 doi: https://doi.org/10.1016/j.csda.2009.05.024

Wang, M., Han, X., Sun, Q., & Chen, X. (2021). Study on soft sensor modeling method for sign of contaminated fermentation broth in chlortetracycline fermentation process. *Preparative Biochemistry and Biotechnology*, *51*, 76-85. doi: https://doi.org/10.1080/10826068.2020.1793173

Wechselberger, P., Sagmeister, P., & Herwig, C. (2013). Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess and Biosystems Engineering volume*, 1205-1218. doi: https://doi.org/10.1007/s00449-012-0848-4

Xie, M., Goh, T., & Kuralmani, V. (2000). On optimal setting of control limits for geometric chart. *International Journal of Reliability*, *7*, 17- 25. doi: https://doi.org/10.1142/S0218539300000031

Xie, M., Goh, T., & Ranjan, P. (2002). Some effective control chart procedures for reliability monitoring. *Reliability Engineering System Safety*, *77*(2), 143 - 150. Retrieved from http://www.sciencedirect.com/science/article/pii/S0951832002000418 doi: https://doi.org/10.1016/S0951-8320(02)00041-8

Yang, J., Chen, X., & Jin, H. (2015). Online prediction for contamination of chlortetracycline fermentation based on dezert-smarandache theory. *Chinese Journal of Chemical Engineering*, *23*, 1009-1016. doi: https://doi.org/10.1016/j.cjche.2014.06.0433

Yang, Y., & Tjia, R. (2010). Process modeling and optimization of batch fractional distillation to increase throughput and yield in manufacture of active pharmaceutical ingredient (api). *Computers Chemical*

Engineering, *34*(7), 1030-1035. Retrieved from https://www.sciencedirect.com/science/article/pii/S0098135410001304 (Process Modeling and Control in Drug Development and Manufacturing) doi: https://doi.org/10.1016/j.compchemeng.2010.03.019

Yardley, J., Kell, D., Barrett, J., & Davey, C. (2000). On-line, real-time measurements of cellular biomass using dielectric spectroscopy. *Biotechnology and Genetic Engineering Reviews*, *17*(1), 3-36. Retrieved from https://doi.org/10.1080/02648725.2000.10647986 doi: 10.1080/02648725.2000.10647986

Yu, L., Amidon, G., Khan, M., Hoag, S., Polli, J., Raju, G., & Woodcock, J. (2014). Understanding pharmaceutical quality by design. *The AAPS Journal*, *16*(4), 771–783. doi: 10.1208/s12248-014-9598-3

Zhang, C., Xie, M., & Goh, T. (2006). Design of exponential control charts using a sequential sampling scheme. *IIE Transactions*, *38*(12), 1105-1116. Retrieved from https://doi.org/10.1080/07408170600728905 doi: 10.1080/07408170600728905

Zhang, M., Peng, Y., Schuh-Renner, A., Megahed, F., & Woodall, W. (2013, 03). Geometric charts with estimated control limits. *Quality and Reliability Engineering International*, *29*, 209-223. doi: 10.1002/qre.1304

Ziegel, E. (1992). *Understanding industrial experimentation* (2nd ed.). SPC Press.

Zou, X., Li, W., & Wang, K. (2013). Ergodic method on optimal harvesting for a stochastic gompertz-type diffusion process. *Applied Mathematics Letters*, *26*(1), 170-174. Retrieved from https://www.sciencedirect.com/science/article/pii/S0893965912003564 doi: https://doi.org/10.1016/j.aml.2012.08.006

Zwietering, M., Jongenburger, I., Rombouts, F., & van 't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, *56*(6), 1875-1881. doi: 10.1128/AEM.56.6

Appendix A

Biomass at time t is denoted by m(t). The derivative of m(t), m'(t), in combination with the biomass at time t can be used to determine the relative growth rate at time t, referred to as $\mu(m(t))$.

