NATIONAL TECHNICAL UNIVERSITY OF ATHENS

CHEMICAL ENGINEERING

Design Space Identification and Data-Driven Modelling of Downstream Bioprocesses

Thesis Submitted for the degree of

Master's in chemical engineering

by

Foteini Michalopoulou

Supervisors: Professor Antonis Kokossis (NTUA)

Dr. Maria Papathanasiou (Imperial College London)

Athens

June 29, 2021

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε σε συνεργασία με τη Dr Μαρία Παπαθανασίου, Λέκτορα στο Imperial College of London, την οποία ευχαριστώ θερμά για την καθοδήγηση και τη στήριξή της, καθώς και για την ευκαιρία που μου έδωσε να εργαστώ σε ένα τόσο καινοτόμο αντικείμενο.

Κυρίως όμως, οφείλω ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή μου Dr Αντώνη Κοκόση, Καθηγητή ΕΜΠ, που με εμπιστεύτηκε και μου έδωσε την ευκαιρία να εκπονήσω τη διπλωματική εργασία μου σε συνεργασία με το Imperial College, όπως επίσης και για την πολύτιμη βοήθεια και καθοδήγησή του, χωρίς την οποία δεν θα μπορούσα να ανταποκριθώ στις απαιτήσεις της παρούσας εργασίας.

Θα ήθελα ακόμη να ευχαριστήσω τον υποψήφιο διδάκτορα του Imperial College Sachio Steven για την σημαντική του βοήθεια στο κομμάτι της επεξεργασίας του μοντέλου, καθώς και τον Μεξή Κωσταντίνο, Ηλεκτρολόγο Μηχανικό ΕΜΠ, για την κατασκευή των νευρωνικών μοντέλων.

Τέλος θα ήθελα να ευχαριστήσω για την στήριξή τους, τους φίλους μου καθώς και όλη την ομάδα των μεταπτυχιακών φοιτητών του Imperial College στην οποία είχα την μοναδική ευκαιρία να συμμετέχω και τους γονείς μου για την υποστήριξη και την ευκαιρία που μου δίνουν να κυνηγήσω τα όνειρά μου.

Abstract

The rising popularity of monoclonal antibodies used as therapeutic proteins has created a demand for agile, robust bioprocesses. However, the biological and modelling challenges presented in upstream and downstream processes, as well as the interplay between them demand an extensive knowledge of the processes that can only be achieved through a quality by design approach. More specifically, downstream purification processes involved in monoclonal antibody biomanufacturing present many challenges in modeling, optimization, and control, due to the increased complexity of the phenomena describing them. In this thesis, the development of computational tools is proposed as a solution to those challenges, for a specific twin-column countercurrent solvent gradient purification ion-exchange chromatography process, in order to overcome the complexity and time consumption of the processes, as well as analyze and optimize over the degrees of freedom. A data-based approach is deployed to simplify the complex model into generic data-driven models that reduce complexity with excellent accuracy and capabilities to reproduce the physical process. Through a series of model simulations and sensitivity analyses, a design space for the multicolumn countercurrent solvent gradient purification ion-exchange chromatography is identified and a data-driven model of the process is created. The dissertation outlines the motivation and the incentives, presents the proposed methodology, and explains the application of the approach with a detailed discussion of the results and their significance.

Περίληψη

Η αυξανόμενη δημοτικότητα των μονοκλωνικών αντισωμάτων που χρησιμοποιούνται ως θεραπευτικές πρωτεΐνες έχει δημιουργήσει μια ζήτηση για ευέλικτες και σταθερές βιοδιεργασίες. Ωστόσο, οι βιολογικές προκλήσεις και οι προκλήσεις μοντελοποίησης που παρουσιάζονται στις διεργασίες παραγωγής και επεξεργασίας των αντισωμάτων, καθώς και η αλληλεπίδραση μεταξύ τους, απαιτούν εκτεταμένη γνώση των διαδικασιών, η οποία μπορεί να αποκτηθεί μόνο μέσω μιας ποιοτικής προσέγγισης σχεδιασμού. Πιο συγκεκριμένα, οι διαδικασίες καθαρισμού που εμπλέκονται στη βιοπαραγωγή μονοκλωνικών αντισωμάτων παρουσιάζουν πολλές προκλήσεις στη μοντελοποίηση, τη βελτιστοποίηση και τον έλεγχο, λόγω της αυξημένης πολυπλοκότητας των φαινομένων που τις διέπουν. Σε αυτή τη διπλωματική εργασία, προτείνεται η ανάπτυξη υπολογιστικών εργαλείων ως μέσο αντιμετώπισης των προκλήσεων αυτών, για να ξεπεραστεί η πολυπλοκότητα και ο απαιτούμενος υπολογιστικός χρόνος μιας χρωματογραφίας ιοντοανταλλαγής δύο στηλών, καθώς και να επιτευχθεί η ανάλυση και βελτιστοποίηση της διεργασίας με βάση τους βαθμούς ελευθερίας της. Χρησιμοποιείται μια προσέγγιση βάσει δεδομένων για την απλοποίηση του σύνθετου μοντέλου σε γενικά μοντέλα βάσει δεδομένων, που μειώνουν την πολυπλοκότητα του συστήματος, διατηρώντας εξαιρετική ακρίβεια, και επιδεικνύουν δυνατότητες αναπαραγωγής της φυσικής διεργασίας. Μέσω μιας σειράς προσομοιώσεων και αναλύσεων ευαισθησίας, καταστρώνεται ένας «χώρος» βέλτιστων συνθηκών λειτουργίας για την χρωματογραφία ιοντοανταλλαγής που χρησιμοποιείται για τον καθαρισμό των μονοκλωνικών αντισωμάτων και κατασκευάζεται ένα μοντέλο βάσει δεδομένων της διαδικασίας με τη χρήση μηχανικής μάθησης. Η διπλωματική αυτή εργασία περιγράφει τα κίνητρα που οδήγησαν σε αυτή την προσέγγιση, παρουσιάζει την προτεινόμενη μεθοδολογία και εξηγεί την εφαρμογή της προσέγγισης, με μια λεπτομερή συζήτηση των αποτελεσμάτων και της σημασίας τους.

Contents

List of Figures

List of Tables

Nomenclature

mAbs	Monoclonal Antibodies
UPS	Upstream process
DSP	Downstream process
CHO	Chinese Hamster Ovary
MCSGP	Multi-column Counter-current Solvent Gradient Purification
proA	Protein A Affinity Chromatography
CEX	Cation Exchange Chromatography
AEX	Anion Exchange Chromatography
HIC	Hydrophobic Interaction chromatography
FDA	Food and Drugs Administration
QbC	Quality by Control
QbD	Quality by Design
CPPs	Critical Process Parameters
CQA	Critical Quality Attributes
HCP	Host Cell Proteins
PAT	Process Analytical Technology
DoE	Design of Experiments
CSS	Cyclic Steady State
LSA	Local Sensitivity Analysis
GSA	Global Sensitivity Analysis
ANN	Artificial Neural Network
FAST	Fourier Amplitude Sensitivity Test
RS-HDMR	Random Sampling - High Dimensional Model Representation

1. Introduction

Monoclonal antibodies (mAbs) constitute one of the most rapidly increasing fields in pharmaceutical industry. MAbs are "Y" shaped molecules, consisting of two light and two heavy chains connected by disulfide bonds. They are able to identify or induce a neutralizing immune response when they identify foreign bodies such as bacteria, viruses, or tumour cells. Owing to their specificity, mAbs have a broad use in tumoral therapy and diagnosis, for example anti-cancer applications [1]. MAbs used against cancer cells are modified to deliver a toxin, radioisotope, cytokine or other active conjugate [2]. Furthermore, they are used for the treatment of autoimmune diseases, like Crohn's disease and sexually transmitted infections (STIs) [3]. It is a rapidly evolving class of drugs, that many scientists are working on, by targeting the creation of new therapeutic molecules, as well as the improvement of the existing ones [4].

A significant disadvantage however of monoclonal antibodies is their production cost coupled with the problems of mass production. To be effective they must be administered in large quantities, while their production requires huge amounts of cultures from mammalian cells followed by expensive techniques, and specialized facilities [5]. Despite that, they are among the top-ranked products in the on-going increasing market of high value biologics. In fact, in comparison to other biotechnological drugs, their sales are considered to build twice as quick. As per market projections, the worth of the mAb market is anticipated to increase up to 75 billion USD by 2025 [6]. Their demand relies profoundly upon the patient populace and their expense can rise up to $35000 P/A per patient for mAbs treating cancer conditions [7].

Additionally, their demand in the past year has skyrocketed, since several neutralizing mAbs have proven to be effective in dealing with the COVID-19 (SARS-CoV-2) pandemic [8]. Amidst the current COVID-19 pandemic, an assortment of prophylactic and therapeutic treatments is being created or repurposed to battle COVID-19. In fact, quite a few neutralizing monoclonal antibodies (mAbs) have passed the developmental state and are currently under assessment in clinical trials. They are considered to be one of the most efficient and innovate ways to guarantee passive immunization from the infectious disease. Furthermore, scientists are currently trying to develop treatments based on specific mAbs to block and/or neutralize SARS-CoV-2 in infected patients. Some of these treatments have been already administered in hospitals whereas others are still being evaluated [9].

In light of these recent events, the optimization of the mAbs manufacturing process is of utmost importance, in order to increase the production and reduce the high costs presented. However, experimental optimization of the process, especially in the case of downstream processes, proves to be extremely expensive and challenging, which leads to the search of alternative methods [10]. It is for that reason that in the past years the focus has shifted to the development of computational tools to help optimize and control the mAbs manufacturing process.

2. Literature Review

2.1. Monoclonal Antibody Biomanufacturing

Due to the ever-increasing demand of monoclonal antibodies, it is most essential to reduce their production costs. For that matter, emphasis has been given to the optimisation of both upstream and downstream processes. The upstream process (USP) involves all the steps included in the production of monoclonal antibodies in cell culture systems. Consequently, the downstream process (DSP) focuses mainly on the purification of the product collected from the bioreactor and the removal of any impurities from the final product. Both processes are to be investigated in the following section and for that reason a schematic approach of a typical mAb manufacturing is presented:

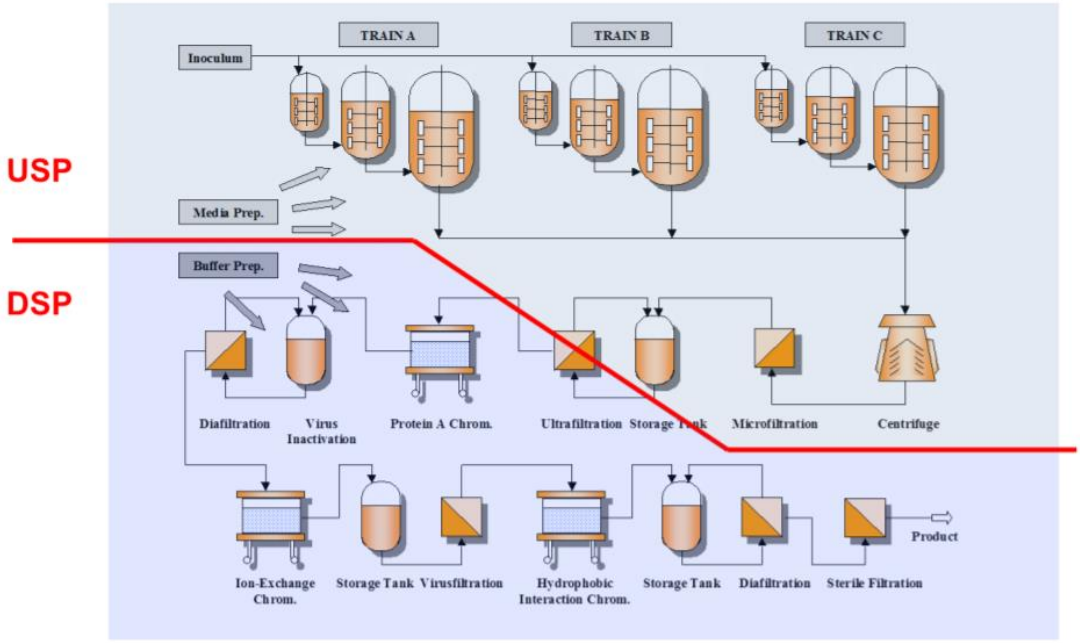

Figure 1: Schematic manufacturing process of monoclonal antibodies from cell culture, adapted from Sommerfeld and Strube [11]

2.1.1. Upstream Process

Upstream processing refers to the first step of mAb production, in which biomolecules are grown in bioreactors. Mostly utilised for their production are mammalian cell lines. Using batch and fed batch cultures, monoclonal antibodies reach a desired density and are then harvested and in order to enter the downstream section of the bioprocess [12]. The

antibodies are collected from the bioreactor using industrial continuous centrifuges and are clarified by using depth and membrane filters [13].

A diagram of a typical upstream process is presented below [13]:

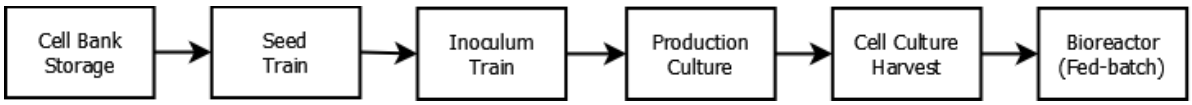

Figure 2: A typical upstream process

For the optimisation of the upstream process the main decision-factors to be considered are [14]:

- Cell lines able to synthesize the required antibodies quickly, at high productivities and low cost
- Culture media and bioreactor culture conditions that ensure the desired productivity and also meet product quality requirements. This includes type of reactor, medium conditions and operation type
- Continuous process monitoring via appropriate sensors
- Excessive understanding of the culture and its performance at different scales, so smooth scale-up is achieved

A summary of the parameters that should be consider when modelling upstream process is presented in the following graph [15]:

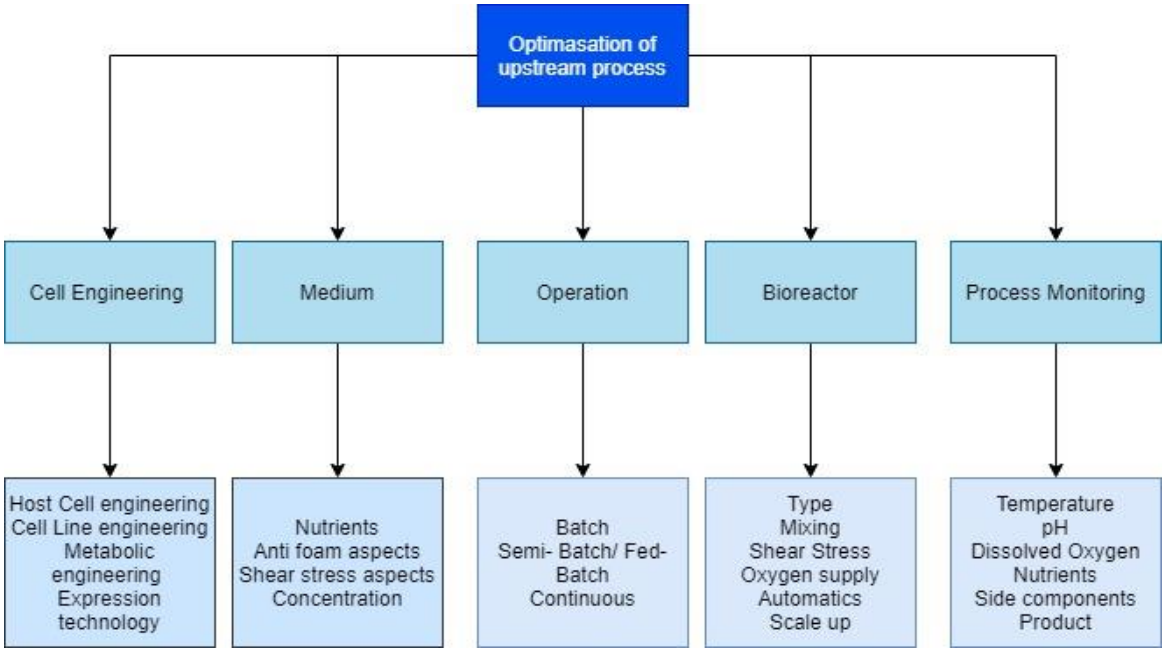

Figure 3: Summary of upstream process' optimization parameters

Upstream modelling and optimisation depend typically on biological limits, like the nature of the cell line or medium used. Thus, higher titers in upstream processes do not require other equipment and they can still occur in the same reactor set-ups as lower titers. Henceforth, upstream processes can be altered without an increase in costs since the processed volumes are the same, but the amount of antibody is increased. However, the limiting factor is downstream processing, where the products of upstream process enter after they created. Downstream processing has its physical limits, and it is intended for much lower quantities of antibody than those that are entered. In other words, upstream capacity can be increased without any further economic burden, whereas downstream capacity, due to its physical limitations for purifying, scales up linearly with the expenses [16].

Initially, the most important step is to choose the most productive cell line. For recombinant proteins, like antibodies, mammalian cells are mostly utilized since they can yield protein folding and post-translational modifications similar to human ones [14]. However, the usage of yeast systems, like Pichia Pastoris, is under investigation. Yeast systems have not only faster growth rates than mammalian cells, but they are also cheaper and simpler systems, due to their lack of complex biological routes. Nevertheless, microorganisms do not have the required physicochemical and biological characteristics needed for the appropriate post-translational mechanisms required for mAbs [17]. Moreover, modified plants have been searched as potential candidates to host mAbs, since plants are easy to cultivate, require low-cost medium and maintenance and achieve relatively high production yields cell cultures. However, there are some limitations in their utilization for mAbs manufacturing, since plants lack in post-translational mechanisms, such as glycosylation. As already stressed glycosylation is the most crucial biological aspect to consider when picking a proper candidate for the manufacturing of monoclonal antibodies [18].

Apart from all the potential cells for mAb production mentioned above, the most broadly used mammalian cell is Chinese hamster ovaries (CHO) [19]. Mammalian cells require complex culture medium, as they are in need for many nutrients including many trace elements. The use of serum provides the cells with those essential nutrients [1]. However, animal sourced ingredients tend to be avoided in the manufacture of mAbs due to concerns about bovine spongiform encephalitis (BSE).

The main reactors used for mAbs manufacturing, each with their own advantages and disadvantages, are [12]:

- Membrane reactors
- Stirred Tank reactors
- Fluidised beds
- Airlift reactors

The operations used for mAbs production are batch, semi-batch (or fed-batched), continuous and perfusion. Batch systems are already prepared with a medium containing all the nutrients

required. The main drawback of this method is that it is hard to keep the initial conditions and waste metabolites are accumulated in the medium, resulting to lower cell density and productivity [20]. The second alternative is continuous systems, but the quantity of medium used increases and the recovery is harder, and the same goes for perfusion systems. The most popular operation is fed-batch systems, where fresh medium is fed, and it is easier to control the medium conditions. As a result, higher titers can be achieved [16].

2.1.2. Downstream Process

Downstream process revolves around the purification of the final product after its production in the bioreactor. This series of processes, aiming to the recovery and purification of mAbs from the cell culture media, is one of the most essential parts of the whole system of manufacturing antibodies. Moreover, downstream process has a strong impact on the economics and, thus, optimisation is vital to decrease costs [21].

Undoubtedly the main purpose of downstream process is to achieve a high level of purity, regarding the quality by designed that has been devised beforehand. In addition to that, speed is another factor that should be considered, bearing in mind that all these processes take place before an antibody is released for clinical trials and then released for commercial use. In the light of these, some approaches have been made to use universal platforms for purification and continuous operations instead of the well-established batch processes. More information on these breakthroughs is given in section 2.3.4. below [22].

As it has been already deliberated, downstream processing has more modelling challenges and is harder to be improved due to its physical limitation. To enumerate, a scale-up is usually followed by a similar scale-up in costs. Henceforth, downstream processing is less eager to alterations [21].

A flowchart describing the downstream process is presented here [23]:

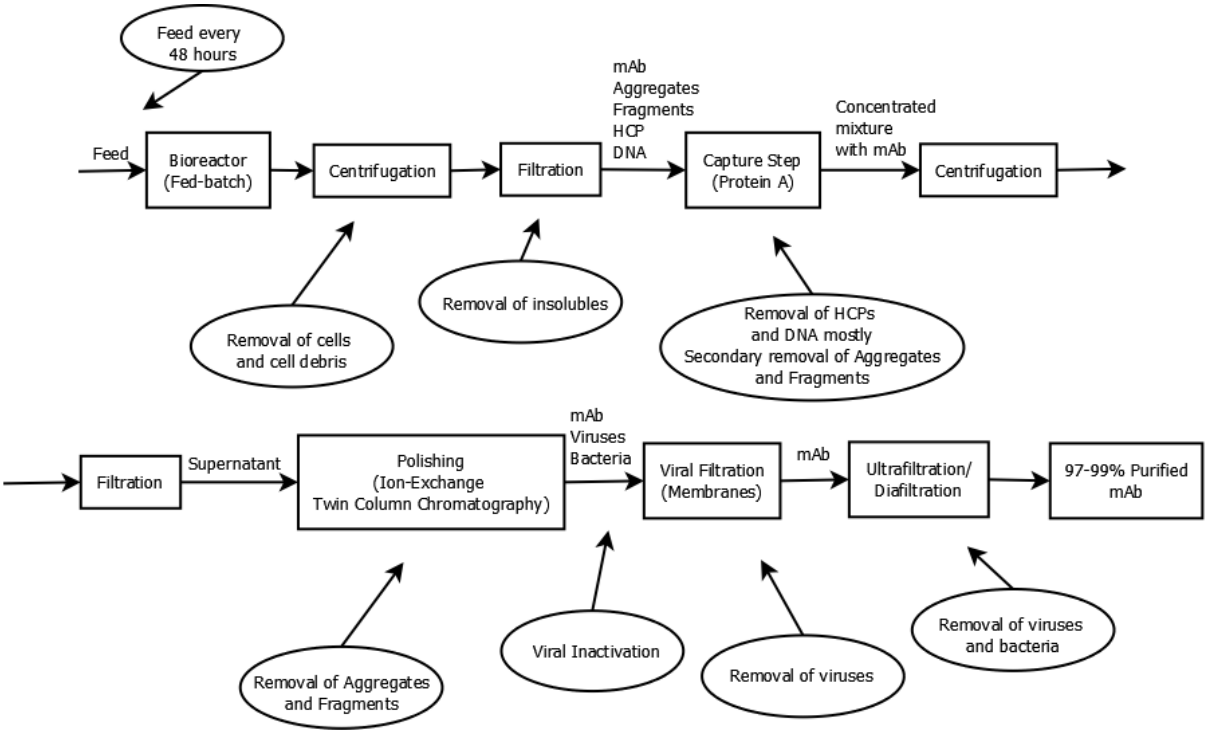

Figure 4: Flowchart of downstream processing

First, antibodies should be separated from the culture broth. This broth has namely high cell density, in order for costs to be reduced in the upstream process [24]. A mean concentration of solids in the culture broth is 40% to 50%, but after the clarification steps solids should be negligible. For that reason, centrifugation, depth filtration and sterile filtration are utilised, with the most predominant practice being the centrifugation. The reason for that is centrifugation's ability to scale up easily in industrial scale and work with large volumes (typically between 2-15,000L per batch) [22]. This separation step, although it seems trivial, is vital and it can rise up to one quarter of the total costs of the whole downstream process [25].

After centrifugation, the mAbs must be captured from the liquid. The most popular approach is using protein A (proA) affinity chromatography [26]. ProA is a polypeptide that is naturally fastened in the membrane of the bacteria Staphylococcus aureus [27]. This chromatography method is the most preferable because it can achieve up to 99% percent purity of the desired product in a single step [22].

Despite the advantages of mammalian cells in production of complex protein structures where post translational modifications are indispensable, endogenous retroviruses can be produced by them. As a result, a viral inactivation step in acidic environment is much needed. During proA chromatography the elusion takes place either way in low pH values, so the existing conditions are exploited to reduce costs.

Even after the affinity chromatography, trace impurities like leached proA, HCP or endotoxins exist. Henceforth, polishing steps such as cation-exchange chromatography (CEX), anion-exchange chromatography (AEX) or hydrophobic interaction chromatography (HIC) are utilised [22]. Undoubtedly the polishing step used is affected by the mAb present and its impurities. However, the existence of at least one ion-exchange chromatography step is observed in most mAb purification processes, since it allows the removal of high molecular weight impurities [24]. The optimum pH for these methods is as further as possible from the antibodies' pI, so that they are charged. However, mAbs have usually high pI values and, thus, AEX columns are preferred. The reason for that is that even in mild pH conditions mAbs are positively charged and consequently easily eluted from the AEX matrix, while negatively charged impurities are anchored [28].

Finally, the purification process is finalised with a viral clearance and an ultrafiltration. The ultrafiltration's parameters, like the imposed pressure across the two sides of the membrane, the type of the membrane used, the flow velocity, and the concentration, can be modelled easily and generalised for all the mAbs [24].

2.1.3. Biological Challenges

The most important biological challenge that occurs in antibodies revolves around the impurities involved. Since it is already stressed that production can hit a peak without overfloating costs too, the real trial lies on the purification.

Both upstream and downstream processes affect the product purity. In the case of the former, the cell line selected, the fermentation process or the culture time and conditions are vital for the impurities content of the broth. Regarding the latter, the methods used on the downstream process can affect both the selectivity and the yield of purification and the impurities or not of the final product. This goes vice versa too, as the impurities nature and content determine the downstream [29].