$$\mu(m(t)) = \frac{m'(t)}{m(t)}, \qquad t \geq 0 \tag{8.1}$$

In case a constant growth rate is assumed, the expression for m(t) becomes $e^{\mu * t}$ for some value of μ. Assuming that at the start of the process the biomass quantity is non-zero because for growth to start bacteria have to be present in the process, the biomass at the start of the fermentation process m(0) is denoted with parameter b_0. This result is the mathematical expression for m(t) as presented in 8.2.

$$m(t) = m(0) * e^{\mu * t}, \qquad t \geq 0 \tag{8.2}$$

Appendix B

The batch is defined as a false positive when one of the following three scenarios occurs (assuming two process parameters):

1. The batch is pure and both critical process parameters are outside their control limits
2. The batch is pure and critical process parameter 1 is outside its control limit but critical process parameter 2 is within its control limit
3. The batch is pure and critical process parameter 1 is within its control limit but critical process parameter 2 is outside its control limit

Outside the limit can happen in two ways, the parameter value is below the lower control limit or is above the upper limit. A number of combinations exist for two critical process parameters, for example for the first scenario critical process parameter 1 can be above its limit and critical process parameter 2 can be below its limit. The possible situations for each scenario are illustrated in Table 8.1. $v_i < x_i - k_i$ models

the situation where the parameter value is below the lower control limit. $v_i > x_i + k_i$ models the situation where the parameter value is above the upper control limit. $x_i - k_i < v_i < x_i + k_i$ models the situation where the parameter value is within the control limits.

The probability function for having a false positive can be modelled by the situations for each scenario in which a false positive occurs.

Scenario	$v_1 < x_1 - k_1$	$v_1 > x_1 + k_1$	$x_1 - k_1 < v_1 < x_1 + k_1$	$v_2 < x_2 - k_2$	$v_2 > x_2 + k_2$	$x_2 - k_2 < v_2 < x_2 + k_2$
1	X			X		
	X				X	
		X			X	
		X		X		
2	X					X
		X				X
3			X	X		
			X		X	

Table 8.1: The situations for each scenario in which a false positive occurs when classifying purity using two process parameters

$$p_{FP}(k_1,k_2) = \int_{q=a_1}^{x_1-k_1}\int_{h=a_2}^{x_2-k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=a_1}^{x_1-k_1}\int_{h=x_2+k_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh +$$
$$\int_{q=x_1+k_1}^{b_1}\int_{h=a_2}^{x_2-k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1+k_1}^{b_1}\int_{h=x_2+k_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh +$$
$$\int_{q=a_1}^{x_1-k_1}\int_{h=x_2-k_2}^{x_2+k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1+k_1}^{b_1}\int_{h=x_2-k_2}^{x_2+k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh +$$
$$\int_{q=x_1-k_1}^{x_1+k_1}\int_{h=a_2}^{x_2-k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1-k_1}^{x_1+k_1}\int_{h=x_2+k_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh$$

The probability function can be further simplified to the equation below.

$$p_{FP}(k_1,k_2) = \int_{q=a_1}^{x_1-k_1}\int_{h=a_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1+k_1}^{b_1}\int_{h=a_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh +$$
$$\int_{q=x_1-k_1}^{x_1+k_1}\int_{h=a_2}^{x_2-k_2} p(q,h)*f_1(q)*f_2(h)*dq*dh + \int_{q=x_1-k_1}^{x_1+k_1}\int_{h=x_2+k_2}^{b_2} p(q,h)*f_1(q)*f_2(h)*dq*dh$$

Appendix C

A process parameter can either have a negative or a positive relation to the probability of having a pure batch. This appendix presents the equation for the probability on having a pure batch (p(v)) for a process parameter that has a positive relation to the purity of a batch.

p(v) models the probability distribution for a batch being pure for a critical process parameter which is determined with a binary logistics regression. For a critical process parameter with a positive relation on p(v), the distribution function is formulated as follows.

$$p(v) = \frac{e^{\alpha+\beta*v}}{1-e^{\alpha+\beta*v}} \qquad \forall\ v \in a, ..., b$$

where α and β are constants that are determined by maximum likelihood estimation.