2.1.4. Modelling Challenges

With regards to bioprocesses, computational techniques have turned into a valuable tool due to their calculation power and their capability to shorten the processing time, when contrasted with the costly and monotonous experiments. The significance of mechanistic modelling lies in its ability to tackle challenges like the integration of process analytical tools, the frequency in which data are obtained and the real-time data measurement and control. Nevertheless, new issues can emerge. The dynamic profile of the cell cultures is susceptible

by multiscale changes all through the bioprocess. This complex nature represents an obstacle to the bioprocess monitoring approaches. Under that premise, the potential to monitor the non-linear profile of the microorganisms and the multiscale interaction within the chromatographic steps by non-linear ordinary differential equations has been arisen [30]. However, in this case the complexity of the system is greatly increased due to the non-linear equations compared to the ordinary differential counterparts. More specifically, the number of parameters constituting the mathematical expression of the system is significantly increased due to the fact that the variables are as a function of multi scales (i.e., time and space). In this manner, approaches such as discretization or other techniques of transformation from the partial to ordinary differential equations ought to be taken into consideration before the application of such mechanistic model [31].

2.1.4.1. Process control and digital twins

After the process model has been developed, arises the issue of process control. In order to move towards a quality by control (QbC) environment in biopharmaceutical manufacturing, there are many significant challenges regarding process data and models that need to be resolved. The principal challenge is to accomplish a better understanding of different Critical Quality Attributes (CQAs) and the causes of their variation, as well as the ability to measure them in real-time. The development of comprehensive predictive dynamic models can be achieved through advancement in process modeling. Models utilized for process control ought to be sufficiently precise and avoid irreversible negative effects through inappropriate control actions, but simple enough for quick optimization of control actions. The heterogeneity of the collected data in both time and type is another identified challenge. In time scale it is collected at different sampling intervals, whereas in type, for instance, at-line, and in-line collected data [32].

A valuable tool when tackling the challenges of process control appear to be the digital twins. Digital twins are models using real-time data along with machine learning techniques, in order to predict the results of a process. They can be used both for producing data prior to the process, so as to optimize the process, and for on-line control of the process, due to their decreased complexity and time demand [33].

2.1.4.2. Adaptive design space and integrated process design

Any adjustment in process operation or design leads to variation in performance and CQAs that can propagate all through the drug substance production chain. The experimental effort needed to identify the impact and optimal levels of each component in a bioprocess could be

exhaustive. However, sophisticated statistical methods can then be utilized to detect compounds that could individually have negligible impacts but can have strong impacts in combination with others.

Considering the effect of upstream process conditions on downstream performance, process optimization ought to be done on the overall process assessment [34]. The first attempt for holistic risk and quantification for impurity propagation across multiple unit operations using an integrated process model was presented by Zahel et al. [35]. The risk for an out of specification final drug substance of a biopharmaceutical product was measured and the model was utilized to pinpoint potential changes in process parameters for risk reduction.

2.2. Quality by Design

Quality by design is a systematic approach towards the development of drug products and processes using rational and scientific based methodology. The aim of these approaches is the manufacture of final products with the desired quality standards. The term “quality by design” or QbD for short has been introduced in the 1970s by J.M. Juran, an American Engineer, and as a practice has been widely used in automotive and aviation industries, where high end products are manufactured [36]. These methods of QbD have been used in pharmaceutics since the 1990s to ensure better end products and the highest possible profit for the companies [36]. To enumerate, QbD relies on careful planning and not on luck to achieve high-quality product and lessen shortcomings [37].

Hence two are the main recipients of QbD, patients and pharma industries. On the one hand the production of drugs with the minimum number of defects and implications is at the most beneficial for patients, which are the main addressees of QbD. On the other hand, the benefits for companies are not to be overlooked. Indisputably, QbD offers a better understanding of both the nature of the product and the processes critical parameters. As a result, less batch failures occur, the scale-up is more feasible and efficient than before and there is a better return of investment. All this contribute to cost savings and, thus, a more lucrative opportunity for the pharmaceutical company [38].

Figure 5 recapitulates the main doctrines of QbD:

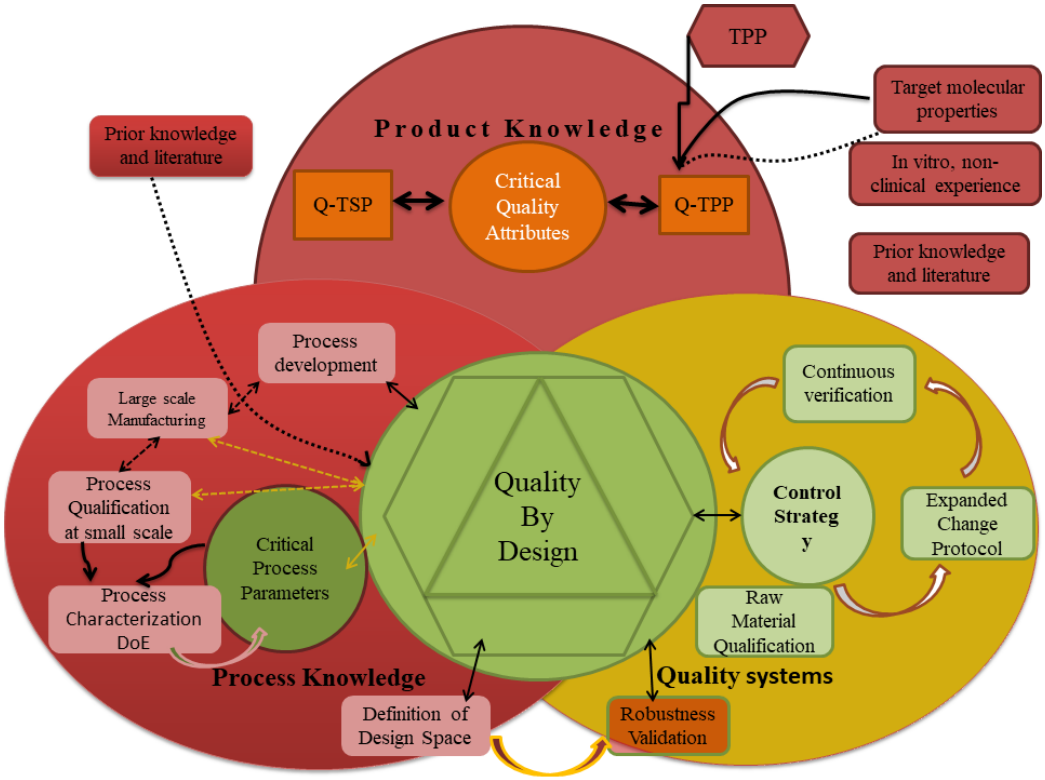

Figure 5: Description of the Quality by Design (QbD) approach

The main steps of QbD are critical quality attributes, process parameters, process controls and product specifications/ control strategy. These categories are presented in the figure below and are further analyzed in the upcoming sections.

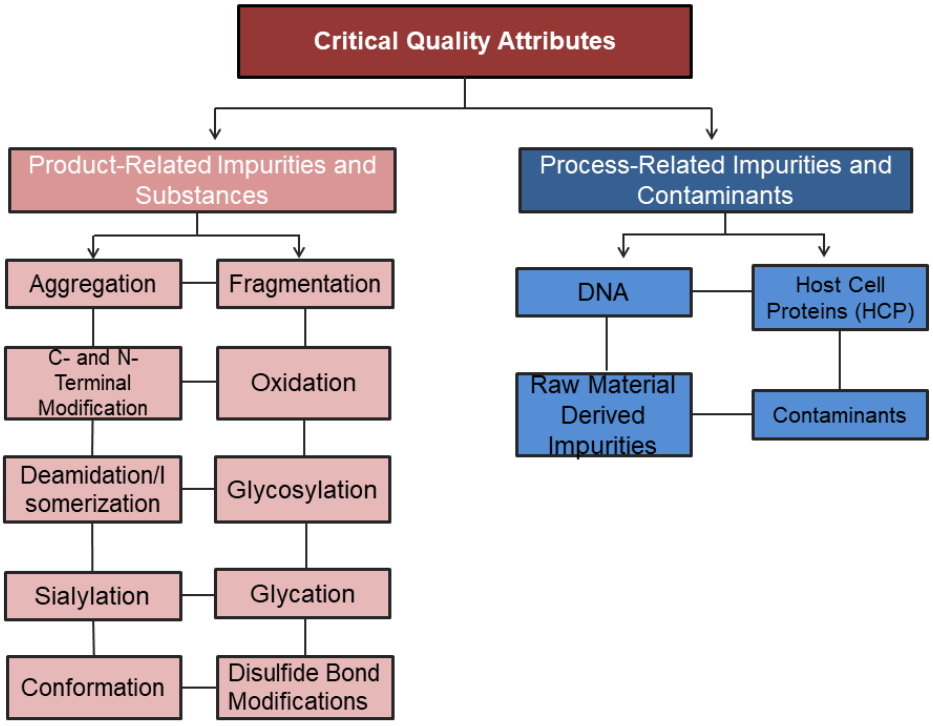

Figure 6: Presentation of Critical Quality Attributes

2.3.1. Critical Quality Attributes

CQAs are physical, chemical, biological, or microbiological properties or characteristics that should be within an appropriate limit, range, or distribution to ensure the desired product quality [39]. They are selected through a risk-based analysis, during which an evaluation of the product quality attributes takes place, in order to assess their potential impact on patient safety and product efficacy.

Proteins are often subjected to a variety of post-translational modifications, which can be linked to diseases such as inflammatory and autoimmune disorders. However, only a relatively limited number are commonly encountered in biopharmaceutical proteins. The quality attributes that can be generally observed in recombinant therapeutic proteins can be divided into two categories: the product-related impurities and substances and the process-related impurities and contaminants [40].

2.3.1.1. Product-Related Impurities and Substances

Product-related impurities are molecular variants created during manufacture and storage that have properties different from those of the desired product with respect to activity, efficacy, and safety [41].

Aggregation

Proteins are marginally stable given their three-dimensional structure. In addition, they are particularly prone to forming aggregates when they are in stages where they have not yet folded or partially folded. It is therefore possible at these stages to form stable aggregates, which are a major problem in the production of therapeutic molecules and are responsible for inducing immune responses in patients [42].

Protein aggregation is one of the major challenges faced during the development protein-based drug products as it is a common source of protein instability [43]. Protein aggregation is observed in all stages of protein product development and manufacturing processes, as well as during storage and handling.

There are various reasons that can cause protein aggregation and each one can result in different properties for the product, according to the type of physical or chemical interaction. As such, the aggregates can have different solubilities and sizes from the initial product.

Additionally, they may be covalently or noncovalently associated and the interaction may be reversible or irreversible [44].

Factors that can result in protein aggregation are primary and secondary structure, the presence of hydrophobic patches on the surface of the molecule, temperature, buffer type, pH and ionic strength, protein concentration, physical stresses such as shaking or stirring, shearing, freezing, thawing, and refolding [40].

The various ways of describing the aggregation of therapeutic molecules that have been proposed are summarized in five main mechanisms [45]:

- Reversible association of the native monomer, where the surface of the native protein monomer is self-complementary, meaning that it will form reversible small oligomers by self-associating.
- Aggregation of conformationally-altered monomer, where aggregation will be promoted by triggering an initial conformational change of the protein using stresses such as heat or shear.
- Aggregation of chemically modified product, where the difference in covalent structure causes a change in protein conformation, which in turn leads to the protein aggregation that precedes aggregation.
- Nucleation-controlled aggregation, where an area on the surface of the monomer can form a nucleus to initiate aggregation with other monomers.
- Surface-induced aggregation, where patches on the surface of the molecules lead to the formation of stable aggregates.

Protein aggregation can have a significantly negative impact on the quality of the product, as it may cause reduced biological activity, which in turn minimizes the product's efficacy [43]. Additionally, there is a chance they may induce immune responses, such as neutralizing antibodies that limit efficacy or causing severe immediate hypersensitivity responses (e.g., anaphylaxis). Consequently, the presence of any insoluble aggregates is unacceptable when it comes to therapeutical proteins [46].

Fragmentation

Despite the remarkable stability of the peptide bond, its enzymatic or chemical disruption leads to protein fragments. The resulting proteins differ according to how much and where cleavage occurs [40]. In the case of monoclonal antibodies, fragmentation is a CQA that needs to be monitored, since it gravelly affects the product purity and integrity. It can occur during protein production in the cell culture, which is then modulated by the purification process. However, chances are it will reoccur during storage or circulation in the blood [47].

Fragmentation rates of monoclonal antibodies are affected by many factors, such as pH, temperature, solvent composition and the presence of metals or radicals [47]. As a result, each case is unique, and the impact of fragmentation is product dependent. The most common issues that can arise include loss of biological activity, reduction of half-life and decreased immunogenicity caused by the generation of novel epitopes.

The effect of fragmentation on the function of monoclonal antibodies can vary depending on the cleavage site. Fragmentation in the complementarity-determining regions (CDRs) is likely to have an effect on the monoclonal antibody's ability to bind to the target and, consequently, have an effect on its potency. Fragmentation in the hinge region may have more implications on the function of a monoclonal antibody molecule, such as reduced circulation half-time or complete loss of biological activity [47].

C- and N- Terminal Modification

One of the most common modifications of recombinant monoclonal antibodies is C-terminal lysine processing of the heavy chain. Carboxypeptidase activity during mammalian cell culture can cause truncation of the C-terminal lysine of antibody heavy chains, which usually leads to a heavily or even completely truncated product [48]. This modification can be important since it is sensitive to the production process [49].

Cyclization of glutamine of IgG antibody heavy chain can result to pyroglutamate, due to instability of N-terminal. This can also occur with Lambda light chains can start with a glutamine residue [48]. However, any impact of pyroglutamate on the safety and efficacy of mAbs is yet to be proven [40].

Oxidation

Protein oxidation is a covalent modification of an amino acid side chain, or a protein backbone caused by two types of reactions. More specifically, it can be caused through direct reactions with reactive oxygen species (ROS), as well as through indirect reactions with secondary by-products of oxidative stress. It can result in protein fragmentation or protein-protein cross-linkages [50]. Although the modification by ROS reaction is possible for all amino acids, the predominant reaction is the conversion of methionine (Met) residues to methionine sulfoxide, due to the high reaction susceptibility presented in the sulfur group in this amino acid.

Oxidative modifications of proteins can lead to changes in their physical and chemical properties, such as conformation, structure, solubility, susceptibility to proteolysis, and

enzyme activities. More specifically, for some Met residues, oxidation has an impact on conformation, stability, and biological activity, whereas for other Met residues, oxidation has little or no impact [40]. The possible negative impacts of oxidation result in loss of biological activity and can affect immunogenicity, often through the induction of aggregation. Although it is yet unclear at what stage these residues are oxidized, under most circumstances, oxidation can be readily detected in monoclonal antibodies after long-term storage, incubation at elevated temperatures or incubation with oxidizing reagents [48].

Deamidation/Isomerization

Another common post-translational protein modification is the nonenzymatic deamidation and isomerization of asparagine (Asn) residues. In the case of recombinant monoclonal antibodies, it can occur during at any stage of the process, e.g., during cell culture, protein purification and storage [48]. However, its impact on biological activity and immunogenicity is product dependent [40].

Both deamidation and isomerization occur naturally and in a constant rate in the bloodstream and can be increased with a rise in pH or temperature. It is therefore expected for therapeutic proteins to undergo deamidation after being administrated in the bloodstream, especially for those who have a long circulatory half-life and are consequently more likely to be affected [40].

Glycosylation

The post-translational modification by which oligosaccharide structures covalently bind to the polypeptide backbone of a protein is called protein glycosylation [51]. It is most likely induced during oligosaccharide synthesis and processing, where the oligosaccharide heterogeneity can cause monoclonal antibody heterogeneity [48].

According to the type of link created, protein glycosylation can be categorized into two types: serine/threonine-linked glycosylation (O-linked) and asparagine-linked glycosylation (N-linked) [51]. However, between the two, O-linked glycosylation is rarely presented in mAbs, and therefore, when studying therapeutic proteins, the focus falls on N-linked glycosylation. This type of glycosylation begins in the endoplasmic reticulum and is completed when the protein reaches the Golgi apparatus [52].

The two major levels of glycoprotein heterogeneity associated with glycosylation are macroheterogeneity and microheterogeneity. The former begins in the endoplasmic reticulum and is triggered by the presence or absence of specific sequons in the structure of

the oligosaccharide. The latter takes place in the Golgi apparatus, where different levels of processing occur [51].

The effects of glycosylation vary and can greatly affect the biological activity of the protein. The protein's resistance to proteolysis, as well as its tendency to aggregate in vitro are increased by the invariant glycan at N297 [53] in the Fc domain [54]. Additionally, studies of aglycosylated mutants (containing N297Q or N297H mutations) of IgG1 have shown that their sensitivity to digestion by pepsin, trypsin, chymotrypsin, and pronase was greater than the one of their glycosylated counterparts [55]. The only exception to that was papain, which was less effective at digesting the aglycosylated antibody. Glycosylation also affects thermal stability, given that protein's protection to proteolysis is subject to the presence of certain glycan patterns [54] and that the removal of glycosylation results in thermal stability loss of both full-length antibodies and Fc fragments [56].

During the culture of mammalian cells, a great a variety of antibody glycoforms is encountered, that are very difficult to separate. Consequently, most clinical antibodies comprise of a population of glycoforms whose relative abundance in the therapeutic preparation dictates their effector function [57]. This high degree of heterogeneity is determined by the cell line, the manufacturing process and the cell culture conditions and need to be taken into account when evaluating the impact of glycosylation on the safety and efficacy of a particular therapeutic protein [40].

Sialylation

The ultimate sugar residue of the oligosaccharide chain of N-linked glycans is sialic acid. Be that as it may, sialic acid site occupancy and composition is specific for each protein and species. This means that 60–95% of glycoproteins and about 10% of antibodies in human serum are sialylated, whereas these numbers are reduced in recombinant proteins [58]. Sialyation can have a positive (e.g., in the case of rhIFN-β1a) [59] or negative (for EPO) [60] biological activity, which makes its role in effector functions of monoclonal antibodies unclear.

Both the importance of sialytion on circulatory half-life and the elimination of nonsialylated proteins from the circulatory system by the asialo-glycoprotein receptors in the liver are important aspects of sialytion [61]. An additional advantage of the lower clearance provided by sialyation is the improvement of the in vivo efficacy, as in the case of tetrasialylated EPO [60]. A great number of glycoengineered therapeutic proteins have been developed thanks to this effect [62].

Even though sialic acids have the tendency to cover antigenic determinants and, subsequently, potentially lessen immunogenicity [63], exceptionally sialylated proteins tend to produce weak immune responses [64].

There are two forms of sialic acids found in the cells of mammas that are usually utilized for the creation of recombinant proteins. These two are the N-acetyl- and N-glycolyl-neuraminic acid, however, just one (the N-acetyl structure) is normally found in adult human glycoproteins. Thus, in humans, N-glycolylneuraminic acid (NeuGc) is antigenic [65]. Furthermore, in CHO cell expressed proteins the NeuGc levels are usually low and therefore are not believed to be a be an immunogenicity concern [66], but higher levels may enhance antigenicity [67]. Moreover, the presence of terminal NeuGc on a recombinant protein is by all accounts corresponded with a quicker expulsion of the molecule from the bloodstream [68]. Finally, another factor for the efficacy of the protein is type of linkage between the terminal sialic acid and the preceding galactose residue [69].

Glycation

Glycation is the result of a nonenzymatic reaction between reducing sugars and the N-terminal primary amine or the amine group of lysine side chains [48]. It is observed when the protein is incubated in presence of reducing sugars. However, even though fructose and galactose have higher rates of glycation, it occurs mostly in the presence of glucose, which is the main carbon source used during cell culture of recombinant proteins [70].

Studies have been made to determine the impact of glycation on proteins, but its effects are not yet clear. It is shown that glycation may have an impact on the biological activity (increased or decreased) of a protein as well as its PK, but its impact on potency is expected to remain limited, since glycation sites are distributed at lysine residues over the entire molecule [71]. Finally, given that human plasma proteins are 10–20% glycated due to the presence of glucose in serum, it is not expected that glycation has an impact on immunogenicity and safety [40].

Conformation

Proteins are complex, three-dimensional structures that exist as an ensemble of different conformations in equilibrium instead of a single rigid structure [72]. This was first proven by Foote and Milstein, who studied the complex binding kinetics of hapten at various concentrations to three specific antibodies and proposed that monoclonal antibodies exist as more than one conformation, at least at the combining site, in equilibrium.

Changes in conformation can have a number of causes, including chemical modifications such as oxidation and deamidation, and can occur throughout the manufacturing process [73]. They can result in the loss of biological activity and the formation of new epitopes with immunogenic potential. Proteins that are particularly at risk of conformational changes during the refolding steps are those that are denatured as part of their manufacturing process such as inclusion bodies produced by microbial expression. This misfolding of therapeutic proteins can be responsible for enhanced immunogenicity of biopharmaceuticals and breaking of tolerance [74].

Disulfide Bond Modifications

IgG molecules are composed of two heavy chains and two light chains, all connected with disulfide bonds. Specifically, each light chain is connected to each heavy chain by one disulfide bond and each heavy chain is connected to the other heavy chain by two to four disulfide bonds depending on the subtypes of the antibodies. It is therefore quite clear that disulfide bond formation is important for the assembly and maintenance of the structural integrity of antibodies [48].

Heterogeneity related to disulfide bonds can be introduced at different stages, given that disulfide bonds may dissociate, leading to the generation of half molecules. Thus, protein folding might be affected, potentially resulting in a change in protein structure and function. A common alteration of disulfide bonds in IgG antibodies is the creation of trisulfide bonds, resulting from insertion of a sulfur atom within a disulfide bond. However, this particular change doesn't seem to affect biological activity [40], [75].

Under denaturing conditions, the presence of incomplete disulfide bonding can trigger disulfide bond scrambling, which can result in the formation of disulfide bond related fragments [76]. These resulting artifacts are different combinations of complete heavy chains and light chains linked by interchain disulfide bonds and constitute a serious problem for IgG4 because of the instability of the hinge region interchain disulfide bonds [77], [78].

2.3.1.2. Process-Related Impurities and Contaminants

Process-related impurities are derived from the materials used throughout manufacturing process, as well as from the host cell itself.

DNA

As mentioned above, the cell lines used for the production of therapeutic proteins have an infinite life span. This leads to a great concern for tumourigenic activity conferred by the DNA released during cell culture by the host cell. This can occur by direct integration of a DNA sequence capable of expressing activated oncogenes or by integration next to a dominant proto-oncogene thus activating its expression [40].

The levels of host cell DNA per dose that can be considered as acceptable, according to the Expert Committee of the World Health Organization, are up to 10 ng [79]. However, studies have shown that the actual risk of tumourigenic activity appears to be very small. More specifically, an experiment including the regular injection of much higher amounts of DNA than the recommended level in nonhuman primates did not result in any tumours during an evaluation period of 10 years, despite the fact that the DNA injected contained an activated oncogene from human tumour cells [80].

Host Cell Proteins (HCP)

During culture, in the occurrence of cell apoptosis, physical breakage of cells, or secretion, endogenous proteins of the host cell can be released. Since most therapeutic proteins are expressed in non-human mammalian cell lines such as CHO, NS0, or SP2/0, the immune system of human patients is expected to recognize endogenous proteins of these cells as foreign. However, until this point in time, only a fraction of clinical adverse events has been attributed to HCP impurities. In any case, since there is a possibility of allergic reaction to HCPs or a possible risk of adjuvant effects, it is considered necessary to minimize their levels.

Raw Material Derived Impurities

The utilization of raw materials in manufacturing is a well-known fact. However, they can sometimes co-elute with the product, raising concern for the product's safety, in the case that they are toxic in humans. These materials include:

- Chemically defined and nonchemically defined components of cell culture media
- Feed solutions and cell culture additives
- Components of purification buffers and solutions
- Excipients of the drug substance and the drug product

- Toxic components of process consumables or equipment in contact with the product (such as chromatography resins, ultrafiltration membranes, filters, containers, tubing, O-rings for instance)

Additionally, the leaching of toxic components or release of components from drug substance and drug product containers and container closure systems can occur and can cause product alteration. Therefore, raw materials and their components should be assessed for the probability of negatively impacting the health of patients through a risk assessment [40].

Contaminants

The possibility of contaminants of recombinant protein therapeutics, the production of which takes place in the cell expression systems of mammals, being unintentionally introduced at any time all through the manufacturing process is known and can be caused by the introduction of contaminated raw materials, the utilization of defective equipment or the operation under nonaseptic conditions. Some common contaminants are: viruses, living organisms, bacteria, fungi, mycoplasma, as well as their toxic by-products (for instance endotoxins). Contaminants can have harmful consequences for patient wellbeing. Furthermore, endotoxins, which are a significant component of the outer cell wall of bacteria that are Gram negative, have high levels of toxicity, which can cause inflammations or even sepsis, with the latter being usually lethal [81]. There two types of viruses based on their origins: exogenous or endogenous. Recombinant therapeutic proteins commonly expressed by mammalian cell lines often include endogenous retroviruses. These retroviruses can either be infectious as in the case of NS0 cells or defective and noninfectious as in the case of CHO cells [82]. Therefore, validation studies are of the outmost importance in order to ensure sufficient clearance of endogenous viruses through the purification process [83].

2.3.2. Critical Process Parameters

Critical process parameters (CPPs) are variables that have an effect on critical quality attributes (CQAs), that have been investigated before. Thus, they should be calculated and manipulated in order for the desired quality of the final product to be achieved. Without a shadow of doubt, defining the CPPs requires a great understanding of both how the processes work and how they impact the CQAs [84].

Another term used when studying the CPPs is the design space. As has been defined by the International Conference on Harmonisation (2008) design space is the multidimensional combination and interaction of process parameters and input variables that have been

demonstrated to provide assurance of quality [85]. Hence, a deep understanding of the nature of the process and how each variable affects the CQAs are perquisites to define the design space. A change within this design space has little to no effect on the final product, where a change outside the boundaries of this space leads to tremendous changes that may worsen its quality. Moreover, a design space can be limited to just an operation, a series of operations or the whole factory [37].

Both the terms CPPs and design space are used to determine the acceptable range of the processes' variables, without affecting the product quality. Of course, prior knowledge of the processes and the desired product in laboratory or pilot level are required in order for the CPPs to be defined and the design space to be created [86].

2.3.3. Control Strategy and Control System

Control strategy is a methodology which focuses on the creation of a control system. A control system is a series of tests and controls which stems from the deep knowledge of the process and the nature of the desired final products. Control strategy has two aspects: control system testing and process monitoring [37], [87].

Control system testing involves stability tests, tests about endotoxins and inhibitors, as well as tests for substances that may reduce the yield and the quality of the product. Also observed during control system testing are decomposition and secondary actions [37].

Process monitoring is about the process itself rather than its products. On the one hand, process is assured that works withing the limits of design space and thus it is kept in check. On the other hand, process monitoring offers a better understanding on the process' gimmicks. As a result, opportunities to enhance the productivity of the process are presented [37].

3. Process System Engineering Background and Motivation

3.1. Process System Engineering Advances in Bioprocess Modelling

The requirement for a change in perspective in the biopharmaceutical industry for quality assurance has been recognized for quite some time. This was made possible, by the introduction of the process analytical technology (PAT) guide by the U.S. Food and Drug Administration (FDA) [88].

Traditionally, off-line and off-site laboratories have been used in order to perform the chemical analyses needed in downstream operations during manufacturing. Because of their remote location, aftereffects of these assays are regularly delayed (going from hours to days) and therefore cannot be used to control and monitor the downstream process in real time. Instead, they are simply relegated to a compilation of databases. It is for that reason that on-line control techniques have long been considered as credible opportunities to improve product quality and efficiency in pharmaceutical production by the FDA [89].

This PAT guide requires an enhanced process understanding that underlines a new quality by design (QbD) approach. In this design approach the mentioned item quality is guaranteed by the actual process and does not have to be tested afterward. Design of experiments (DoE), which is used in studies to gather process knowledge, is among the most prominent approaches to enable QbD [90]. In order to set up a DoE for a product, critical process parameters (CPP) which impact the products' critical quality attributes (CQA) must be defined [91]. To systematically investigate these CPP combinations and their multifactorial influence as well as to keep the number of experiments manageable, different designs can be applied. Such designs are usually evaluated by application of process modelling.

3.1.1. Process Modelling

The response surface methodology is the most well-known modelling technique, used in combination with DoE studies. In this technique, the experimental results of a design space are represented on a surface as responses of the CPPs. This, in turn, is utilized to discover the optima for the investigated conditions [92]. This particular technique is generally utilized in media development and optimization for production processes.

The models created that utilise this technique can be separated into two categories, descriptive and predictive models. As a matter of fact, descriptive models provide real-time information, that is, certain information for only up to the current time point of the process. Comparatively, predictive models can predict future values of state variables and, since the input data can be simulated for the future, give an educated guess about the trajectory [93].

Moreover, the expansion of the existing process knowledge can be achieved with the appliance of exploratory data analysis, for example, principal component analysis (PCA) and parallel factor analysis (PARAFAC). With these methods, hidden structures and latent variables, the so-called principal components (PC), in the investigated spectra can be determined [94]. In general, there are two diverse modelling approaches. The first one includes nonparametric (black-box) models, while the second one includes parametric (white-box) models. Nonparametric models are based on experimental data only and do not need any further process knowledge. Therefore, various regression techniques are available and commonly applied to develop nonparametric models [95]. Conversely, parametric models utilize empirical knowledge and first principles, that is, their structure is well defined and transparent [96]. Because of their respective model structures, both modelling approaches have separate distinctive advantages just as disadvantages and limitations. Finally, the combination of a nonparametric and a parametric model into a single semi-parametric model structure is defined as hybrid (grey-box) modelling and it permits the fuse of both process knowledge and data-driven information.

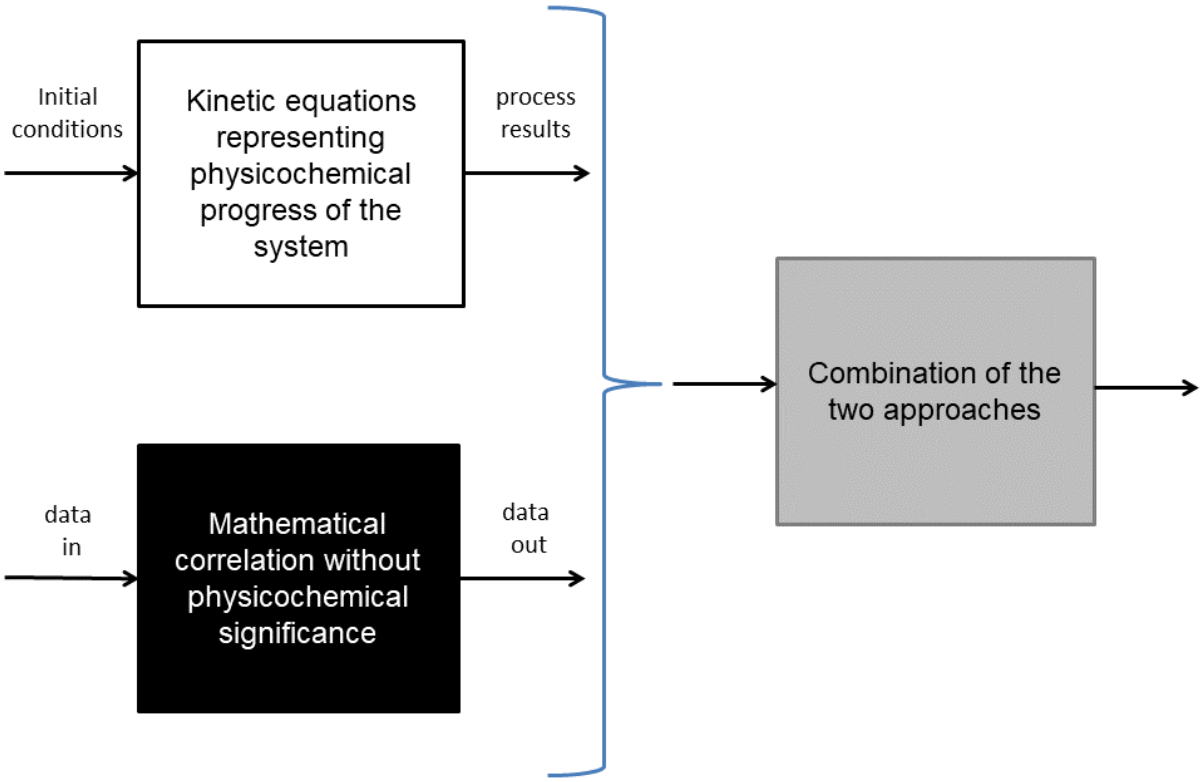

Figure 7: Schematic representation of white-box, black-box and grey-box models

In the remainder of this chapter an indicative but not exhaustive list of main contributions in the field of product and process modelling are discussed.

High Fidelity Modelling

In high fidelity models or mechanistic models, the process is described with a series of complex equations. These equations derive from the fundamental scientific laws. Physical and biochemical principles establish these model equations. However, a variety of experimental data are still required to regulate the model and determine some unidentified model parameters. Of course, dynamic systems respond to a system of differential equations [97]. Mechanistic models regarding both the upstream and the downstream process are worth mentioning. To begin with, advances on the upstream process should be presented.

In 2019 Kotidis et al. conducted a research where the impact of glycosylation precursor feeding on cellular growth and metabolism and on antibody productivity and glycoform distribution was quantified [98]. On the one hand, the cell culture and nucleotide sugar donor metabolic models were namely utilized. Thus, 102 unknown parameters were involved, from which 17 parameters were allocated fixed values based on bibliography available and the rest 85 were calculated by fitting the model equations to the existing experimental data using gPROMS. On the other hand, in the n-linked Glycosylation model, the other model used, unidentified parameters were valued using the work by del Val et al. [97].

Moreover, advances were made on the already known fed-batch systems. Namely Amribt et al. (2013) suggesting a macroscopic modelling for hybridoma cell cultures where the three metabolism states: respiratory metabolism, overflow metabolism and critical metabolism were taken into account. The parameters in this study were estimated with the least square method and were validated with experimental data. Hence, the model effectively predicted the cell growth and death, substrate consumption and lactate and ammonia production, which are the main metabolites, as the time passes [99].

Nevertheless, fed-batch reactors are not that productive and tend to be preplaced with perfusion reactors. In 2019 a study was performed by Shirahata et al., comparing a fed-batch and a perfusion bioreactor of CHO culture. The mechanistic model used was once again validated with the experimental data available in literature and the cell, substrate and product concentrations were successfully predicted. Even the by-product concentrations were accurate. Economic factors were also considered, as an economic analysis was also conducted [100].

Undoubtedly process has been made regarding the downstream process too. Müller-Spath et al. in 2011 uses a lumped kinetic model to describe the cation-exchange capture step of a monoclonal antibody (mAb) at the purification process using multicolumn continuous chromatography (its nomenclature is MCSGP). Model parameters including porosity, mass transfer of purified mAbs and retention factors were experimentally determined. Nevertheless, the saturation capacity was calculated through peak fitting, assuming a Langmuir-type adsorption isotherm. In the end, the model was substantiated using linear

batch gradient elution. This model approach was precisely predicting product concentration. However, it failed to provide information regarding the product purity [101].

An earlier publication of Hahn et. al in 2005 was focused on different protein-A affinity resins for mAb purification. Mass transfer in these resins was designated by a model including film and pore diffusion. The study was shown that all the media in were suited for capture of feed stocks with high antibody content, which is a prerequisite in nowadays mAb production [102].

Additionally, Steinebach et al. in 2016 implemented a semi-continuous twin protein A chromatography column model. All in all, 10 equations are utilized in this model. This mechanistic model namely considers both the specific adsorption as well as transport through the resin beads, which was considered a novelty for that time. To enumerate, the adsorption mechanism is defined with a two-site Langmuir isotherm and transport phenomena in the resin pores are described using lumped driving force mass transfer [103].

In this study was noted that the internal pore diffusion is the rate-limiting phase. This result was supported too in the work of Grom et al. in 2018. where five different types of affinity protein A chromatography resins for CHO cell culture were investigated. In this study, the absorbance of mAbs onto the protein A ligands was designed based again on single site Langmuir adsorption kinetics [104].

Moreover, Close et al. suggesting a mechanistic model regarding industrial hydrophobic interaction chromatography (HIC). The main aim of this study was to underline how helpful may modelling be to get a better understanding on industrial scale processes. Even though purification of products is a well-studied subject monomer subunit ratio, which affects product quality, may arise challenges. This model predicted namely this ratio for a wide range of inlet concentrations and inlet product distributions. The results of this studies were validated with a series of scale-down experiments [105].

Lastly, one of the earliest works on chromatography-based separation was made by Aumann et al. in 2007. Multicolumn ion-exchange chromatography is needed when high purity levels are on demand. This chromatography technique was modelled and validated in this study. Of course, the adsorption of both strong and weak impurities was taken into account [106].

Data-Driven Modelling

Pharmaceutical processes are well known for their complexity. Hence, theoretical models, based on mathematical expression, describing the behaviour of the product, the cells or the substrate are difficult to originate from literature or are computationally impossible to be calculated. These highly multifaceted processes are considered as black-box processes [107].

In these cases, data-driven modelling is utilised, where model equations used derive not from a physical or chemical approach but from data fitting. The main advantage of data-driven models is that a great understanding of the process is not required [108].

Data-driven models are more popular in downstream process where real-time predictions are needed, and mechanistic modelling is more challenging. The reason for that is that upstream process operates in a timelapse of days, whereas downstream process is completed in a matter of hours.

Some of the advantages that have been namely made in downstream process were focused on individual process units. For example, the dynamic behaviour of protein A chromatography columns has been designed with data-based models by Rüdt et al. in 2017. In this research, Partial Least Squares Regression (PLS) modelling was exploited to measure mAbs after protein A capture step, during the load phase. This PLS model predicted accurately the mAbs' concentration [109].

PLS modelling is one of the most wide-spread used data driven approaches. Walch et al. in 2019 managed to predict protein concentration, high molecular weight impurities (like host cell protein and DNA impurities) during elution phase by online data from UV, pH, conductivity, light scattering, and refractive index detectors, as well as a fluorescence and a mid-infrared spectrometer. PLS modelling was used in this case too. Even though in this study no mechanistic modelling is used, real- time predictions were rather accurate [110].

Furthermore, another interesting study where both parametric and non-parametric approaches were used is that of Joshi et al. in 2017. In this study a mechanistic model is used to describe separation mechanism. The mechanistic model in question is a combination of modified Langmuir binding kinetics, which is capable of displaying the effect of even nonlinear changes. However, a data- based approach is helpful to simplify a bit the optimization process [111].

Aside PLS, other approaches are used too. For instance, Liu and Papageorgiou in 2019 preferred the piecewise linear regression modelling, because it's simpler and more accurate that other methods available. This study was against focused on chromatography-based processes and loaded mass, flow velocity, and column bed height where the inputs. The data provided were sprung from microscale experiments [112].

Hybrid Modelling

Hybrid modelling is the idea of combining a nonparametric and a parametric model into a single semi-parametric model structure. This permits the fuse of both process knowledge and data-driven information. The hybrid model structure is able to overcome the shortcomings of

each separate modelling technique. To give an illustration of what that means, the black box can be utilized to calculate parameters in the white box, which therefore do not have to be solely assumed, and that reduces errors at the cost of increased complexity [113]. The last example highlights the fact that the values of specific rate expressions are known unknowns a priori for a bioprocess and should first be determined, for instance, by process modelling. Having said that, by solely utilizing a white-box model these rate values should be assumed from data using a defined causal method, however in a hybrid model, they initially can be estimated in a defined black box and afterward transferred to the white box. Not to mention that, by utilizing process variables that have an influence on these rates as input to the black box, the incorporation of this impact can also be taken into account in the white box, and therefore generate hybrid model predictions closer to the analytical values. For this process artificial neural networks (ANN) are regularly utilized [114]. For robust bioprocess modelling, accurate rate estimations are of great importance and accomplishing these as precisely as possible is of high interest [115].

Despite the fact that a hybrid model gives improved performance contrasted with other approaches [116], the chance of misprediction actually exists. To determine the chance of such model uncertainty, cross validation is regularly performed in machine learning for the calculation of the average misprediction possibility [117].

Owing to these benefits, hybrid modelling is gaining in popularity for bioprocess modelling. By combining non-parametric and parametric modelling, the complex, time-consuming, equations describing the model can be eliminated. Consequently, the resulting model will have high accuracy and precision when it comes to the results produced, while simultaneously demanding only a fraction of the original computational time, compared to the parametric model. This could prove to be the key to on-line control for downstream processes.

Some indicative studies about both the upstream and downstream progress are to be presented, even though hybrid modelling in downstream process is a newly-hatched field and further progress is about to be made.

Regarding the upstream process, Narayanan et al. in 2020 presented a hybrid modelling of a mammalian cell culture in perfusion reactor. The aim of this study was a better precision in real-time monitoring and online decision making. In this work, two different models were considered, a hybrid-EKF (namely the extended Kalman filter) and the popular PLS data driven-model. The hybrid model consists of both ANNs (artificial neural networks) and a parametric system of mass balance differential equations. The result of this study shows 35% better prediction than the standard PLS model, an upgrade that clearly indicates the potential of hybrid modelling [118].

ANN modelling was utilised too for the prediction of protein glycosylation by Kotidis and Kontoravdi in 2020. This modelling was applied on 4 different glycoproteins. Undoubtedly this

accuracy of the model was outstanding; namely 1.1 % average error in prediction of glycoform distributions [119].

Nevertheless, about the downstream process little information could be found on the literature. An intriguing study was made in 2021 by Narayanan et al. proposed a hybrid ANN model to predict the absorbance of the chromatographic processes for protein capture. This study is considered a novelty, because in chromatography processes mechanistic methods are more popular. However, in hybrid models there is no need for a deep understanding of the process' nature, thus hybrid models are less intensive in terms of effort. Not only that, but they are also more accurate too. The hybrid model had three times more accurate results that the mechanistic Lumped kinetic model that was put in test. As a result, hybrid models are not only simpler to apply but also much more precise than its widespread mechanistic counterparts [120].

3.2. Motivation, Aim and Objectives

The importance of DSP in antibody biomanufacturing has been made clear, given its role in the purification of the product. However, despite the significance of the process, fewer advances have been made regarding the optimization and on-line control of the process, when compared to the ones reported in USP. The higher complexity of DSP, along with its dependency on USP performance pose great challenges, that have yet to be overcome in order to achieve efficient on-line control of the process.

The aim of this thesis is two-fold and focuses on the development of computational tools with the potential to provide a solution to the aforementioned challenges. The purpose of these tools will be to (a) aid in the design of an agile, robust bioprocess and (b) be used as digital twins for on-line applications.

MAb manufacturing is particularly challenged by the interplay between USP and DSP. USP performance may be variable, rendering the design of a robust DSP process a challenging task. For this, thorough understanding of the USP-DSP interactions is of key importance and will aid in the creation of a QbD approach. Given the fact that the DSP studied in this thesis is a purification process, the QbD approach is centred around the maximization of the product purity, while maintaining an efficient process yield. In order to achieve maximum purity, a thorough understanding of the impurities throughout the process is needed, along with a clear knowledge of the desired quality attributes and their allowed thresholds. In addition, the interplay between process parameters and quality attributes needs to be carefully studied, in order to ensure that the process will be operating within the optimal range. The set of feasible and/or optimal process conditions that yield within-spec products is also known as "design space". The identification of such a design space can be the key to the creation of

many similar processes, without the need for complex calculations and with a beforehand knowledge of the expected product purity and process yield.

Beyond design space identification and optimisation, digital tools offer cost-efficient platforms for *in silico* testing, monitoring and control. The capabilities and ease of use of such tools depends highly on the complexity of the process model, which in separation processes is often prohibiting. More specifically, downstream purification processes are described by complex partial differential equations that are challenging to simulate and use for post-modelling analysis, such as dynamic optimisation. Computer-based models describing such processes are often complex and computationally expensive to use. For this reason, the second pillar of this thesis is focusing on alleviating this complexity of the model through data-driven modelling. By replacing the complex partial differential equations of the model with simpler algebraic equations derived from the results of an initial, experimentally validated, simulation model, a simplified data-driven model can be created. Consequently, this model will require less computational power and time to complete a simulation, enabling it to act as a possible controller for the purification process.

The specific objectives of this thesis can be summarised in three main parts (a) the modelling of the MCSGP periodic process presented in section 4.2.1, (b) the execution of a sensitivity analysis and identification of a design space for said process and (c) the development of a digital twin of the process.

4. Proposed Methodology and Model Description

4.1. Methodology

A flowchart of the proposed methodology presented in sections 4.1.1 - 4.1.3 is depicted in Figure 8.

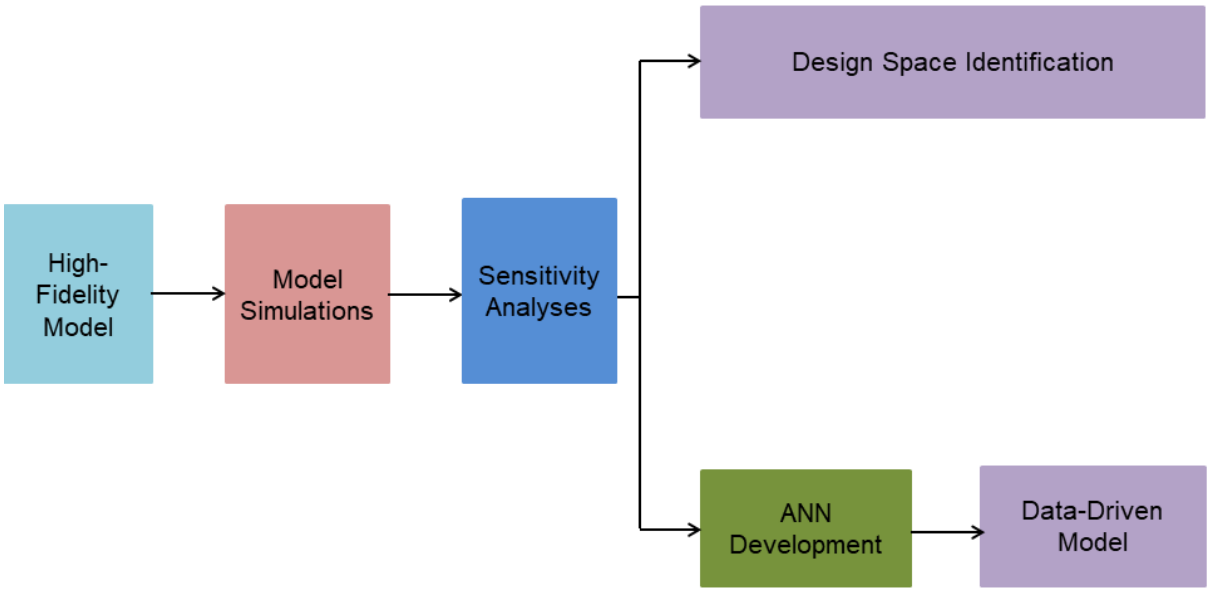

Figure 8: Methodology

4.1.1. Model Development and Simulation

The framework developed and followed in this work begins with the development and simulation of a high-fidelity process model describing the system at hand. Here, gPROMS® ModelBuilder v.7.0.7 is used for the computational experiments.

Once the model structure has been finalized, a preliminary sensitivity analysis is performed to determine the optimal operation parameters of the model, followed by simulations, executed to provide further insight on the behaviour of the process.

4.1.2. Sensitivity Analysis

Sensitivity Analysis (SA) is performed for the identification of the uncertainty introduced by parameter values to the model outputs. SA can be either local (LSA) or global (GSA). LSA is

assesses the local impact of input factors' variation on model response by concentrating on the sensitivity in vicinity of a set of factor values [121]. However, due to the interactions between inputs, observed in many complex and nonlinear spatial phenomenon or processes, it is inappropriate to only evaluate the impact of one input factor on the model output with other factors being constant. It is for that reason that a global sensitivity analysis is also executed.

GSA focuses on estimating the uncertainty in outputs to the uncertainty in each input factor over their entire range of interest, for a certain process. In this case, all the input factors are varied simultaneously, and the sensitivity is evaluated over the entire range of each input factor [122]. Using GSA, the importance of model inputs and their interactions with respect to model output can be quantified, and the result is more realistic to the real world since it allows all input factors to be varied simultaneously. Few of the many GSA methods available include the Sobol's sensitivity estimates, the Fourier amplitude sensitivity test (FAST), and the Monte-Carlo-based regression–correlation indices [122].

In this work, the simulations needed to later perform the local and global sensitivity analyses are going to be executed using gPROMS's feature: Global System Analysis. After that, the results of the global system analysis will be processed using the SobolGSA Software. As indicated by its name, the software uses the Sobol indices approach, which is in turn is based on the ANOVA decomposition, which is generated by a method called Random Sampling - High Dimensional Model Representation (RS-HDMR). The advantage of the RS-HDMR technique is that the model is constructed from randomly sampled input-output data and is represented as a sum of component functions with increasing dimensionality [123]. The reason behind the selection of this particular method rests on the fact that the Sobol indices approach has a better performance with regards to nonlinear models and introduces the individual and total sensitivity index of each input. The two different Sobol indices which are mainly considered are the first-order effect and total effect indices [124].

The first-order effect index indicates the main contribution of each individual input factor on the variance of an output. Therefore, the larger the value of the index, the more impactful the input is on the monitored output. What needs to be noted is that the total sum of indices across the inputs for a particular response is always less than or equal to one. When the total sum is equal to one, then the model is additive. When it is close to one, a small higher order interaction between the inputs is implied, and when it is much smaller than one, it is an indication of significant higher order interactions between the inputs [124].

In the cases where there is an indication of significant higher order interactions between the inputs the total effect index needs to be calculated. This index considers the total contribution including the individual contribution and all of the higher order effects due to its interaction with other inputs. When the total effect index of a factor is equal (or close) to its first-order index, then this factor has no interactions with the other factors. However, if it is larger than

the first-order index, that is an indication of significant interactions. By definition the total effect index is always larger or equal to one [124].

Finally, after the sensitivity analyses have been completed and the Sobol indices calculated, a design space identification will be performed, using the information gathered in the previous steps.

4.1.3. Data-Driven Modelling

Following the sensitivity analyses, an assessment of the equations describing the model will be performed. The complex equations detected will then be replaced by an Artificial Neural Network (ANN), that will be developed on Python using TensorFlow.

Artificial neural network models are developed by training the network to represent the relationships and processes that are inherent within the data. Since they are non-linear regression models, they perform an input–output mapping using a set of interconnected simple processing nodes or neurons. That means that each neuron takes in inputs (entry data important for predicting the outputs) either externally or from other neurons and passes it through an activation or transfer function such as a logistic or sigmoid curve. These data entries enter the network through the input units arranged in an input layer and are then fed forward through successive layers (including a hidden layer) to emerge from the output layer [125].

In this case, after determining the equations to be removed, the variables used to calculate these equations will be identified. Using the results of the previously executed GSA, the values of these variables will be used as inputs for the ANN and the values previously calculated by the equations will become the outputs.

After the training of the ANN is complete, a new set of inputs and outputs will be generated using the original high-fidelity model on gPROMS. This set will be used in turn for the assessment and the validation of the model.

4.2. Model Analysis

4.2.1. The Multicolumn Countercurrent Solvent Gradient Purification

This work focuses on the simulation and post-modelling analysis of the Multicolumn Countercurrent Solvent Gradient Purification (MCSGP). MCSGP is based on ion-exchange chromatography. It was developed at the Swiss Federal Institute of Technology (ETH) Zürich

by Aumann and Morbidelli, and is used to separate or purify biomolecules from complex mixtures [106]. The process consists of two to six chromatographic columns, which are switched in position opposite to the flow direction. They are connected to one another in such a way that as the mixture moves through the columns the compound is purified into several fractions [126].