Appendix D

The mathematical parameters a, b and c from the standard Gompertz model as presented in Equation 8.3 can be written as λ, A and μ_m (Zwietering et al., 1990).

$$y(t) = a * e^{-e^{b-c*t}} \tag{8.3}$$

The maximum growth rate (μ_m)

The maximum growth rate can be obtained using the inflection point of the curve which can be derived from the second derivative (Equation 8.5) with respect to t. For this, the first derivative (Equation 8.4) has to be calculated.

$$\frac{dy}{dt} = ac * e^{-e^{(b-c*t)}} * e^{(b-c*t)} \tag{8.4}$$

$$\frac{d^2y}{dt^2} = ac^2 * e^{-e^{(b-c*t)}} * e^{(b-c*t)} * (e^{(b-c*t)} - 1) \tag{8.5}$$

The inflection point is the t for which the second derivative is equal to zero; this t is called t_i (Equation 8.6). The maximum growth rate is equal to the first derivative at t_i (Equation 8.7).

$$\frac{d^2y}{dt^2} = 0 \rightarrow t_i = \frac{b}{c} \tag{8.6}$$

$$\mu_m = \left(\frac{dy}{dt}\right)_{t_i} = \frac{ac}{e} \tag{8.7}$$

Lambda (λ)

Lambda models the time at which the process transitions from the lag phase into the exponential growth phase. The duration of the lag phase can be found by the t-axis intercept of the tangent line through t_i. This tangent line is presented in Equation 8.8. The t-axis intercept is where the y of the tangent line is zero. In combination with Equation 8.7, it follows that lambda is equal to $\frac{b-1}{c}$.

$$y = \mu_m * t + \frac{a}{e} - \mu_m * t_i \tag{8.8}$$

Asymptote (A)

The asymptotic value is reached for t $\rightarrow \infty$. In this case, y becomes equal to the mathematical parameter a. In other words, A is equal to a.

Appendix E

The improvement of fit realized by adapting the standard Gompertz model can be quantified by comparing measures of Goodness of fit. Also, the gains as a consequence of this improvement of fit are quantified by comparing the prediction accuracy for yield when using different growth models for generating input. (Note: numbers are corrected for confidentiality purposes).

Quantitative Analysis of Model Fit using R-square

Starting with the model adapted for an indirectly controlled process parameter, the difference in goodness of fit between the adapted model and the standard Gompertz model is quantified using R-square. The results are presented in Table 8.2. The results show that the difference in R-square is small. The difference between the two models only affects the very small period of time at the end of the process having a

Model	R-square
Adapted model	0.9997
Standard Gompertz model	0.9928

Table 8.2: Quantitative comparison of the goodness of fit using R-square for modelling the curve of an indirectly controlled process parameter

Model	R-square
Adapted model	0.9997
Standard Gompertz model	0.9987
Model Chatterjee et al. (2015)	0.9946

Table 8.3: Quantitative comparison of the goodness of fit using R-square for modelling the curve of a process parameter that includes a death phase

limited effect on the total goodness of fit measure. For validation purposes, the numerical analysis is conducted for multiple products at MSD resulting in similar conclusions in regards to fit.

The model adapted for process parameters including a death phase is also compared with existing models using R-square (see Table 8.3). The differences are small since only the modelling of the death phase is affected by the model change which covers a very small period of the process duration. For validation purposes, the numerical analysis is conducted for multiple products at MSD resulting in similar conclusions in regards to fit.

Improvement in Prediction Accuracy of Yield Model

The adaption of this model can also influence the prediction accuracy of the yield or purity model. In this research, the model for process parameters including a death phase is used as input for the yield model. The adapted growth model allows for incorporating the death phase hereby potentially increasing prediction accuracy for yield. The increase in prediction accuracy is quantified by comparing the prediction error of the yield model based on the adapted Gompertz model to the prediction error of the yield model based on the standard Gompertz model. Using the adapted Gompertz model instead of the standard Gompertz model has led to a increase in prediction accuracy resulting in a 4 percent decrease in average prediction error.