Generally, a single column process is made up of four parts [106]:

- Equilibration
- Load
- Elution
- Stripping

In turn, the elution of the feed can be divided in five steps [106]:

- Elution of weak impurities (W).
- Overlapping region containing weak impurities and product (W/P).
- Window where the product (P) is pure.
- Overlapping region containing product and strong impurities (P/S).
- Elution of strong impurities (S).

In batch chromatography, weak and strong impurities (R1 and R2 respectively as depicted in Figure 9) often have to be discarded to reach a certain product purity. This is due to the fact that the overlapping windows are undoubtedly contaminated. However, they do contain a large quantity of target product, that cannot be wasted and needs to be recovered, in order to obtain a satisfactory process yield. In batch processes, this is performed through an external recycle, which increases the process time and the risk of errors made by the operator. The automation of this step is achieved through the MCSGP technique [127].

With MCSGP, weak and strong impurities are internally recycled in a periodic closed loop process, while pure product is continuously extracted. As a result, little product is lost, and the yield of pure product is maximized without any accumulation of impurities, whilst retaining target purity. This leads to a significant improvement in process economics, since MCSGP gives higher yield and purity than batch chromatography [128].

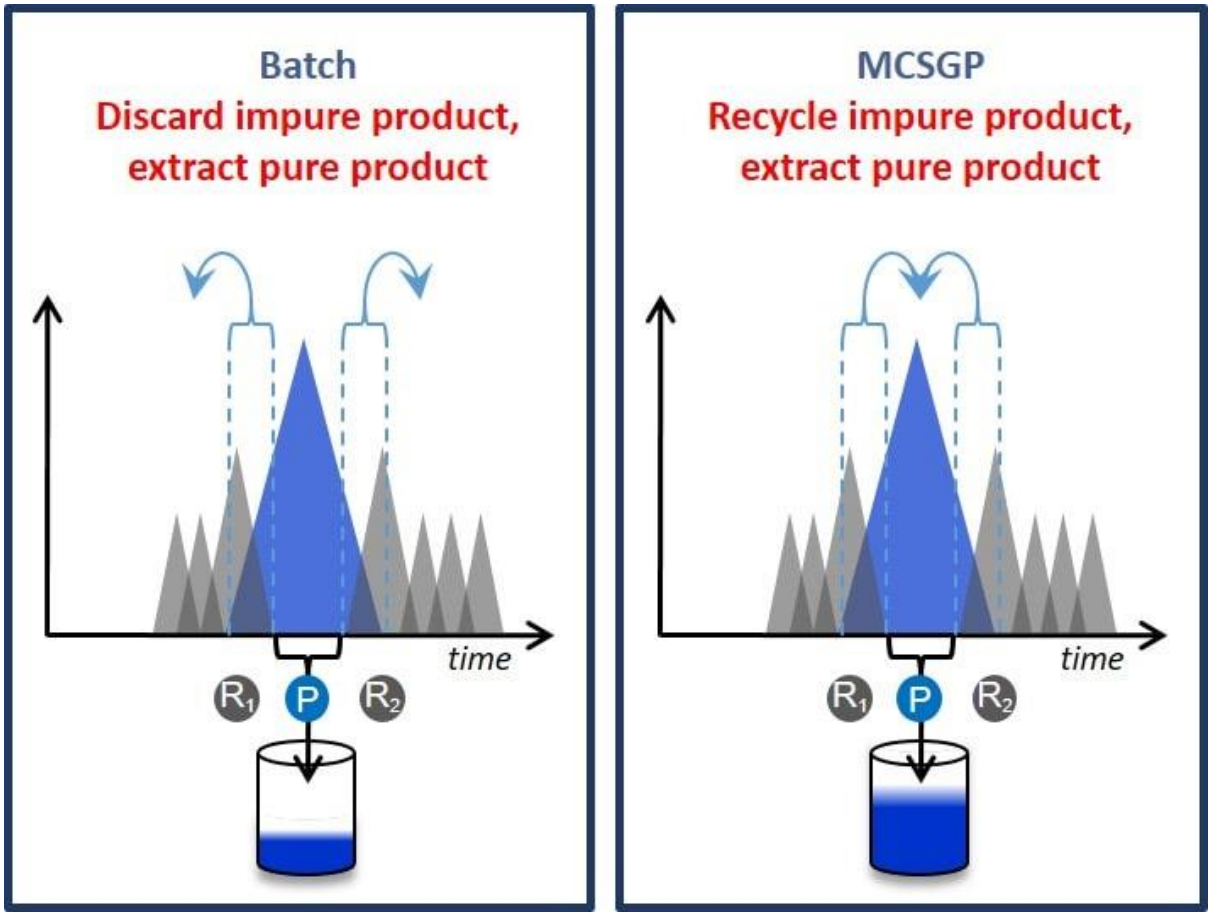

Figure 9: Comparison between batch and MCSGP Chromatographic processes.

Twin-Column MCSGP

The twin-column MCSGP system comprises two identical, ion-exchange, chromatographic columns, operating in countercurrent mode, in order to purify a mAb from a ternary mixture, composed by weak impurities (W), the product (P), and strong adsorbing impurities (S), as mentioned above. Alternated between batch (phases B1 and B2) and interconnected (phases I1 and I2) state, these two columns operate in a semicontinuous fashion, as shown in Figure 10 [127].

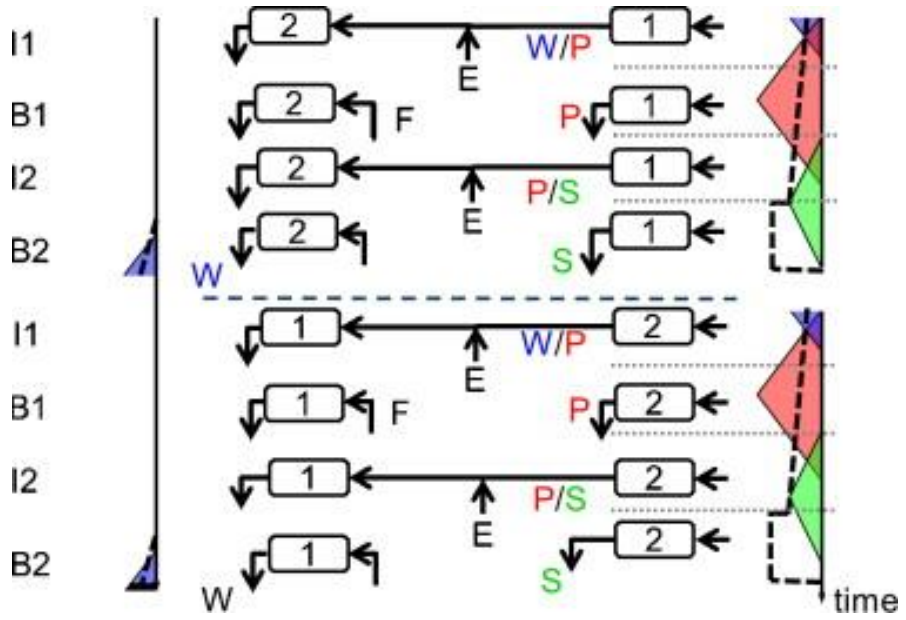

Figure 10: Schematic overview of a complete cycle of the twin-column MCSGP process, adapted from Krättli et al. [127]

Figure 10 illustrates the main process setup of the twin column MCSGP. The general approach of the MCSGP process is to divide the product collection into three different steps, where only the central fraction containing the product (phase B1) within specifications is withdrawn from the unit, while the two overlapping regions are recycled within the unit. These regions contain product contaminated by weak (phase I1) or strong (phase I2) impurities and therefore are recycled onto a second column [129].

More specifically, at the beginning of phase I1, column 2 starts empty and equilibrated, in order to be filled with the outlet flow of column 1 mixed with an additional fraction of adsorbing eluent (E). This is done to favor adsorption since elution occurs at higher modifier concentration/solvent strength than adsorption. By the end of I1 and at the beginning of phase B1, the feed (F) is introduced to column 2 and the product is eluted from column 1. This is the only step during which product collection occurs. Similarly, in phase I2 the outlet flow exiting column 1 enters column 2, containing the impure fraction of product mixed with strong impurities. By the end of I2 phase and at the beginning of phase B2, column 1 starts eluting pure S, whilst column 2 starts eluting pure W. When the overlapping region of W and P reach the end of column 2, phase B2 ends and phase I1 starts again, with the columns switching positions. This means that column 1 will go through the recycling and feeding tasks as described above, while column 2 will continue with the gradient elution [130].

The Process Model

The model used in this work has been previously developed and validated by the Morbidelli Group [131]. Due to confidentiality agreement between Imperial College London and the Morbidelli Group (ETH Zürich) the exact values for parameters and outputs cannot be disclosed and therefore only estimates are provided.

Liquid Phase Concentration

$$\frac{\partial c(z,t)_{i,h}}{\partial t} = D_{ax}\frac{\partial^2 c(z,t)_{i,h}}{\partial t^2} - \frac{Q_h}{A_{col}\varepsilon_i}\frac{\partial c(z,t)_{i,h}}{\partial z} - \frac{(1-\varepsilon_i)}{\varepsilon_i}\frac{\partial q(z,t)_{i,h}}{\partial t}$$

Solid Phase Concentration

$$\frac{\partial q(z,t)_{i,h}}{\partial t} = k_i\left(q^*\left(c(z,t)_{i,h}\right) - q(z,t)_{i,h}\right)$$

Initial Conditions

$$c(z,t=0)_{i,h} = c^0_{i,h}$$

Boundary Conditions

$$c(z,t=0)_{i,h} = c^{in}_{i,h}$$

$$\frac{\partial c(z=L_{col},t)_{i,h}}{\partial z} = 0$$

Where, $t \in [0, t_{end}]$, the time

$z \in [0, L_{col}]$, the column length

$i = 1,\ldots, n_{comp}$, the component

$h = 1,\ldots, n_{col}$, the column index

Isotherms

Competitive Bi-Langmuir Isotherm

$$q^*\left(c(z,t)_{i,h}\right) = \frac{c_{i,h}\cdot H^I_{i,h}}{1+\sum_{i=2}^{n_{comp}}\frac{c_{i,h}\cdot H^I_{i,h}}{q^I_{i,h}}} + \frac{c_{i,h}\cdot H^{II}_{i,h}}{1+\sum_{i=2}^{n_{comp}}\frac{c_{i,h}\cdot H^{II}_{i,h}}{q^{II}_{i,h}}}$$

Henry Constants

$$H_{i,h} = a_{i,1} \cdot \left(c(z,t)_{i,h}\right)^{a_{i,2}}$$

$$H_{i,h}^{II} = a_{i,3} \cdot \left(H_{i,h}\right)^{a_{i,4}}$$

$$H_{i,h}^{I} = H_{i,h} - H_{i,h}^{II}$$

Saturation Capacities

$$q_{i,h}^{I} = a_{i,5} \cdot \left(H_{i,h}^{I}\right)^{a_{i,6}}$$

$$q_{i,h}^{II} = a_{i,7} \cdot \left(H_{i,h}^{II}\right)^{a_{i,8}}$$

Equilibrium Solid Phase Concentration of Modifier

$$q^{*}\left(c(z,t)_{i,h}\right) = H_{mod} \cdot c(z,t)_{h}$$

Purity

$$Pur_{av,j} = \frac{C_{avPs,j}}{C_{avWs,j} + C_{avPs,j} + C_{avSs,j}}$$

Yield

$$Y_{j} = \frac{C_{avPs,j}}{C_{P}^{feed}}$$

Where, j = 1,…, n_{cycle}, the cycle index

s = 1,…, n_{outlet}, the outlet stream

4.2.2. Cyclic Steady State Operation

One of the critical difficulties in developing a chromatographic system is the challenge in optimizing and controlling it. Chromatographic systems are classified into the greater category of periodic systems. As a result of the motion of the sorbate concentration of the liquid and solid phase axially along the bed, their concentration profiles are time and space dependent. In that regard, the continuous model is as a function of time that can reach a pseudo steady state, the cyclic steady state (CSS) [132]. What is basically anticipated in the responses of the model is a periodic nature, meaning that in each cycle after CSS has been reached, the profiles will be indistinguishable from each previous and subsequent cycle.

5. Results and Discussion

5.1. Model Simulation

During the first step of the framework presented, i.e., the model simulation, a preliminary sensitivity analysis on the model (section **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**) is performed to determine the optimal collocation points, at which the model has an acceptable precision and accuracy, without increasing greatly the computational time. The number of collocation points considered ranges from 10 to 70, while the original model has been experimentally optimised for 50 collocation points.

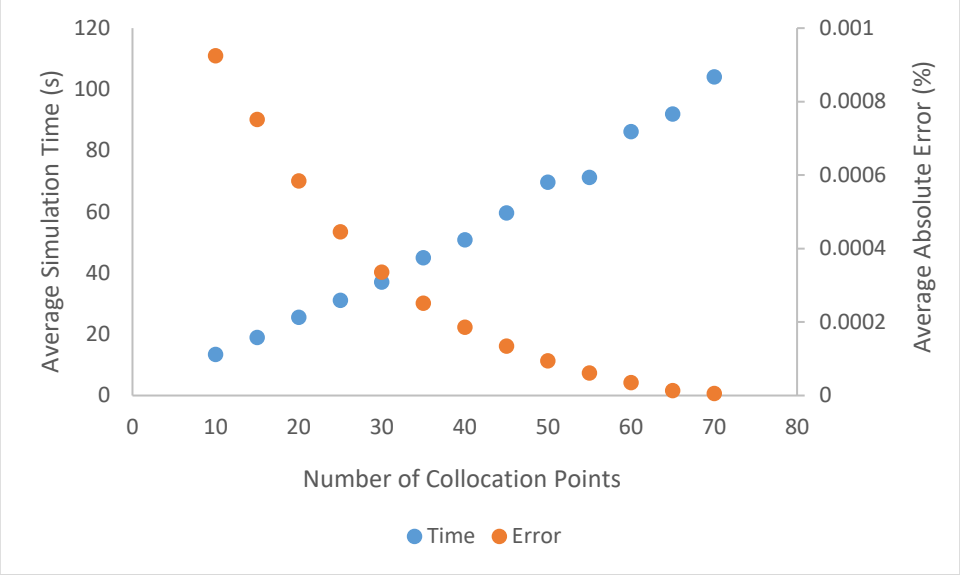

Figure 11: Average error (orange) on the monitored process variables across the samples and average time taken (blue) to complete one simulation of a sample with different number of collocation points.

As can been seen in Figure 11, the error indicated by the orange dots decreases as the number of collocation points increases. However, the simulation time increases at a faster rate as the number of collocation point increases. Essentially, there is a trade-off between speed and accuracy as the number of collocation points is increased. The results indicate that after 50 collocation points, any increase in their number causes a very small respective decrease in error. Therefore, given that the model was already validated using 50 collocation points, it was decided that for the rest of the analysis, 50 collocation points will be used.

After determining the collocation points, the cycle of the process, during which cyclic steady state is achieved, needs to be identified. Using 50 discretization points, as dictated before, a simulation is run completing 10 process cycles. The results of said simulation are presented in the figure below, where the liquid concentration of product at the end of each column was measured against time.

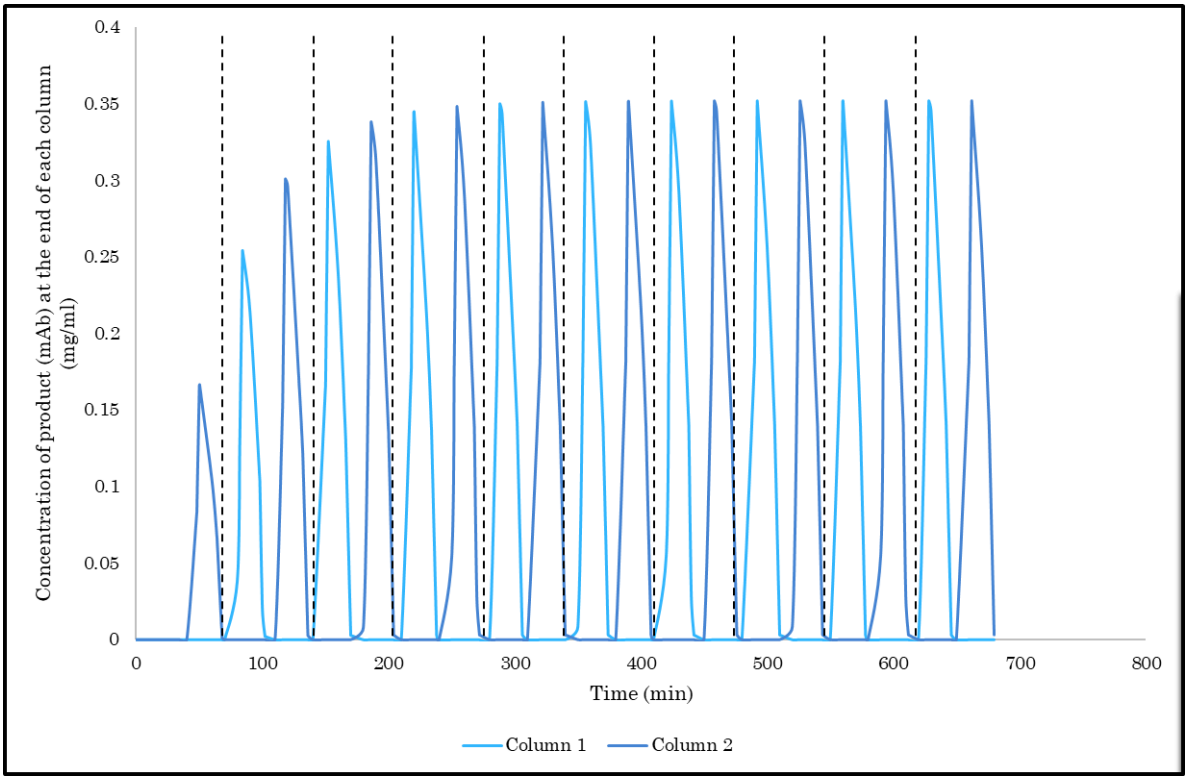

Figure 12: Concentration of product at the end of the column against time for 10 process cycles. The black dash-dotted lines represent the end and start of a new cycle.

As can be easily observed, each process cycle is separated by a dotted black line. To declare that the process has reached CSS, the total mAb concentration at the end of the column is recorded at a time just after elution has started in each cycle (beginning of phase I1). Then the value recorded is compared to the value of the previous cycle for the exact same situation time. This process is repeated for every cycle and when the error between the two values is below 0.1 % (absolute percentage error) or correct to the fourth significant figure, it can be assumed that the process had reached CSS. It is therefore clear from Figure 12 that after the fourth process cycle the peaks of each column become almost identical and cycle steady state is reached.

Having determined the collocation points and the process cycle where CCS begins, the model assessment is complete and its optimal operation parameters are identified. Using these, the global sensitivity analysis can be, in turn, executed.

5.2. Digital Design Space Identification

Here the model was used for the design of a digital operating space (*digital design space*). For that, computational sensitivity analysis was conducted, divided two parts. Firstly, the inputs

analysed, i.e., the concentration of the feed, the flowrate, the concentration of the modifier, the column length, and the column porosity, are analysed separately in a local sensitivity analysis. This gives great insight to the effect of each input regarding the process. Secondly, a global sensitivity analysis is executed, in order to assess the total effect of each input on the process as well as their interactions.

The most important values of the process are the process yield and the product purity. It is for that reason that, in all the sensitivity analyses, these two are the main constrains. In order for the process yield to be considered satisfactory a minimum of 80% is required. At the same time, given the nature of the product, only a purity of at least 98% is accepted, to assure the safety of the product. Given that the product purity of interest, is that of the collected product, all the sensitivity analyses are executed at the end of each column during the collection phases (B1).

The inputs analysed (Table 1: Inputs considered for the local and global sensitivity analyses.) and the outputs calculated (Table 2) are presented in the tables below.

Due to confidentiality agreement between Imperial College London and the Morbidelli Group (ETH Zürich) the exact values for parameters and outputs cannot be disclosed and therefore only estimates are provided.

Table 1

Table 1: Inputs considered for the local and global sensitivity analyses.

	Input	Baseline Value	Lower Bound	Upper Bound	±%	Units	Description
LSA_1	$C_{feed,W}$	0.07	0.05	0.08	20	mg/ml	Feed concentration of weak impurities
	$C_{feed,P}$	0.4	0.3	0.4	20	mg/ml	Feed concentration of product
	$C_{feed,S}$	0.04	0.03	0.05	20	mg/ml	Feed concentration of strong impurities
LSA_2	Q_{MAX}	1	0.5	1	-	ml/min	Max flowrate
	Q_1	0.1	0.1	1	-	ml/min	Inlet flowrate of the column executing the gradient elution during phase I1
	Q_2	0.1	0.1	1	-	ml/min	Inlet flowrate of the column executing the gradient elution during phase I2
LSA_3	$C_{feed,M}$	2	1.6	2.4	20	mg/ml	Feed concentration of modifier
	$C_{M,1}$	2.5	2	3	20	mg/ml	Initial modifier concentration for the column executing the gradient elution

	$C_{M,2}$	1	0.8	1.2	20	mg/ml	Initial modifier concentration for the column executing the recycling and feeding tasks
LSA_4	L_{col}	15	9	21	40	cm	Column length
	LD_{col}	20	14	26	30	-	Column length to diameter ratio
LSA_5	ε_M	0.8	0.5	1	30	-	Column porosity for modifier
	$E_{W,P,S}$	0.6	0.4	0.7	30	-	Column porosity for product and impurities

Table 2: Outputs calculated in the local and global sensitivity analyses.

Input	Baseline Value	Units	Description
Purity	0.98	mg/ml	Product purity
Yield	0.88	mg/ml	Process yield
C_M	2.6	mg/ml	Modifier concentration at the end of the column during product collection
C_W	0.0003	mg/ml	Weak impurities concentration at the end of the column during product collection
C_P	0.3	mg/ml	Product concentration at the end of the column during product collection
C_S	0.007	mg/ml	Strong impurities concentration at the end of the column during product collection

5.2.1. Local Sensitivity Analyses

Concentration of Feed

The first local sensitivity analysis executed is to determine the effect of the feed concentration, coming from the bioreactor, on the process. As has already been explained in the section above, the feed form the bioreactor is only introduced in the columns during stages B1 and consists of the product and weak and strong impurities.

In this particular case, the feed concentration does not constitute a degree of freedom, since it is predetermined by the upstream process. It is however a disturbance that may vary during

the process. Therefore, the purpose of this sensitivity analysis is to determine the bounds in between which, the product collected has the desired qualities.

In order to study the effect of each feed component on the process, the following diagrams are designed:

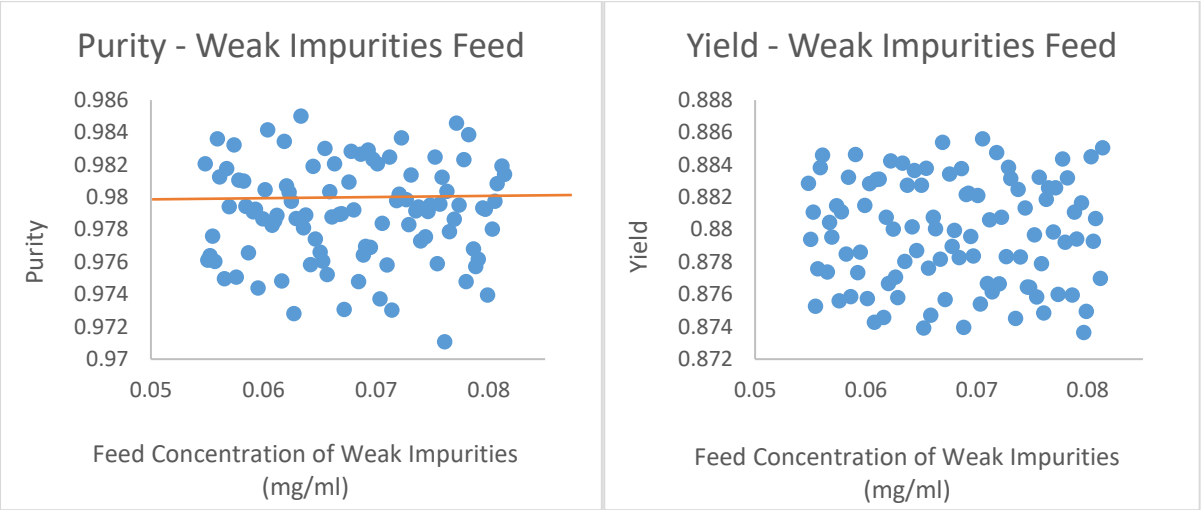

Figure 13: Weak Impurities' feed concentration effect on Purity and Yield

It can easily be observed that the effect of the weak impurities on the product purity and the process yield does not follow a certain pattern. When it comes to the latter, it appears that no matter the concentration of weak impurities, the yield is held above the 80% threshold. This is to be expected if one considers the bounds chosen for the sensitivity analysis. The variation of the concentration of feed was kept at a maximum of 20% (the maximum of the expected disturbance), which is the variation for which the model has been validated, and a yield above 80% is assured [129].

In the case of the product purity, it appears that more than half the concentrations analysed produce a product of unacceptable quality. However, there is no clear indication that the concentration of weak impurities has any effect on it. On the contrary, what should be considered is the fact that the concentration of the total feed was analysed, and not simply of the weak impurities. It is therefore necessary to proceed to an examination of the impact of the concentration of the product, as well as the concentration of the weak impurities, before arriving to further conclusions.

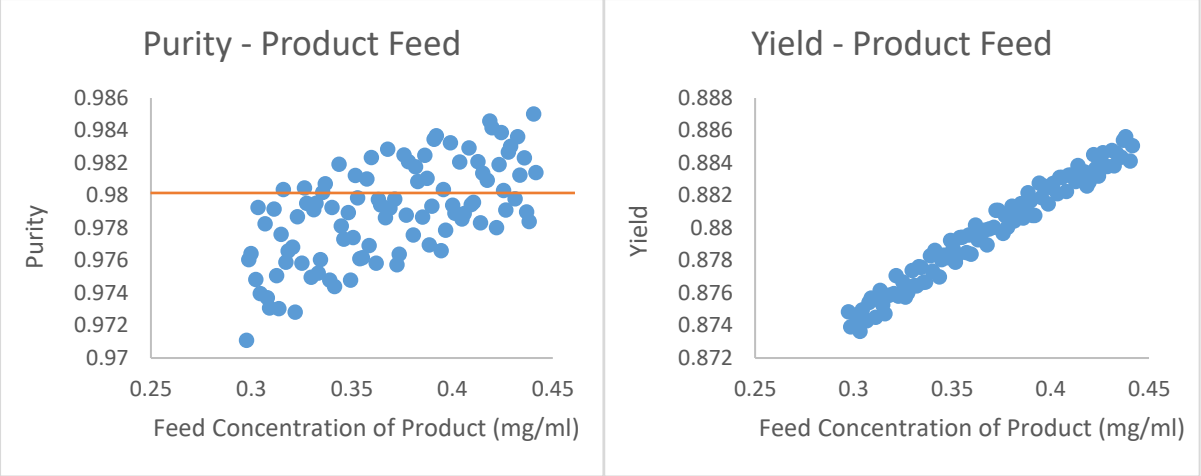

Figure 14: Product's feed concentration effect on Purity and Yield

Contrary to the concentration of weak impurities, the concentration of the product appears to have a significant impact on both the product purity and the process yield. More specifically, an increase in the product concentration of the feed leads directly to an increase in purity and yield. Starting from the latter, this behaviour is expected, since the yield is calculated by dividing the product concentration at the moment of collection with the product concentration of the feed. Considering the principles of ion-exchange chromatography, while it is not entirely possible to completely separate the three feed components, it can be done to a great degree, with only a fraction of the product being eluted with the other two components, outside of the collection points. By increasing the concentration of the feed there is only a small increase of this fraction, and therefore a larger percentage of the product is collected.

Consequently, this increase in the concentration of the product being collected leads to an increase to the product's purity. In this case however, the increase is not as linear as in the case of the process yield. That indicated that there is another factor influencing the purity of the collected product. Since the weak impurities have so far proven to have insignificant impact, the strong impurities need to be examined.

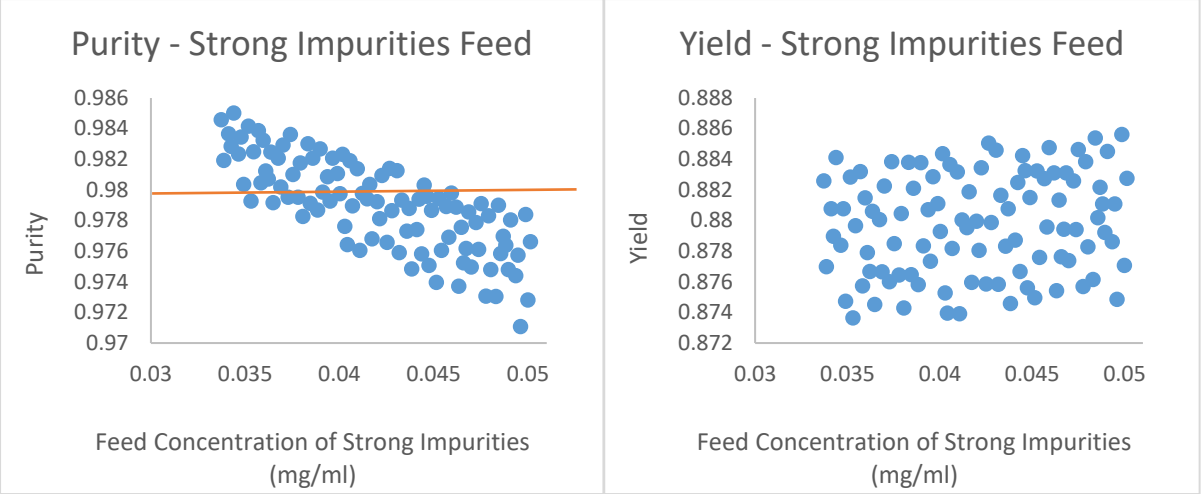

Figure 15: Strong Impurities' feed concentration effect on Purity and Yield

By examining the effect of the concentration of the strong impurities on the product purity and the process yield, the previous theory that the concentration of strong impurities has a large impact on the product purity proves to be correct. The effect that strong impurities have on product purity appears to be the exact opposite of the one the feed concentration of the product has. This means that a larger initial concentration of strong impurities in the feed leads to a reduced product quality.

However, this was not the case with the weak impurities. The reason behind this difference in behaviour is the very nature of the impurities. By definition, weak are the impurities easier to separate, whilst strong are the ones that are the hardest to remove. As a result, a 20% disturbance in the concentration of the weak impurities is quickly corrected by the chromatographic process, leaving no lasting effects on the product. Contrary to that, a 20% disturbance in the concentration of strong impurities, can lead to a reduced product quality by greatly affecting the product's purity.

Finally, similar to weak impurities, strong impurities appear to not have a significant impact on the process yield. This is explained by the fact that the yield is calculated solely by the concentration of the product collected and a variance in the concentration of impurities does not directly affect the concentration of the collected product.

To further understand the interactions between the inputs and the outputs the following first-order and total effect indices are calculated:

Table 3: First-order indices for the feed concentration

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
$C_{feed,W}$	0.05	0.04	0.04	0.18	0.04	0.01
$C_{feed,P}$	0.15	0.11	0.08	0.12	0.25	0.07
$C_{feed,S}$	0.10	0.14	0.14	0.09	0.10	0.51
Sum	0.30	0.29	0.26	0.39	0.39	0.59

Table 4: Total effect indices for the feed concentration

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
$C_{feed,W}$	0.50	0.48	0.52	0.58	0.40	0.29
$C_{feed,P}$	0.58	0.54	0.56	0.48	0.62	0.36
$C_{feed,S}$	0.62	0.69	0.67	0.55	0.59	0.75
Sum	1.70	1.71	1.75	1.61	1.61	1.40

When studying the tables above, only values larger or equal to 0.1 are considered significant. By looking at the first-order indices it is once again confirmed that the feed concentration has no significant impact on any of the outputs, i.e., the product purity, the process yield, and the concentrations at the end of the column at the moment of collection, except of course the concentration of weak impurities during product collection. Additionally, it appears that the feed concentration of the product has a higher impact on the product purity than the strong impurities do, while the strong impurities have a larger impact on the process yield, which was not predicted by the diagrams above. However, it needs to be noted that all those impacts are very small, with their sum being very small in almost every case, which is an indication for significant interactions between the feed concentrations themselves.

Looking at Table 4, the total effect indices of every input are significantly larger, which confirms the existence of input interactions. These interactions are the very reason that the chromatography is needed. By definition, these three components of the feed are difficult to separate due to their similar structure and the interactions taking place between them.

Flowrate

After studying the effect of the feed concentration on the process the focus swifts to the flowrate. Unlike the concentration of the feed, which was approached as a disturbance, the flowrate is a degree of freedom of the process, along with the modifier concentration and the switching times.

Throughout the process the flowrate is kept constant and equal to Q_{MAX}. The only two instances where there is a change of flowrate in one of the columns are during phase I1, where the flowrate of the column executing the gradient elution is set to Q_1, and during phase I2, where the flowrate of the same column is set to Q_2. For that reason, when studying the effect of the flowrate on the process, all three values are taken into consideration, as can be seen from the following diagrams.

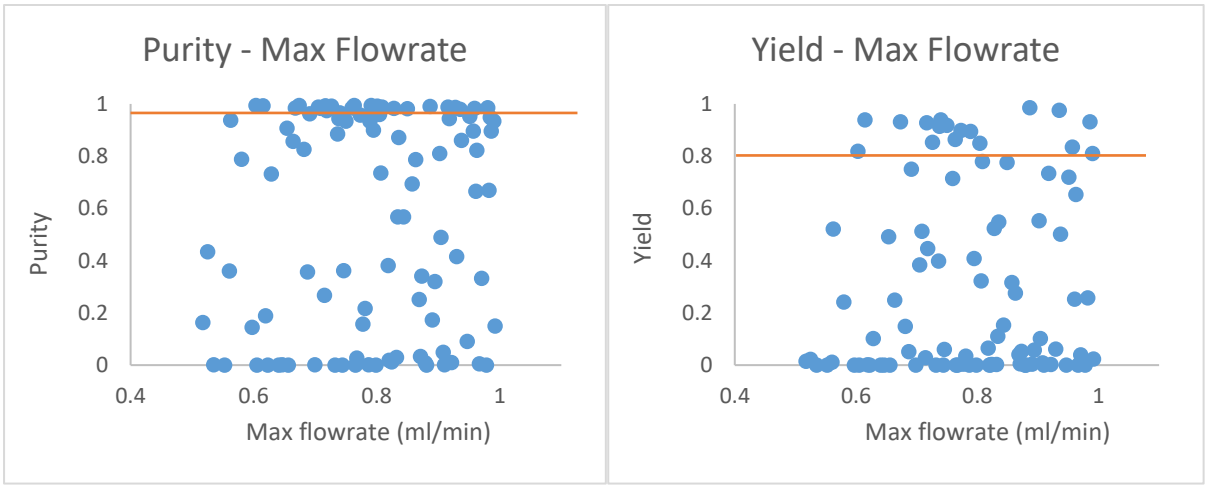

Figure 16: Maximum flowrate's (Q_{MAX}) effect on Purity and Yield

It is clear from the diagrams depicting the impact of the maximum flowrate of the process on the product's purity and the process yield that the flowrate has a significant effect on the process. Certain values can lead to a desired purity and yield, whilst others can lead to a complete failure of the purification process. This is due to the fact that a change in flowrate, given that all the other process parameters are constant (i.e., pH and switching times), affects the velocity of each component inside the column. Starting from the column executing the recycling and feeding tasks, a change in flowrate can affect the initial separation. Having a higher or lower velocity, each component interacts differently with the charged field, meaning that the overlapping regions containing product along with either weak or strong impurities can increase. Additionally, all three feed components arrive at the end of the column at a different time point. Having not changed the switching times and therefore the

product collection time, there is a very high chance that the "product" collected, may not in fact contain any product at all, or if it does, that it will be highly contaminated with impurities, as is depicted in the diagrams above.

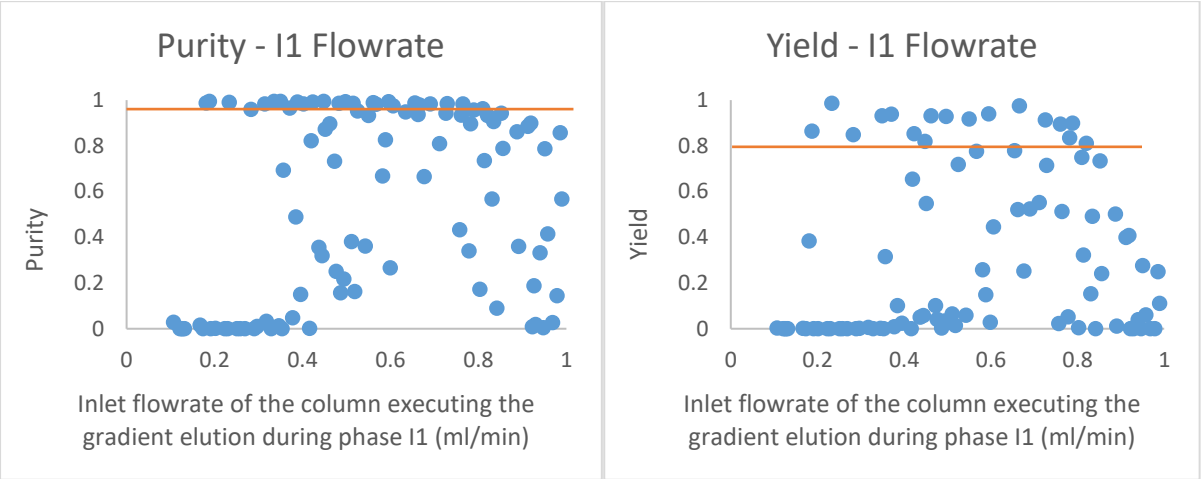

Figure 17: Effect of inlet flowrate of the column executing the gradient elution during phase I1 (Q_1) on Purity and Yield.

Along with the maximum flowrate, the flowrate during the interconnection of the columns of phase I1 (Q_1) appears to have an impact on the outputs. The value of said flowrate determines the amount of mobile phase, supposedly containing the overlapping weak impurities and product region, that is transferred from the elution column to the recycling column for further separation and purification of the product. This flowrate by itself should not have a significant impact on either the process yield or the product purity. However, when coupled with changes in the maximum flowrate of the process, it could lead to a recycling of mostly impurities instead of product, therefore greatly affecting the final product by making the already difficult separation from the weak impurities even harder.

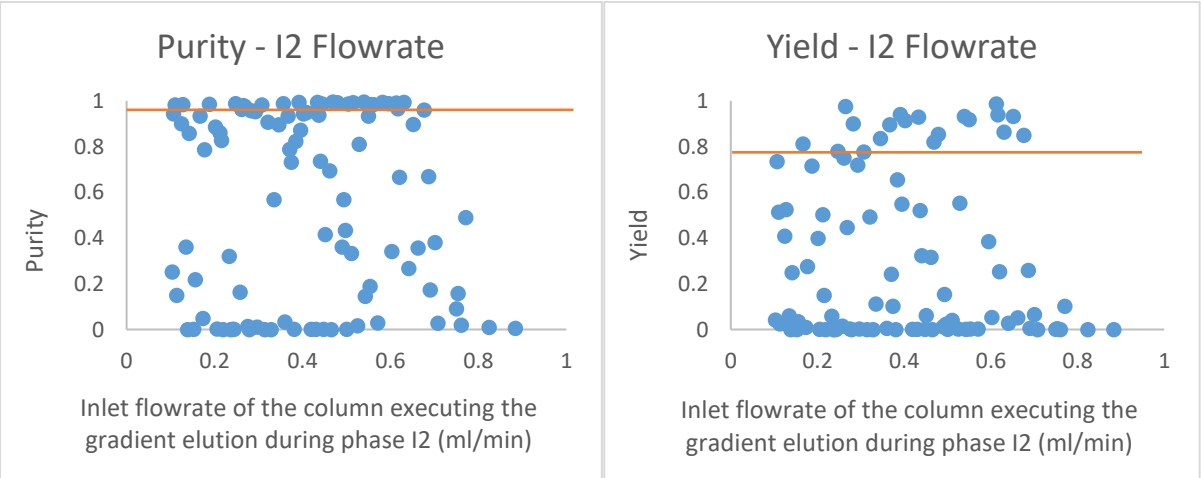

Figure 18: Effect of inlet flowrate of the column executing the gradient elution (Q_2) during phase I2 on Purity and Yield.

Similarly to Q_1, the flowrate during the interconnection of the columns of phase I2 (Q_2) appears to also have an impact on the outputs. A pattern however is once again not indicated by the diagrams. The value of this flowrate determines the amount of mobile phase, containing the overlapping strong impurities and product region, that is transferred from the elution column to the recycling column for further separation and purification of the product. Contrary to Q_1, however, the recycling of this overlapping region has a greater impact on the process due to the fact that it is the strong impurities that are being recycled along with the product. By definition, the strong impurities are the hardest to separate and therefore small miscalculations or changes in the recycling feed can lead to a significant impact on the whole process. Furthermore, it is safe to assume that the already significant impact of Q_2, when coupled with changes in Q_{MAX} and Q_1 as well, will be further increased.

For a better understanding of the significance of each flowrate the first-order and toral effect indices are calculated:

Table 5: First-order indices for the flowrate

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
Q_{MAX}	0.06	0.08	0.19	0.11	0.08	0.10
Q_1	0.09	0.04	0.13	0.10	0.04	0.06
Q_2	0.17	0.24	0.17	0.09	0.24	0.13
Sum	0.32	0.36	0.49	0.30	0.36	0.31

Table 6: Total effect indices for the flowrate

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
Q_{MAX}	0.50	0.57	0.47	0.68	0.57	0.69
Q_1	0.40	0.25	0.41	0.32	0.25	0.26
Q_2	0.78	0.82	0.63	0.70	0.82	0.74
Sum	1.68	1.64	1.51	1.70	1.64	1.69

With a first look at the indices when it comes to the product purity and process yield, it appears that indeed only Q_2 has a significant direct impact on these two outputs, but they all

affect them greatly through the interactions between them. However, after examining them further, their direct and indirect impact on the modifier concentration holds great interest.

By changing the flowrates, it is not only the velocity of the feed components in the column that is changed, but the velocity of the modifier as well. Given that the process is an ion-exchange chromatography, the whole outcome of the process is dependent on the charge inside the columns, i.e., the pH. By changing the flowrates without adjusting the modifier concentration as well, there is a subsequent chance in the columns' pH. This in turn leads to variations in the columns' ability to separate the product from the impurities, as well as the time points at which the product is eluted.

From this point on, it is easy to comprehend that these changes in pH directly affect the concentrations of the product and the impurities at the end of the column during product collection, as can be seen in the tables above. As a result, given the length of the effect the flowrate has on the process, it is preferable for it to be treated as a factor to be optimised, along with the modifier concentration and the switching times, rather than a sensitivity analysis input.

Concentration of Modifier

The last LSA executed for the process parameters concerns the modifier concentration throughout the process. In order to assess the modifier's effect on the process the concentrations are broken down to three parts, the first being the modifier concentration inserted into the column during phase B1 along with the feed from the bioreactor, the second being the initial modifier concentration inserted into the column executing the gradient elution during phase I1, and the third being the initial modifier concentration inserted into the column executing the feed and recycling tasks during phase I1. These three initial concentrations constitute a degree of freedom of the process and dictate the modifier concentration in the columns at any given moment.

In order to study the effect of the beforementioned concentrations on the process, the following diagrams are designed:

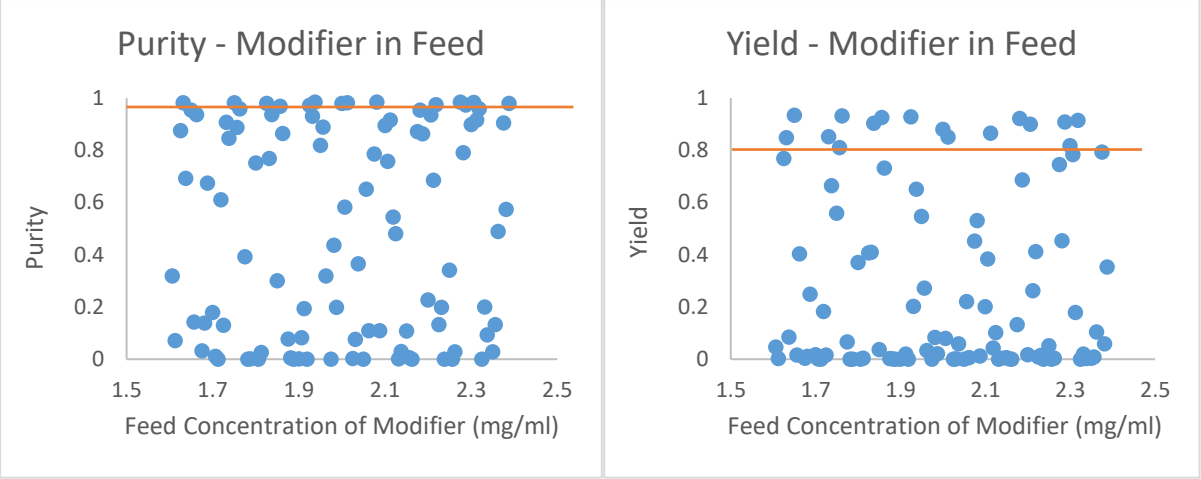

Figure 19: Effect of modifier's feed concentration ($C_{feed,M}$) on Purity and Yield

Based on the two diagrams presented above, the modifier's concentration inserted along with the feed from the bioreactor ($C_{feed,M}$) appears to have little effect on the purity of the product, as well as the yield of the process. This is justified due to the fact this concentration is only inserted in the column executing the feeding and recycling tasks during phase B1, which only lasts for a small fraction of the process cycle. Moreover, right after phase B1 a new dose of modifier is introduced in the column, the concentration of which is dependent on the initial concentration inserted into the column executing the recycling and feeding during phase I1 ($C_{M,2}$). As a result, the concentration of the modifier added with the feed is expected to not have a significant impact on the process outputs, it is expected however to have significant interactions with the rest of the inputs, i.e., the other modifier concentrations added throughout the process.

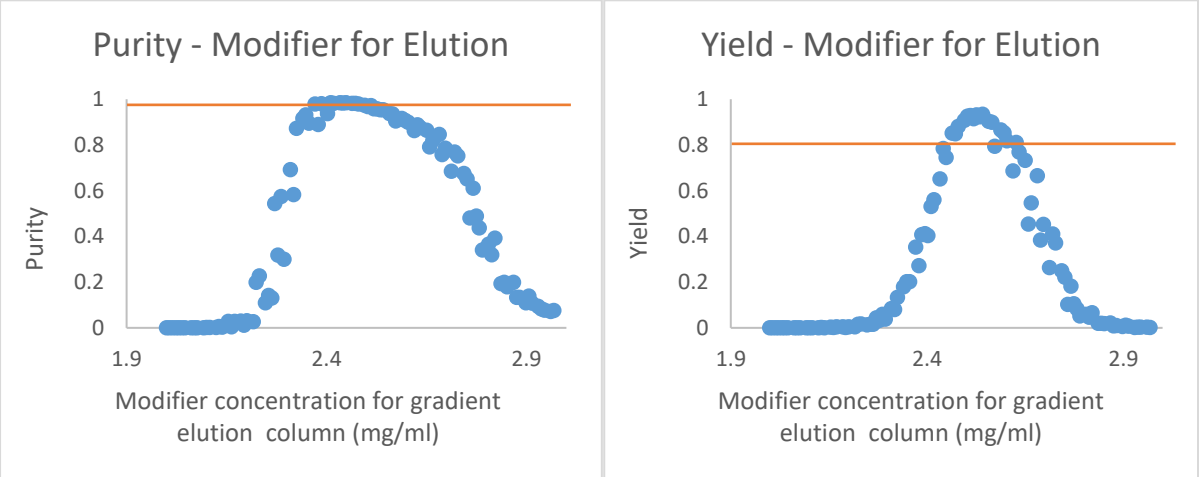

Figure 20: Effect of initial modifier's concentration for the column executing the gradient elution ($C_{M,1}$) on Purity and Yield.

Given that the initial concentration of the modifier entering the column executing the gradient elution ($C_{M,1}$) directly dictates the modifier's concentration during product collection, it is expected to have a large impact. More specifically, the concentration of the modifier affects the column's pH, which in turn affects the ion-exchange chromatography. The principles of ion-exchange chromatography dictate that during product elution, the closer the pH is to the pl of the product, the larger the amount of product that is eluted. However, small deviations in pH during the designated time for product collection, can lead to an electrically charged product that cannot yet be or has already been eluted. At the same time, a delayed or rushed elution of the product can be followed by an increased elution of impurities during collection.

These phenomena can be observed in the diagrams above. Both product purity and process yield present small peaks for a small range of $C_{M,1}$, and outside of those peaks the values of purity and yield quickly drop to zero. This indicates that the elution of the product happens at a completely different time than that of the collection of the product. Much like in the case of the flowrate, it is clear that the concentration of the modifier should not be optimized by itself, but should be accompanied by a respective optimization of the switching times, in order to ensure that the product elution and the product collection occur at the same time.

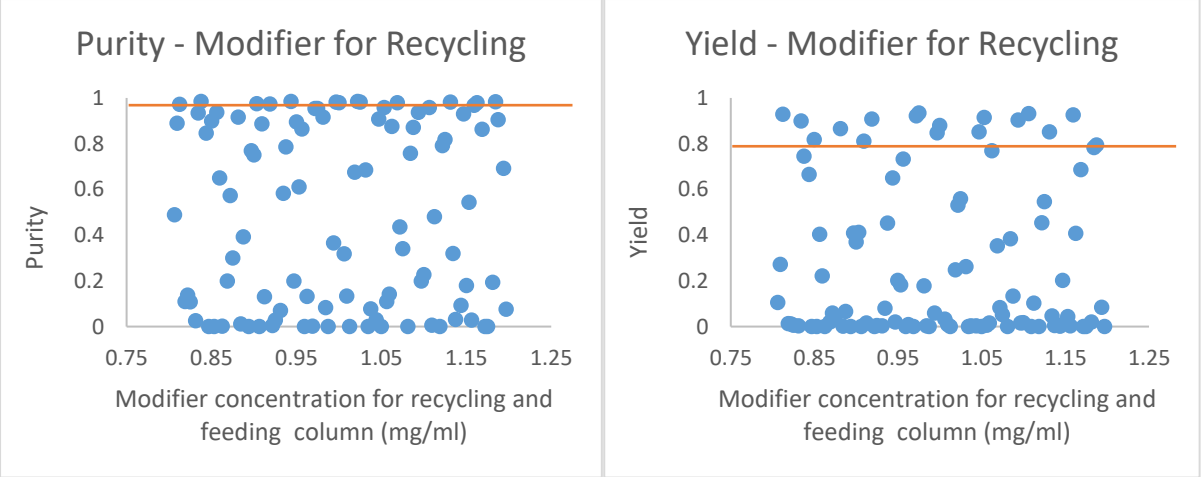

Figure 21: Effect of initial modifier's concentration for the column executing the recycling and feeding tasks ($C_{M,2}$) on Purity and Yield.

The last modifier concentration that can be changed and therefore optimised in the process, is the initial concentration entering the column executing the recycling and feeding tasks during phase I1 ($C_{M,2}$). The diagrams constructed above show no clear sign of a significant impact of $C_{M,2}$ to the product purity and the process yield. However, contrary to the modifier concentration of the feed, in this case it is not expected. $C_{M,2}$ affects the modifier concentration of the column executing the recycling and feeding tasks during phases I1, I2

and B2. It is therefore clear that it controls the pH of the column during the first separation of the feed components. This separation is detrimental to the purity of the process, since if it is not successful, the product eluted afterwards from the other column, will be accompanied by an increased concentration of impurities.

Since the importance of $C_{M,2}$ is not reflected in the diagrams, the first-order and total effect indices of the three modifier concentrations are calculated and presented in the tables bellow:

Table 7: First-order indices for the modifier concentrations

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
$C_{feed,M}$	0.04	0.06	0.11	0.03	0.06	0.00
$C_{M,1}$	0.15	0.40	0.25	0.48	0.40	0.75
$C_{M,2}$	0.26	0.02	0.10	0.11	0.02	0.03
Sum	0.45	0.48	0.46	0.62	0.48	0.78

Table 8: Total effect indices for the modifier concentrations

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
$C_{feed,M}$	0.41	0.38	0.48	0.30	0.38	0.15
$C_{M,1}$	0.44	0.70	0.59	0.74	0.70	0.90
$C_{M,2}$	0.70	0.44	0.48	0.34	0.44	0.17
Sum	1.55	1.52	1.55	1.38	1.52	1.22

Starting from $C_{feed,M}$, the first order indices indicate that it indeed has no direct impact on the process outputs. However, as was expected, it has a great impact on the outlet concentrations, as well as on purity and yield, through its interactions with the other inputs. Similarly, the impact of $C_{M,1}$ on purity is confirmed, along with the larger impact, both directly and through interactions with the rest of the modifier concentrations, that it has on process yield. Finally, the most interesting information derived from these tables is the effect of $C_{M,2}$ on purity and yield. Even though it was not reflected on the diagram, $C_{M,2}$ has the largest

direct impact on the product's purity, as well as the largest indirect (throughout interactions) one.

Column Length

After completing the LSAs concerning the process parameters, the design parameters are studied. In order to assess the shape of the column, a sensitivity analysis for the column length and the length to diameter ratio is executed. The reason behind this approach, instead of a sensitivity analysis over the length and the diameter of the column, is to ensure that the shape of the column remains, otherwise the results produced have no physical meaning and therefore no actual applications.

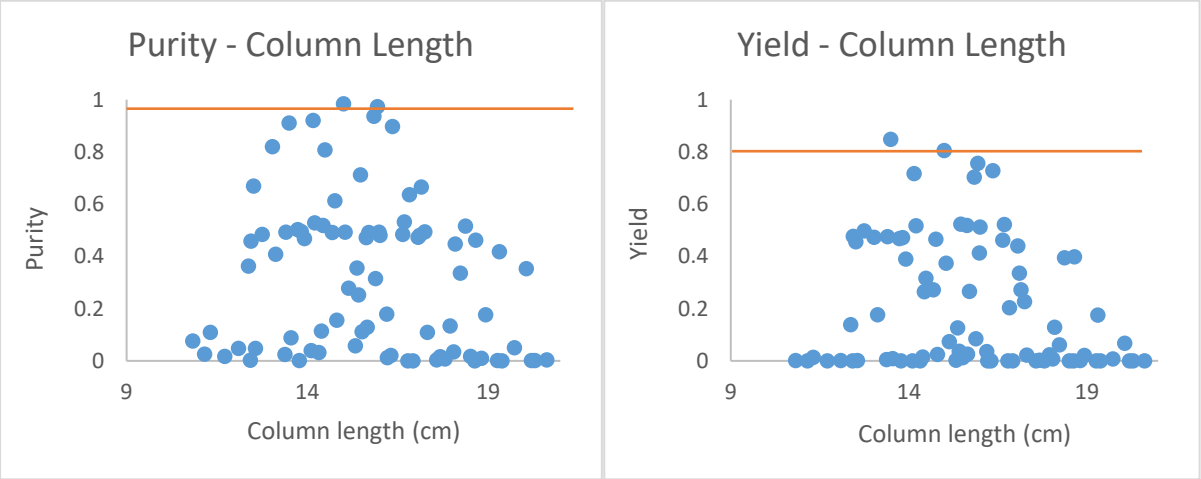

Figure 22: Column length's effect on Purity and Yield

As shown in Figure 22, column length has a significant impact of the process yield and product purity. Keeping in mind that the flowrates are kept constant along with the product collection points, changes in the column length lead to changes in the elution time points of the product. Consequently, a smaller length would lead to the product arriving at the end of the column before the designated time, whilst a bigger length would find it still crossing the column at the moments of collection.

If, however, the switching times were optimized along with the changes in the length of the column, then the results would be different. A bigger length could result in more time for the separation of components to be completed and therefore in higher product purity. Unfortunately though, this increase in product purity would be accompanied by an increase in the cost of the column.

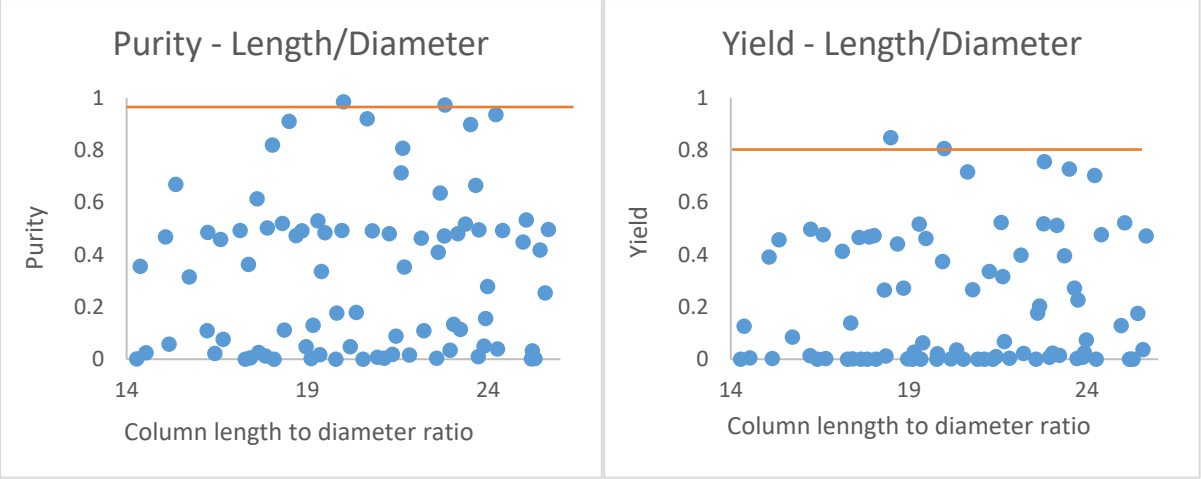

Figure 23: Column length to diameter rate's effect on Purity and Yield

When changing the length of the column, the changes in the length to diameter ratio of the column represent the changes of the column's diameter. Therefore, by studying the diagrams of Figure 23, the effect of the column's diameter can be estimated. Since by changing the diameter of the column there are subsequent changes in the column's surface, it would be expected for the diameter to affect the concentrations of the components inside the column. By increasing or decreasing the adsorption sites there is an accompanying increase or decrease in the effectiveness of the separation process. However, these are merely estimations and are not clearly depicted in Figure 23. To further understand the effect of the length and diameter of the column, the first-order and total effect indices are calculated.

Table 9: First-order indices for the column length

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
L_{col}	0.16	0.16	0.48	0.16	0.16	0.31
LD_{col}	0.02	0.10	0.05	0.06	0.10	0.14
Sum	0.18	0.26	0.53	0.22	0.26	0.45

Table 10: Total effect indices for the column length

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
L_{col}	0.98	0.90	0.95	0.94	0.90	0.86
LD_{col}	0.84	0.84	0.52	0.84	0.84	0.69
Sum	1.82	1.74	1.47	1.78	1.74	1.55

By examining the first-order and total effect indices it is clear that, as expected, the column length has a significant effect on the process yield and the product purity, but an even more significant through its interaction with the column diameter. The actual meaning of the "interaction" effect is that the shape of the column plays a significant part in the purification process. It can therefore be concluded that a change in column would need to be accompanied by an optimization of all the other parameters of the system, in order for the separation process to remain effective.

Column Porosity

The last local sensitivity analysis conducted studies the effect of the column porosity on the product purity and yield. This analysis is relatively simple, given that the porosities of the column are only two, one meant for the modifier, and one meant for the components of the feed, due to their similar size and structure.

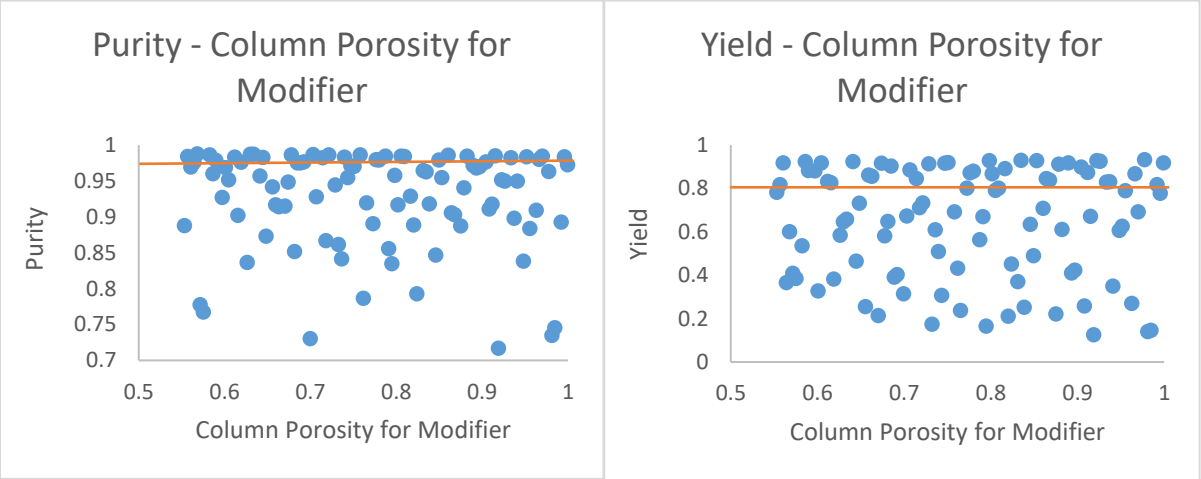

Figure 24: Effect of the column porosity for the modifier on Purity and Yield

The column porosity for the modifier, by definition only affects the concentration of the modifier. It is therefore not expected to have a significant (or any) direct effect on the product purity and the process yield. However, by affecting the concentration of the modifier inside the column, it can lead to unpredicted changes in the pH of the column, with in turn greatly affects the separation process, as well as the elution times of the components. As a result, even though there are no direct effect, a change in the column's porosity for the modifier could indirectly have a significant impact on the collected product.

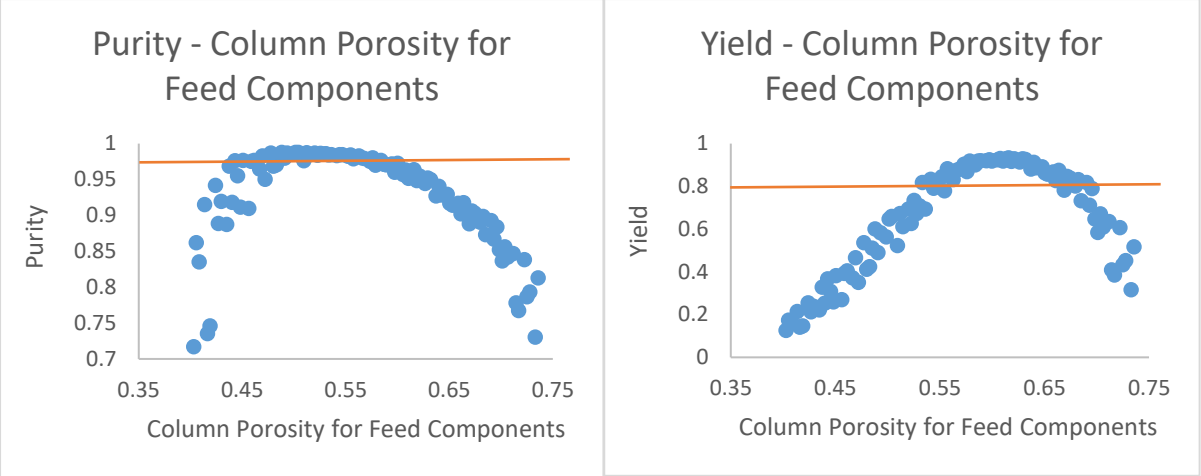

Figure 25: Effect of the column porosity for the feed components on Purity and Yield

Contrary to the porosity for the modifier, the column's porosity for the feed components has significant impact on the products purity, as well as the process yield. The purpose of the column's porosity is to be able to withhold the components of the feed during their journey through the column. Therefore, slight changes to the porosity result in a reduced ability to separate the product from the impurities.

For more insight into the effects of the porosity, the first-order and total effect indices are calculated:

Table 11: First-order indices for the column porosity

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
ε_M	0.01	0.00	1	0.07	0.00	0
$\varepsilon_{W,P,S}$	0.82	0.96	0	0.75	0.96	1
Sum	0.83	0.96	1	0.82	0.96	1

Table 12: Total effect indices for the column porosity

	Pur	**Y**	**C_M**	**C_W**	**C_P**	**C_S**
ε_M	0.18	0.04	1	0.25	0.04	0
$\varepsilon_{W,P,S}$	0.99	1.00	0	0.93	1.00	1
Sum	1.17	1.04	1	1.18	1.04	1

By looking at the indices produced, the greatest interest is held by the values of the first-order and total effect indices of the column porosity for the components and the concentration of the strong impurities. Given that the index is equal to one, there are no other interactions, as are in the case of the product and the weak impurities. This observation leads to the conclusion that the porosity of the column chosen, during the design of the column, mainly aims in capturing the strong impurities and separating them from the rest of the components. Such a strategy can help improve the effectiveness of the process, since the greatest challenge is separating the strong impurities from the product.

5.2.2. Global Sensitivity Analysis and Design Space Identification

Having completed all the local sensitivity analyses and acquired an insight on the inner workings of each variable in the process, a global sensitivity analysis is executed, to study the interactions occurring between the all the process variables. Since an individual study of the effect of each variable would be tedious, a 3D bar chart is created, depicting all the inputs and outputs of the GSA, along with their corresponding indices:

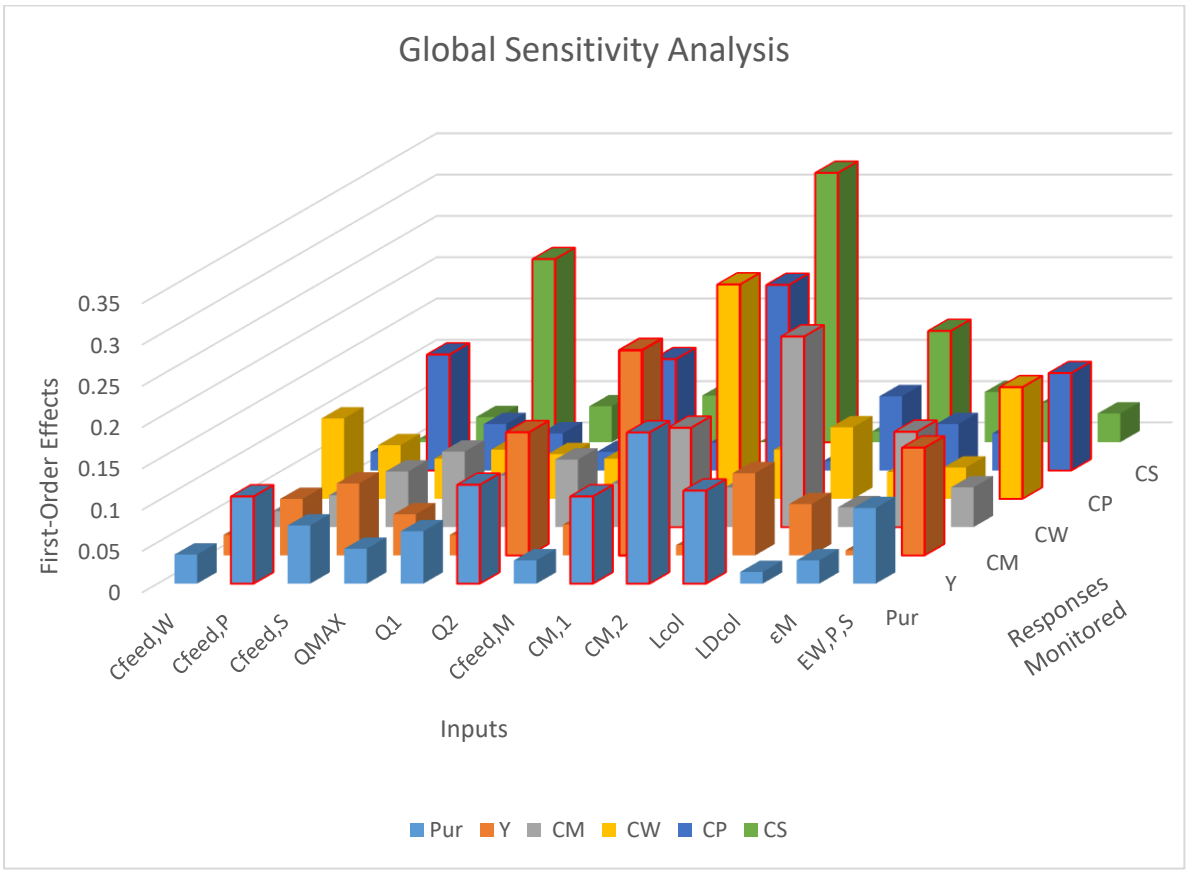

Figure 26: Global sensitivity analysis first-order indices

In Figure 26, the highly complicated effect each input has on the process is presented. Out of all the indices, highlighted with a red outline are the once that are considered significant, i.e., the ones whose value is higher than 0.1. By studying the inputs that have a significant impact on the process, it appears that the only inputs not affecting the process directly (though they could be affecting it indirectly throughout interactions with the rest of the inputs) are the feed concentrations of modifier and weak impurities, the maximum flowrate, and the length to diameter ratio of the column.

To identify the desirable operating spaces, a purity and yield constrain are used. More specifically, the purity required for a product of the desired quality is:

$$\text{Purity} \geq 98\%$$

And the yield required for the process to be considered profitable is:

$$\text{Yield} \geq 80\%$$

Therefore, excluding the variables that do not have a significant direct impact on the process, the following design spaces are identified.

Composition of Feed

For first design space identified, the variables of the feed with significant impact (concentration of product and concentration of strong impurities) are considered., as shown in Figure 27.

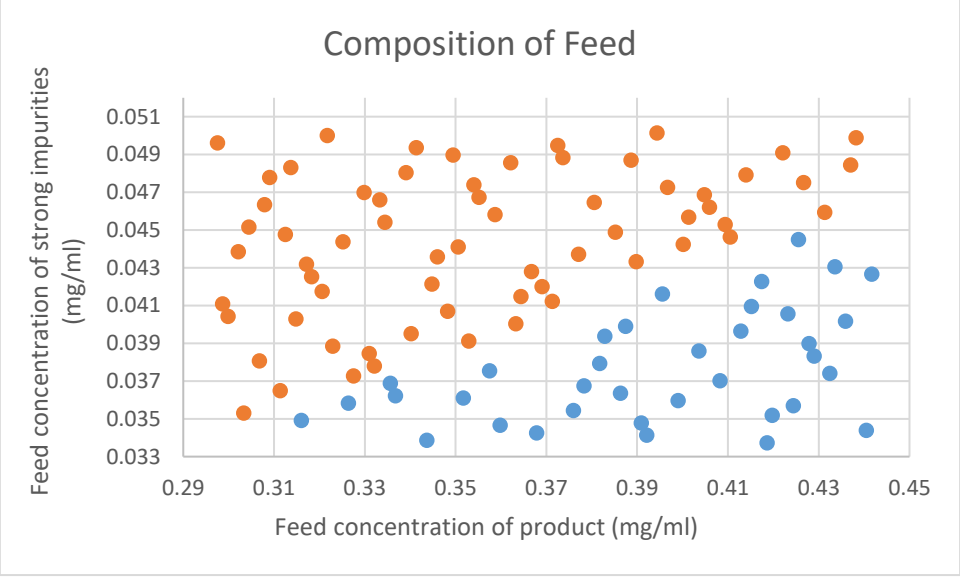

Figure 27: Design Space of feed composition. Blue markers indicate that the constrains are satisfied and orange markers indicate that they are violated.

As can be easily observed in Figure 27, there is a clear line separating the combinations of feed composition that satisfy the set constrains, from the ones that violate them. An increased concentration of the product in the feed, coupled with a decreased concentration of strong impurities, leads to higher chances of the separation process being successful.

Additionally, as was explained in section 5.2.1, the concentration of the weak impurities is not included in the design space, since it has little to no effect in either the product purity or the process yield, when it remains within the expected range of disturbances.

Flowrates During Interconnecting Phases

Continuing with the effect of the flowrate of the process, Figure 26 indicates that the flowrates that have direct impact on the process outputs and the set constrains are the ones used during the interconnecting phases I1 and I2. Therefore, a design space of those two variables is identified to visualize the effect of the flowrate.

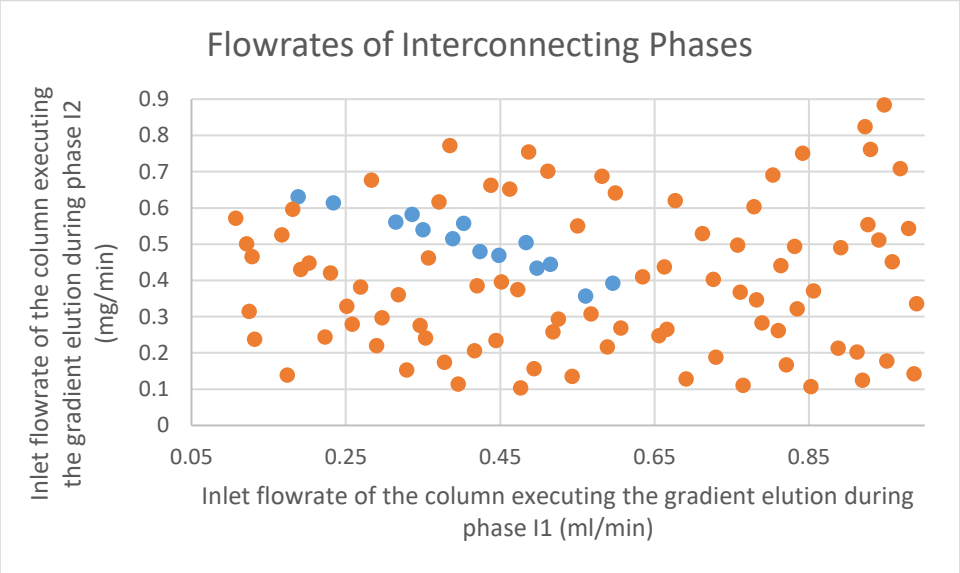

Figure 28: Design Space of the flowrates during the interconnecting phases I1 and I2. Blue markers indicate that the constrains are satisfied and orange markers indicate that they are violated.

Similarly to the design space of the feed composition, in Figure 28 a clear line of values with satisfied constrains is created. However, as can be seen, some of these values are very close with values that violate the constrains of either purity or yield. This is due to the fact that small changes in flowrates, can lead to respective changes in the elution time points of the components. This does not constitute a problem, since the flowrates of the process are

constant and are not subject to unpredicted variations, that may lead to reduced product quality, throughout the process.

Initial Concentrations of Modifier in Each Column

Regarding the concentrations of the modifier, the ones identified on Figure 26 with significant impact on the process outputs are the ones introduced into the column executing the gradient elution tasks during phases I1 and I2. Therefore, a design space of these two concentrations is presented in Figure 29.

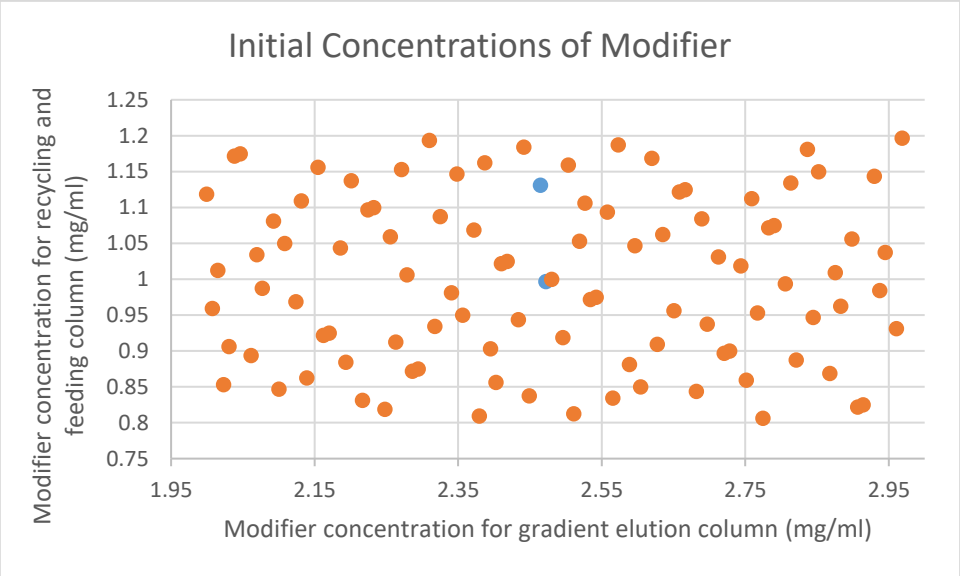

Figure 29: Design Space of the initial concentrations of the modifier for each column. Blue markers indicate that the constrains are satisfied and orange markers indicate that they are violated.

Comparing this design space with the ones concerning the feed composition and the flowrates, it appears that the concentrations of the modifier are more restrictive that the other parameters. Such behavior is expected, given the nature of the chromatographic columns used. Since the separation process in an ion-exchange chromatography, changes in the concentration of the modifier, and consequently in the pH inside the columns, need to be accompanied by process optimization, otherwise the resulting process will not lead to the required product purity and yield. In this case however, the rest of the parameters were kept constant, hence resulting in a very restrictive design space.

Moreover, what holds particular interest is the fact that one of the two operating points suggested from the design space is very close (almost overlapping) to a non-satisfactory operating point. The constrain violated in this case is the product purity, which drops slightly

below the accepted threshold. However, insignificant as it may seem, this small deviation from the desired purity could lead to out-of-spec products that may need to be discarded. It is therefore very important to choose carefully the operating parameters of the process.

Column Dimensions

Having completed the identification of the design spaces of the process parameters, the optimal operating space for the design parameters needs to be identified. Starting from the column dimensions, the only two parameters affecting it are the length of the column and the diameter, which in this case is investigated through the column length to diameter ratio. The design space for these two variables can be seen in Figure 30.

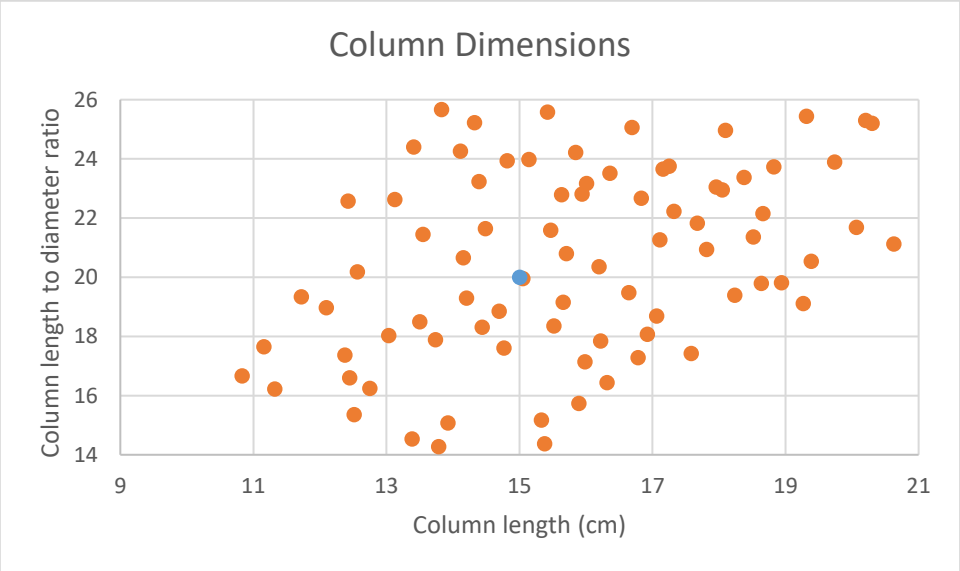

Figure 30: Design Space of the column dimensions. Blue markers indicate that the constrains are satisfied and orange markers indicate that they are violated.

Much like the concentration of the modifier, the column length has very strict operation parameters, as shown in Figure 30. Given the particular parameters of this process (which were kept constant), there is only one column dimension that meats the required criteria for the process, with all the rest of the column dimensions showing a process yield below 80%. This is due to the fact that the process has been optimized for this particular column dimension and not the other way around. As a result, even a slight change in the column length and diameter (as can be seen from the overlapping operating point in violation of the constrains), will lead to a significantly reduced product quality and process yield.

Column Porosity

The last design space created depicts the impact of the column porosity, as can be seen in Figure 31.

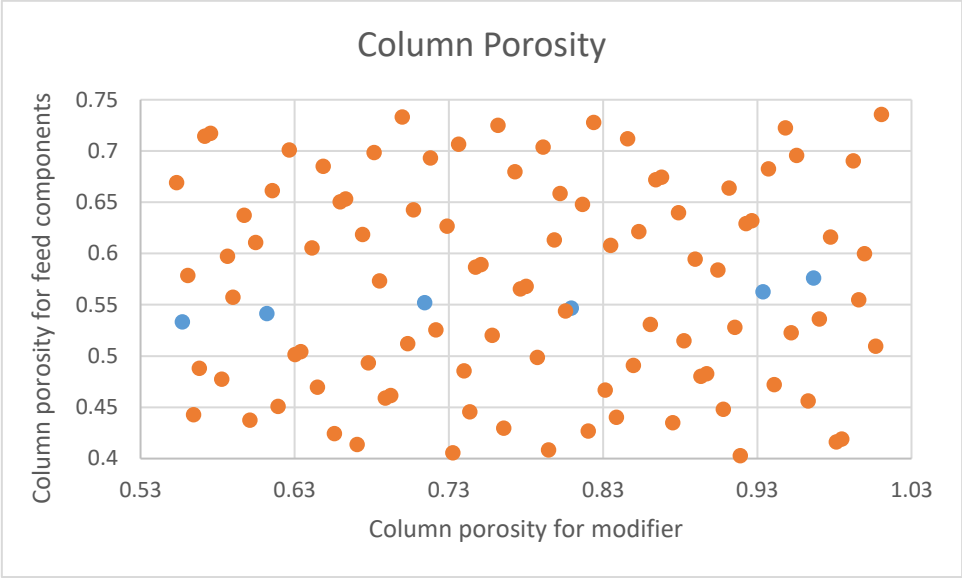

Figure 31: Design Space of the column porosity for each component of the process. Blue markers indicate that the constrains are satisfied and orange markers indicate that they are violated.

The operation points depicted in Figure 31 that satisfy the constrains follow a horizontal line. Therefore, there is a wider number of porosities for that satisfy the constrains, than the one of porosities of the components. This is due to the fact that the porosity for the components has an active role in their adsorption in the column and therefore their separation. Being the components of the feed have a certain size and structure, changes in the porosity of the column would lead to significant decrease in the adsorption capabilities of the column. However, changes in the adsorption of the modifier, that would subsequently cause changes in the pH inside the column, can be optimized through changes in the modifier concentration.

Additionally, is should be noted that the operation points, included in the constraint-satisfying horizontal line, that violate the constrains, show very small deviation from the required yield, whilst still producing a product of sufficient purity.

5.3. Data-Driven Modelling

The first step in order to proceed with the data-driven model is to identify the equations that are going to be replaced by the artificial neural network. By looking at the equations in paragraph 4.2.1, it is clear that the two most complex equations of the system are the partial differential equations describing the liquid and solid phase concentrations of the four components (i.e., the modifier, the weak impurities, the product and the strong impurities). By removing these two equations the rest of the system is turned into a linear model.

However, their removal is not as simple as that. Since the equations are a function of time and space, calculating the values of each concentration at any given time at any given space would demand great computational power as well as great number of GSAs conducted. Since, however, when it comes to controlling the process, the main concern is the concentration of the product collected at the end of the column, it is decided to calculate with the ANN the concentrations at any point in time, at the end of the column.

Moving on to the identification of the variables used in the calculation of the two concentrations, it appears that the only variables used for the calculation of the solid phase concentrations are the equilibrium solid phase concentrations. However, in the case of the liquid phase concentrations, the needed variables are the column cross-section, the column porosity, the flowrate, and the solid phase concentration for each component. This poses a problem since an ANN is unable two calculate two outputs at the same time, when the value (output) of the one is an input for the other.

This problem is overcome by creating two different consecutive ANNs, each one calculating only one concentration. The two ANNs created are identical, except from the inputs given and the outputs calculated. The inputs dataset of each ANN is first shuffled and split into two separate datasets, one used for training (80% of the original dataset) and one used for testing (20% of the original dataset) the model. Afterwards, the training dataset created is used to train the model through three hidden layers using a rectified linear activation function (ReLU), which is described by the equation:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

The results of each individual artificial neural network created are presented in the following paragraphs.

ANN for the Solid Phase Concentrations

The first ANN trained is the one calculating the solid phase concentration of each component. After training the ANN the first indication of how well the training was executed is the diagram presenting the learning curve:

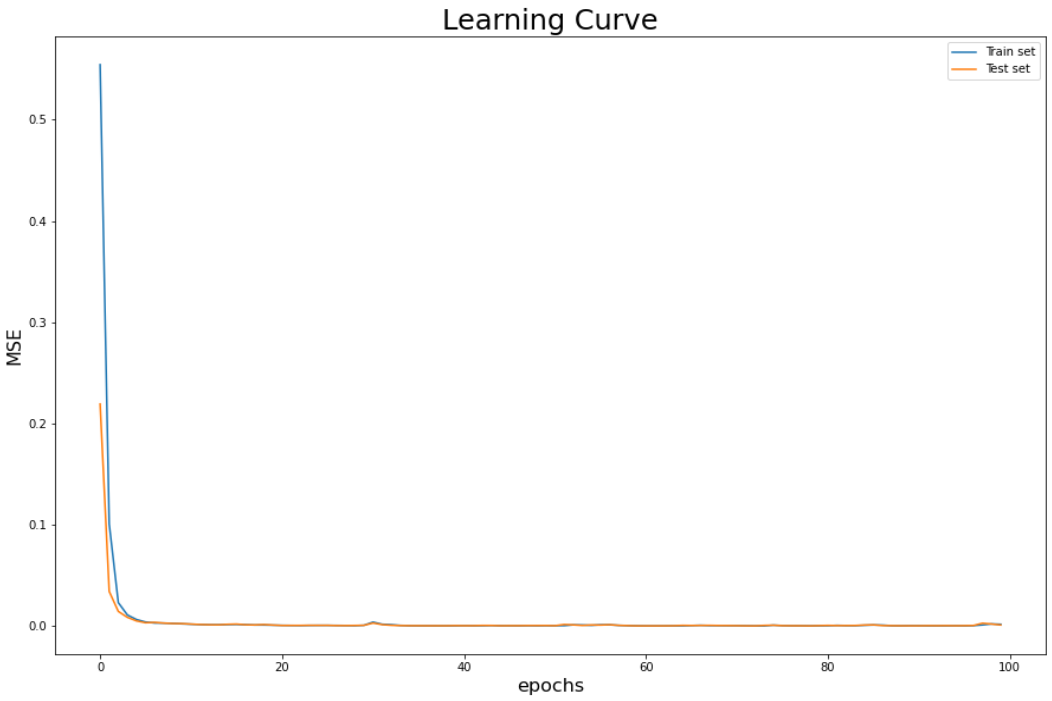

Figure 32: Mean squared error against time for the first ANN.

This diagram depicts the mean squared error of the outputs calculated by the ANN throughout time (epochs). As can be seen from the two overlapping lines the ANN adapts extremely quickly and presents very high accuracy. This is also confirmed by the calculation of the mean squared error (MSE) as well as the coefficient of determination (R^2) for each output (q):

Table 13: Mean squared error and coefficient of determination for each output of the first ANN.

	q_M	q_W	q_P	q_S
MSE	0.0098	0.0021	0.0208	0.0013
R^2	0.9988	0.9995	0.9981	0.9995

The high accuracy of the model presented is expected due to the very nature of the relationship between the inputs and the outputs. Even though the equation replaced was a partial differential equation, each output is only dependant on one input. That leads to a relatively simple connection between the inputs and the outputs, which can be easily determined by the ANN.

Additionally, the accuracy of the model is examined through the diagrams of the predictions of the ANN against the outputs calculated by the model.

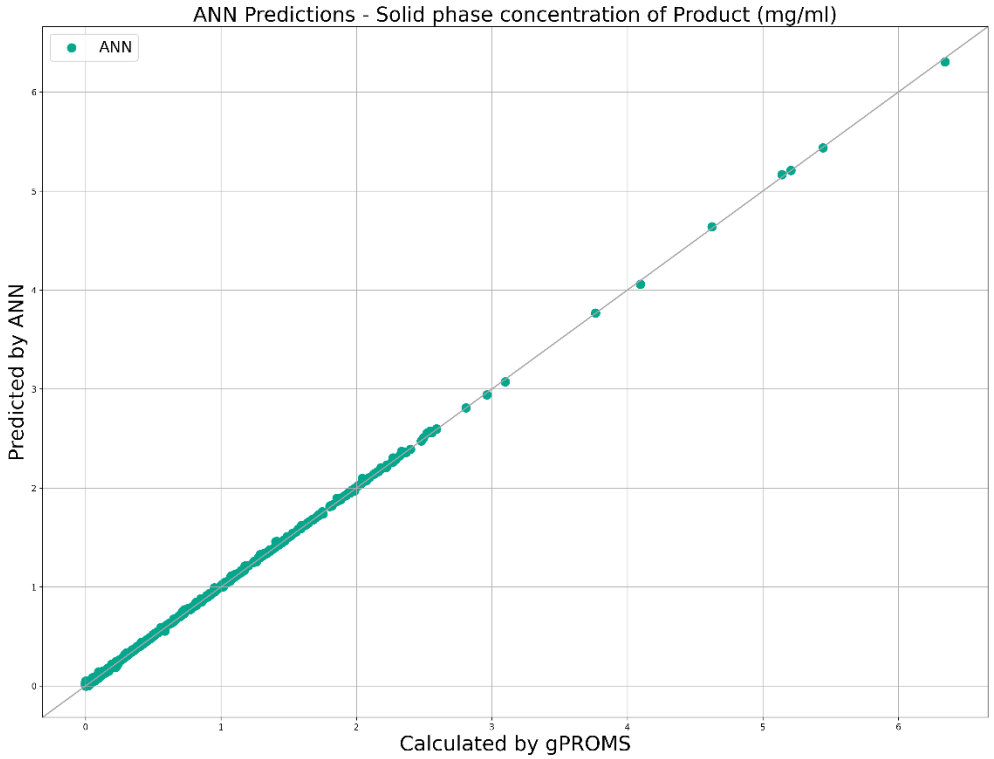

Figure 33: ANN predictions of the solid phase concentration of the product against the actual values calculated by gPROMS.

Since the diagrams for all the outputs have the same behaviour, only the one concerning the solid phase concentration of the product, which is the most important component of the process, is analysed further and the rest of the diagrams can be found in Appendix A. As can be observed from the diagram above, the predictions of the ANN coincide perfectly with the testing values calculated by gPROMS. However, these calculations along with the diagrams presented are not sufficient to validate the model.

As a final step, a new set of input values is generated by the original high-fidelity model on gPROMS and is fed to the ANN model in order to evaluate its performance. Once again, the diagram depicting the values of the solid phase concentration of the product is depicted

below, whilst the rest of the diagrams follow a similar pattern and can be found in Appendix A.

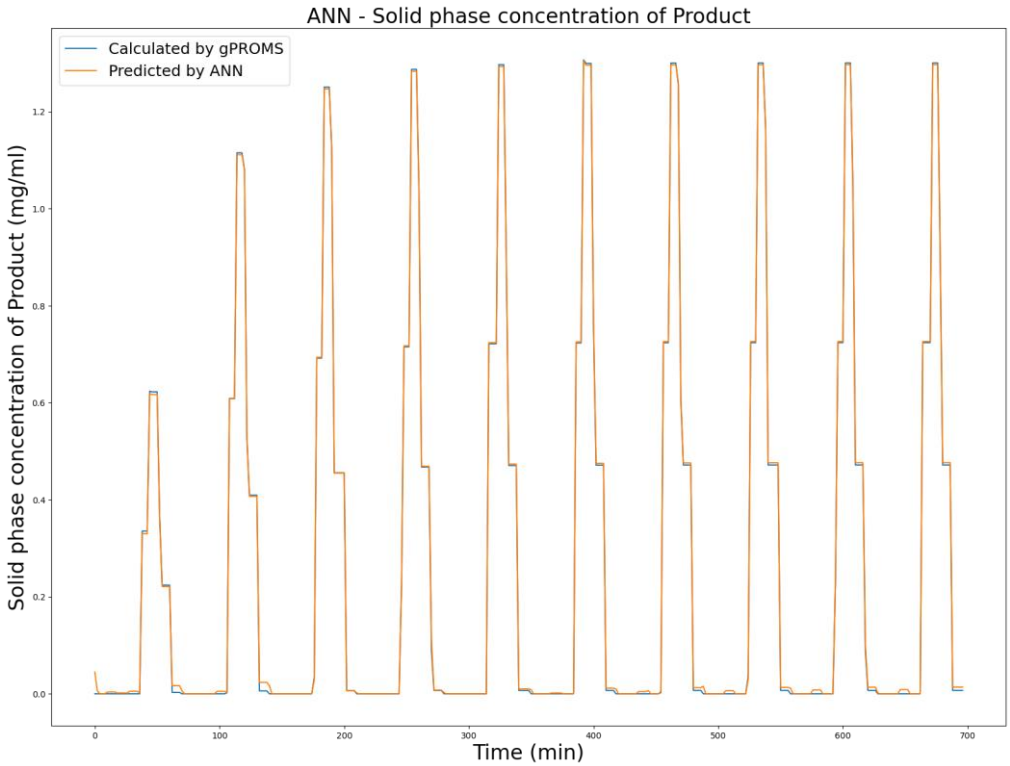

Figure 34: The solid phase concentration of the product throughout 10 process cycles calculated by the ANN and by gPROMS.

This diagram comparing the results generated by the original high-fidelity gPROMS model and by the ANN is concluding proof of the ANN's high performance. As can be easily observed, the two lines of the diagram are almost identical, with only some small deviations in concentrations with values close to zero. However, these deviations are insignificant, given their small values and the even smaller values of the corresponding concentrations.

It is therefore concluded that the ANN can safely replace the partial differential equation describing the solid phase concentration of the modifier, the weak impurities, the product, and the strong impurities., at the end of the column.

ANN for the Liquid Phase Concentrations

The second ANN trained is the one calculating the liquid phase concentration of each component. After training the ANN the first indication of how well the training was executed is the diagram presenting the learning curve:

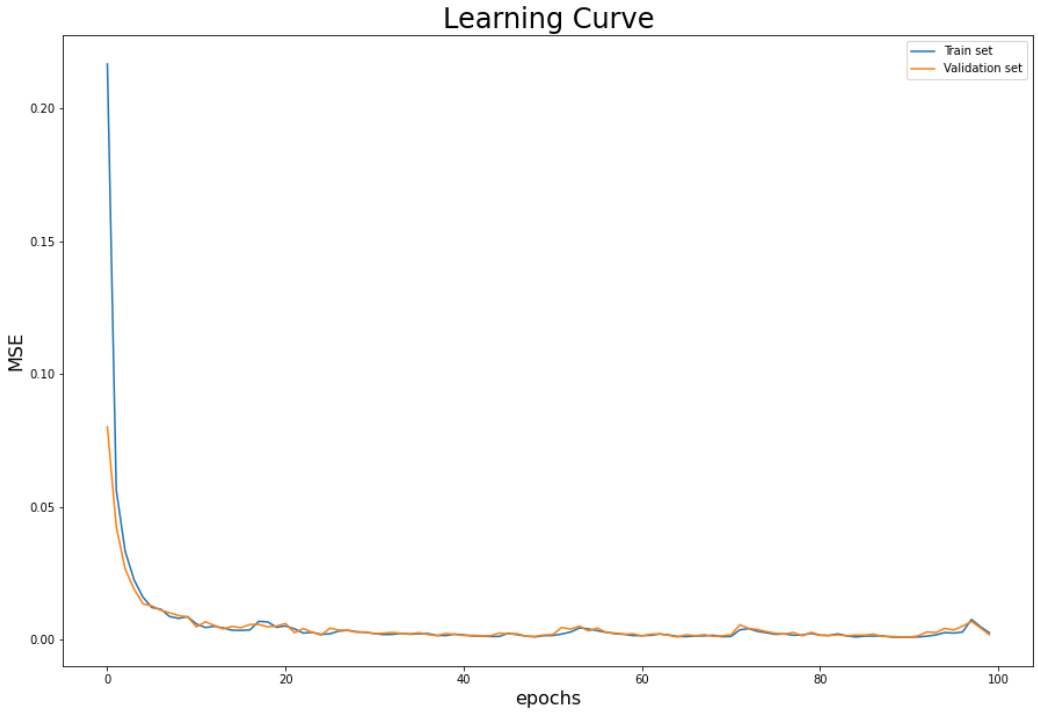

Figure 35: Mean squared error against time for the second ANN.

This diagram depicts the mean squared error of the outputs calculated by the ANN throughout time (epochs). As can be seen from the two overlapping lines the ANN adapts quickly and presents very high accuracy. This is also confirmed by the calculation of the mean squared error (MSE) as well as the coefficient of determination (R^2) for each output (q):

Table 14: Mean squared error and coefficient of determination for each output of the second ANN.

	q_M	q_W	q_P	q_S
MSE	0.0434	0.0012	0.0043	0.0005
R^2	0.9985	0.9979	0.9989	0.9981

Despite the high accuracy of the model, it appears to have a slightly larger error in its predictions compared to the first ANN. however, these errors are still small enough to be considered insignificant (less than 0.1). This small difference in prediction can be explained

by the higher complexity of the model (i.e., the larger number of inputs it receives and the various interactions between those inputs).

In order to further examine the accuracy of the model, the diagrams of the predictions of the ANN against the outputs calculated by the model are created.

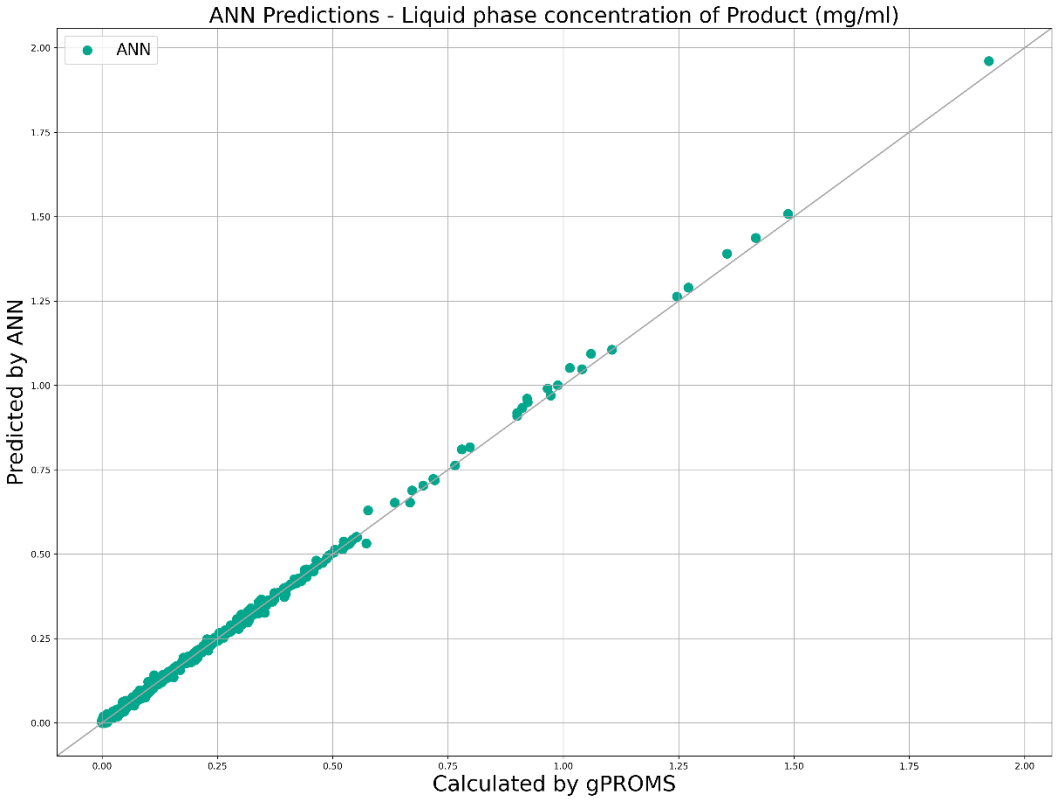

Figure 36: ANN predictions of the liquid phase concentration of the product against the actual values calculated by gPROMS.

Much like in the case of the first ANN, since the diagrams for all the outputs have the same behaviour, only the one concerning the liquid phase concentration of the product is analysed further and the rest of the diagrams can be found in Appendix B. As can be observed from the diagram above, the predictions of the ANN coincide with the testing values calculated by gPROMS, with only small deviations presented when the concentration gets higher. The impact of these deviations, however, is not clear by the prediction diagrams alone.

To assess their impact on the outputs, a new set of input values is generated by the original high-fidelity model on gPROMS and is fed to the ANN. The liquid concentration of each component at the end of one column throughout a time of 10 process cycles is presented in the following diagrams.

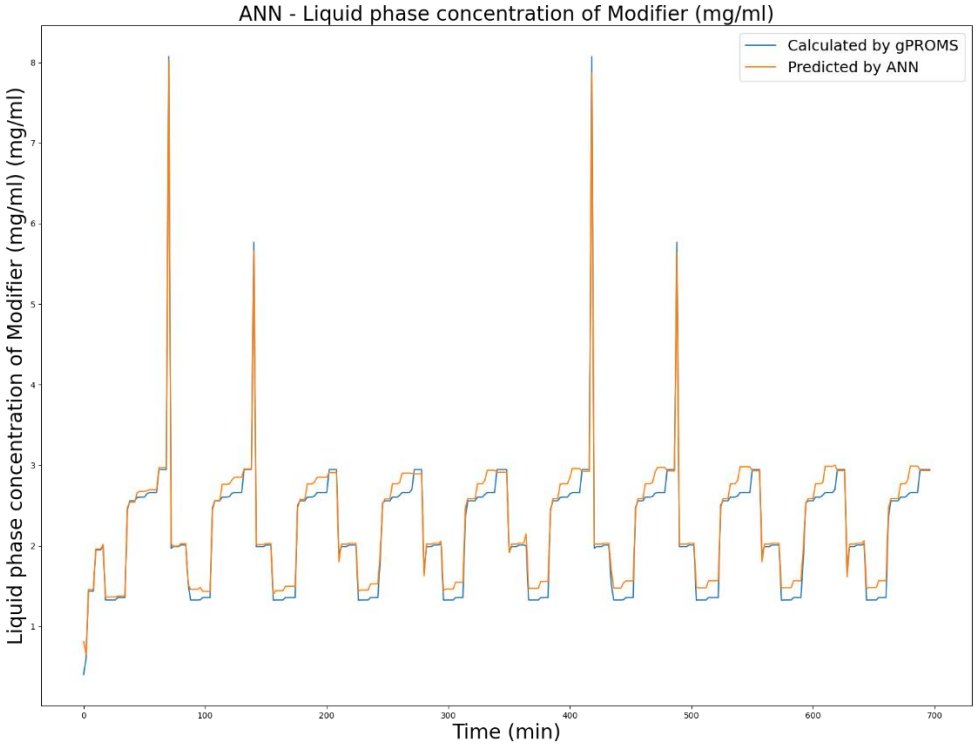

Figure 37: The liquid phase concentration of the modifier throughout 10 process cycles calculated by the ANN and by gPROMS.

The above diagram depicts the changes in the liquid concentration of the modifier throughout the process. It can be easily observed that the two lines (of the ANN prediction and the gPROMS calculations) overlap for the most part, with only some small deviations. Therefore, the ANN can safely predict the liquid concentration of the modifier, with small errors.

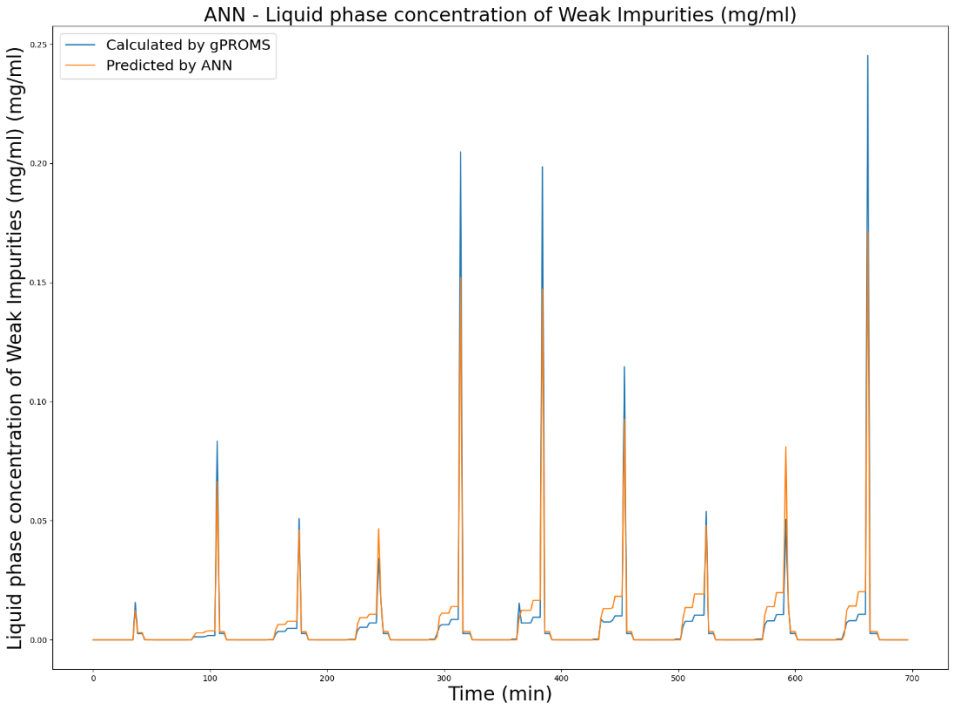

Figure 38: The liquid phase concentration of the weak impurities throughout 10 process cycles calculated by the ANN and by gPROMS.

The second diagram presented depicts the liquid phase concentration of the weak impurities as calculated by gPROMS and as predicted by the ANN. Similarly to the previous diagram, for the most part the two lines overlap. However, in this case, during the peaks in concentration there appear to be significant mispredictions. This can be explained by the irregularity of the peaks in each cycle, which increases the difficulty of training the ANN. However, the concentrations during the peaks are of high importance since they have a greater impact on the model. As a result, if this ANN were to be integrated into a hybrid model it is possible that these concentrations would cause significant problems. A solution to this problem could be a change in the number of layers used for the creation of the ANN or in the activation functions used.

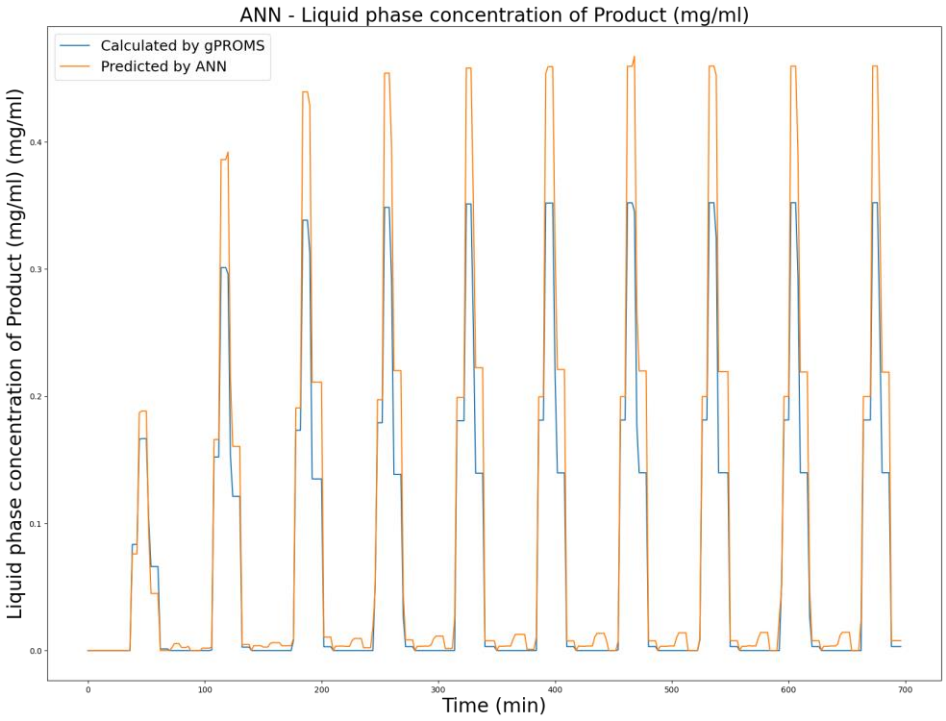

Figure 39: The liquid phase concentration of the product throughout 10 process cycles calculated by the ANN and by gPROMS.

Continuing with the predictions of the liquid phase concentration of the product at the end of the column, it is observed that the ANN predictions are significantly higher from the actual value during the peaks, i.e., during the time points of product collection. However, in this case, this cannot be attributed to irregularities in the high levels of concentration. Nonetheless, it could be explained by the wide range of values of the concentration of the product. Contrary to the impurities, the product's concentration reaches significantly high values for a very small period of time. The fact that the concentration is for the most time close to zero and then presents with a very large value, could be leading the ANN to mispredictions. To resolve this, more experiments on the structure of the ANN should be performed, as was already suggested. If, however, none of them proved to solve this issue,

another approach would be to create an ANN that would only predict the concentrations at the times of product collection. If that where the case, then unfortunately, the ANNs prospect of being integrated to a hybrid model and used for the control of the process would be reduced.

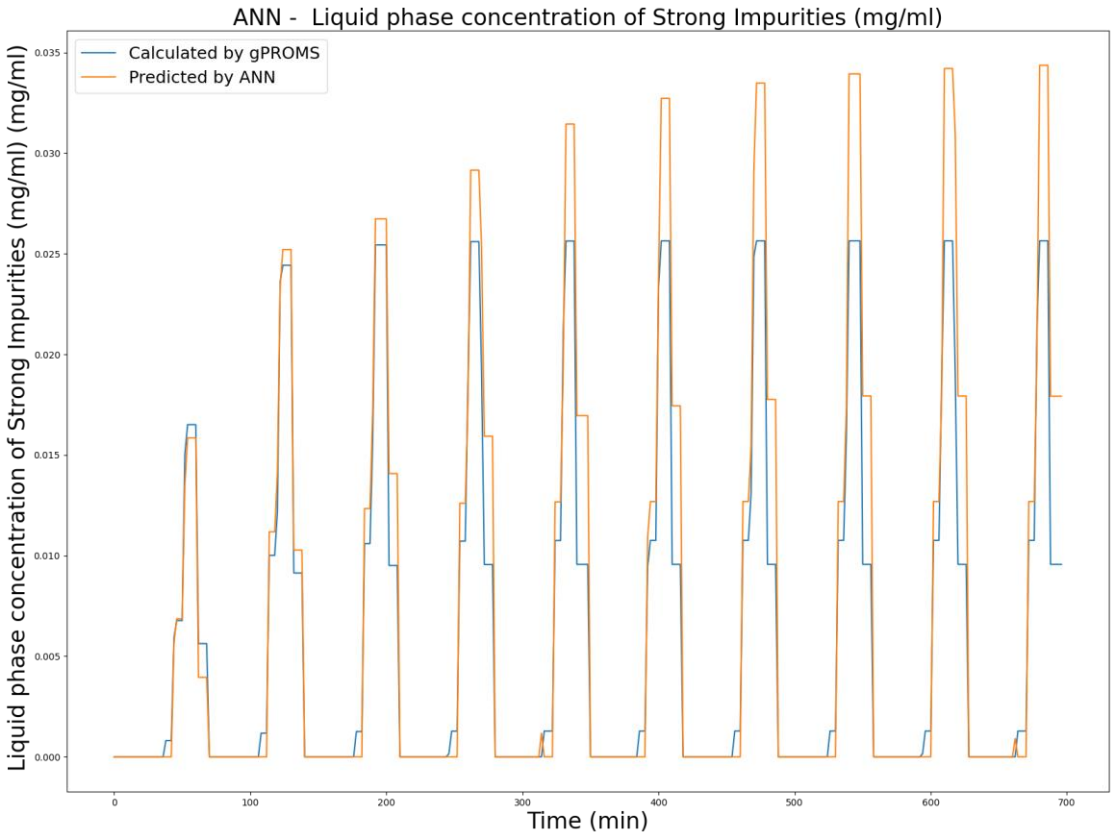

Figure 40: The liquid phase concentration of the strong impurities throughout 10 process cycles calculated by the ANN and by gPROMS.

Finally, the last diagram created depicts the liquid phase concentration of the strong impurities as calculated by the high-fidelity model and as predicted by the ANN. After looking at the ANN predictions, what is baffling in this case, is the fact that even though the actual values of the concentration are constant, after the process has reached cyclic steady state, the predicted values of the ANN show an increase in each process cycle.

It is therefore clear that, unlike the first ANN created that calculates the solid phase concentration of each component, this ANN presents many mispredictions when the values of the concentrations increase, due to the increased complexity of the inputs used and their interactions. It shows however very promising results and has the potential to provide accurate results when integrated into a hybrid model, if the larger concentration values are first regulated.

6. Conclusions and Future Work

In the present work, a high-fidelity model describing the MCSGP chromatographic process was simulated in gPROMS in order to further the understanding of the process. Through a series of sensitivity analyses an attempt was made to identify the optimal operation parameters of the process and develop a digital twin of the model, to be used for on-line applications.

In order to generate data from the high-fidelity model, a preliminary sensitivity analysis was conducted to determine the optimal discretization points of the model, for which maximum accuracy was achieved, but not at the cost of an increased computational time. The analysis showed a sufficient accuracy at 50 collocation points, which were also the collocation points used in the validation of the model, so the decision was made to utilize 50 discretization points. To complete the assessment of the model, a simulation was run to identify the cycle during which cycle steady state is reached. The results of the simulation showed that after the completion of the third cycle, the cycles become identical and therefore CSS is achieved.

Following the assessment of the model, a sensitivity analysis, consisting of five local and one global sensitivity analysis, was executed. The sensitivity analyses conducted proved the high complexity of the system and gave insight on the impact of the input variables of the system on the process, both directly and indirectly (through further interactions among the inputs). The variables with the greatest direct impact were then further studied, to identify the optimal operation spaces for each one given a certain expected level of product purity (higher than 98%) and process yield (higher than 80%). As a result, five separate design spaces where created, concerning the composition of the feed, the flowrates of the interconnected phases, the initial modifier concentrations, the column diameter, and the column porosity.

With the conclusion of the sensitivity analysis, the creation of a data-driven model of the process was attempted. Having identified the partial differential equations of the model to be replaced, the inputs of each equation were evaluated, to distinguish the actual variables from the set parameters. The complexity of the model, however, proved once again an issue due to the correlation between the two equations, and more specifically due to the utilization of the outputs of the first equation as inputs for the second. This complication created the need for two separate artificial neural networks, each one calculating only one of the equations. After the two ANNs were created, an evaluation of their performance against gPROMS's performance was executed. The first ANN, calculating the solid phase concentration of the components, showed great accuracy in the prediction of results and insignificant errors. However, the second ANN, calculating the liquid phase concentration of the components, presented significant errors in the predictions of higher concentrations. A potential solution to the errors presented, would be the creation of a third, separate ANN, calculating only the large concentrations of each component, to work alongside the second one created.

Regarding the future work to be done, the work conducted in this thesis is the first step to developing an effective on-line control model for downstream separation processes. The ANNs created show great potential in accurately predicting the concentrations of the process. By integrating them into the high-fidelity model on gPROMS and replacing the partial differential equations, a hybrid model will be created, that will provide results with great accuracy at a fraction of the original time. With the significant reduction in the computational time needed, the model will be able to provide accurate future predictions of the process that can be used to control it. Additionally, the design space identification can aid in establishing the controller's bounds, so that the purity of the product and the yield of the process are guaranteed throughout the process.

Finally, the use of the ANNs created as a modular unit in gPROMS is only one of the opportunities presented. These alternative and convenient models can be additionally connected with flowsheeting, control and optimization tasks in a large number of processes, downstream or otherwise. Based on the application and the product specs, future use of these simplified models provides with additional and apparent possibilities for model reduction, visualization, and customization.

References

[1] M. Chartrain and L. Chu, "Development and Production of Commercial Therapeutic Monoclonal Antibodies in Mammalian Cell Expression Systems: An Overview of the Current Upstream Technologies," *Curr. Pharm. Biotechnol.*, vol. 9, no. 6, pp. 447–467, Dec. 2008, doi: 10.2174/138920108786786367.

[2] "Principles of Oncologic Pharmacotherapy." https://www.cancernetwork.com/view/principles-oncologic-pharmacotherapy (accessed Jun. 04, 2021).

[3] D. M. Ecker, S. D. Jones, and H. L. Levine, "The therapeutic monoclonal antibody market," *mAbs*, vol. 7, no. 1. Landes Bioscience, pp. 9–14, Jan. 01, 2015, doi: 10.4161/19420862.2015.989042.

[4] M. Z. Siddiqui, "Monoclonal antibodies as diagnostics; An appraisal," *Indian Journal of Pharmaceutical Sciences*, vol. 72, no. 1. Wolters Kluwer -- Medknow Publications, pp. 12–17, Jan. 01, 2010, doi: 10.4103/0250-474X.62229.

[5] P. Chames, M. Van Regenmortel, E. Weiss, and D. Baty, "Therapeutic antibodies: Successes, limitations and hopes for the future," *British Journal of Pharmacology*, vol. 157, no. 2. Wiley-Blackwell, pp. 220–233, May 2009, doi: 10.1111/j.1476-5381.2009.00190.x.

[6] "Biosimilar Monoclonal Antibodies: World Market Prospects 2013-2023." Accessed: Jun. 28, 2021. [Online]. Available: www.visiongain.com.

[7] S. S. Farid, "Process economics of industrial monoclonal antibody manufacture," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, vol. 848, no. 1. J Chromatogr B Analyt Technol Biomed Life Sci, pp. 8–18, Mar. 15, 2007, doi: 10.1016/j.jchromb.2006.07.037.

[8] P. C. Taylor, A. C. Adams, M. M. Hufford, I. de la Torre, K. Winthrop, and R. L. Gottlieb, "Neutralizing monoclonal antibodies for treatment of COVID-19," *Nature Reviews Immunology*, vol. 21, no. 6. Nature Research, pp. 382–393, Jun. 01, 2021, doi: 10.1038/s41577-021-00542-x.

[9] A. Torrente-López, J. Hermosilla, N. Navas, L. Cuadros-Rodríguez, J. Cabeza, and A. Salmerón-García, "The relevance of monoclonal antibodies in the treatment of COVID-19," *Vaccines*, vol. 9, no. 6, p. 557, Jun. 2021, doi: 10.3390/vaccines9060557.

[10] J. Hummel *et al.*, "Modeling the Downstream Processing of Monoclonal Antibodies Reveals Cost Advantages for Continuous Methods for a Broad Range of Manufacturing Scales," *Biotechnol. J.*, vol. 14, no. 2, p. 1700665, Feb. 2019, doi: 10.1002/BIOT.201700665.

[11] S. Sommerfeld and J. Strube, "Challenges in biotechnology production - Generic processes and process optimization for monoclonal antibodies," *Chem. Eng. Process. Process Intensif.*, vol. 44, no. 10, pp. 1123–1137, Oct. 2005, doi:

10.1016/j.cep.2005.03.006.

[12] E. Jain and A. Kumar, “Upstream processes in antibody production: Evaluation of critical parameters,” *Biotechnology Advances*, vol. 26, no. 1. Biotechnol Adv, pp. 46–72, Jan. 2008, doi: 10.1016/j.biotechadv.2007.09.004.

[13] B. Kelley, “Industrialization of mAb production technology: The bioprocessing industry at a crossroads,” *mAbs*, vol. 1, no. 5. Landes Bioscience, pp. 443–452, 2009, doi: 10.4161/mabs.1.5.9448.

[14] F. Li, N. Vijayasankaran, A. Shen, R. Kiss, and A. Amanullah, “Cell culture processes for monoclonal antibody production,” *mAbs*, vol. 2, no. 5. Taylor & Francis, pp. 466–479, Sep. 2010, doi: 10.4161/mabs.2.5.12720.

[15] P. Gronemeyer, R. Ditz, and J. Strube, “Trends in upstream and downstream process development for antibody manufacturing,” *Bioengineering*, vol. 1, no. 4. MDPI AG, pp. 188–212, Dec. 01, 2014, doi: 10.3390/bioengineering1040188.

[16] S. Lucas *et al.*, “Production Processes for Monoclonal Antibodies Provisional chapter Production Processes for Monoclonal Antibodies,” Feb. 2017, doi: 10.5772/64263.

[17] T. I. Potgieter *et al.*, “Production of monoclonal antibodies by glycoengineered Pichia pastoris,” *J. Biotechnol.*, vol. 139, no. 4, pp. 318–325, Feb. 2009, doi: 10.1016/j.jbiotec.2008.12.015.

[18] G. Moussavou, K. Ko, J. H. Lee, and Y. K. Choo, “Production of Monoclonal Antibodies in Plants for Cancer Immunotherapy,” *BioMed Research International*, vol. 2015. Hindawi Limited, 2015, doi: 10.1155/2015/306164.

[19] F. M. Wurm, “Production of recombinant protein therapeutics in cultivated mammalian cells,” *Nature Biotechnology*, vol. 22, no. 11. Nat Biotechnol, pp. 1393–1398, Nov. 2004, doi: 10.1038/nbt1026.

[20] M. Chartrain and L. Chu, “Development and Production of Commercial Therapeutic Monoclonal Antibodies in Mammalian Cell Expression Systems: An Overview of the Current Upstream Technologies,” *Curr. Pharm. Biotechnol.*, vol. 9, no. 6, pp. 447–467, Dec. 2008, doi: 10.2174/138920108786786367.

[21] K. Keller, T. Friedmann, and A. Boxman, “The bioseparation needs for tomorrow,” *Trends in Biotechnology*, vol. 19, no. 11. Elsevier Current Trends, pp. 438–441, Nov. 01, 2001, doi: 10.1016/S0167-7799(01)01803-0.

[22] A. A. Shukla, B. Hubbard, T. Tressel, S. Guhan, and D. Low, “Downstream processing of monoclonal antibodies-Application of platform approaches,” *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, vol. 848, no. 1. Elsevier, pp. 28–39, Mar. 15, 2007, doi: 10.1016/j.jchromb.2006.09.026.

[23] I. F. Pinto, M. R. Aires-Barros, and A. M. Azevedo, “Multimodal chromatography: debottlenecking the downstream processing of monoclonal antibodies,” *Pharm. Bioprocess.*, vol. 3, no. 3, pp. 263–279, Jun. 2015, doi: 10.4155/pbp.15.7.

[24] E. V. Capela, M. R. Aires-Barros, M. G. Freire, and A. M. Azevedo, “Monoclonal Antibodies - Addressing the Challenges on the Manufacturing Processing of an Advanced Class of Therapeutic Agents,” 2017, pp. 142–203.

[25] P. A. Marichal-Gallardo and M. M. Álvarez, “State-of-the-art in downstream processing of monoclonal antibodies: Process trends in design and validation,” *Biotechnol. Prog.*, vol. 28, no. 4, pp. 899–916, Jul. 2012, doi: 10.1002/btpr.1567.

[26] K. Swinnen, A. Krul, I. Van Goidsenhoven, N. Van Tichelt, A. Roosen, and K. Van Houdt, “Performance comparison of protein A affinity resins for the purification of monoclonal antibodies,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 848, no. 1, pp. 97–107, Mar. 2007, doi: 10.1016/j.jchromb.2006.04.050.

[27] H. Hjelm, K. Hjelm, and J. Sjöquist, “Protein a from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins,” *FEBS Lett.*, vol. 28, no. 1, pp. 73–76, Nov. 1972, doi: 10.1016/0014-5793(72)80680-X.

[28] Y. Yigzaw, R. Piper, M. Tran, and A. A. Shukla, “Exploitation of the Adsorptive Properties of Depth Filters for Host Cell Protein Removal during Monoclonal Antibody Purification,” *Biotechnol. Prog.*, vol. 22, no. 1, pp. 288–296, Jan. 2006, doi: 10.1021/bp050274w.

[29] C. E. M. Hogwood, D. G. Bracewell, and C. M. Smales, “Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses,” *Current Opinion in Biotechnology*, vol. 30. Elsevier Ltd, pp. 153–160, 2014, doi: 10.1016/j.copbio.2014.06.017.

[30] J. Gomes, V. R. Chopda, and A. S. Rathore, “Integrating systems analysis and control for implementing process analytical technology in bioprocess development,” *J. Chem. Technol. Biotechnol.*, vol. 90, no. 4, pp. 583–589, Apr. 2015, doi: 10.1002/jctb.4591.

[31] G. Carta and E. X. Perez-Almodovar, “Productivity considerations and design charts for biomolecule capture with periodic countercurrent adsorption systems,” *Sep. Sci. Technol.*, vol. 45, no. 2, pp. 149–154, Jan. 2010, doi: 10.1080/01496390903423865.

[32] W. Sommeregger, B. Sissolak, K. Kandra, M. von Stosch, M. Mayer, and G. Striedner, “Quality by control: Towards model predictive control of mammalian cell culture bioprocesses,” *Biotechnology Journal*, vol. 12, no. 7. Wiley-VCH Verlag, p. 1600546, Jul. 01, 2017, doi: 10.1002/biot.201600546.

[33] A. Fuller, Z. Fan, C. Day, and C. Barlow, “Digital Twin: Enabling Technologies, Challenges and Open Research,” *IEEE Access*, vol. 8, pp. 108952–108971, Oct. 2019, doi: 10.1109/ACCESS.2020.2998358.

[34] E. Puente-Massaguer, L. Badiella, S. Gutiérrez-Granados, L. Cervera, and F. Gòdia, “A statistical approach to improve compound screening in cell culture media,” *Eng. Life Sci.*, vol. 19, no. 4, pp. 315–327, Apr. 2019, doi: 10.1002/elsc.201800168.

[35] T. Zahel *et al.*, “Integrated process modeling—A process validation life cycle companion,” *Bioengineering*, vol. 4, no. 4, p. 86, Dec. 2017, doi: 10.3390/bioengineering4040086.

[36] “Juran on Quality by Design: The New Steps for Planning Quality Into Goods ... - J. M. Juran, JOSEPH M AUTOR JURAN - Βιβλία Google.” https://books.google.co.uk/books?hl=el&lr=&id=KPUXbZ2Hw1EC&oi=fnd&pg=PR5&ots=xMUa5BF6iJ&sig=qsJ-9Hph4IBGKtjs6UGsA0JRuxg&redir_esc=y#v=onepage&q&f=false (accessed Jun. 04, 2021).

[37] R. Taticek and J. Liu, “Definitions and scope of key elements of QbD,” *AAPS Adv. Pharm. Sci. Ser.*, vol. 18, pp. 31–46, 2015, doi: 10.1007/978-1-4939-2316-8_3.

[38] S. Beg, M. S. Hasnain, M. Rahman, and S. Swain, “Introduction to Quality by Design (QbD): Fundamentals, Principles, and Applications,” in *Pharmaceutical Quality by Design*, Elsevier, 2019, pp. 1–17.

[39] “Guidance for Industry Q8(R2) Pharmaceutical Development,” 2009. Accessed: May 08, 2021. [Online]. Available: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

[40] A. Eon-Duval, H. Broly, and R. Gleixner, “Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach,” *Biotechnol. Prog.*, vol. 28, no. 3, pp. 608–622, May 2012, doi: 10.1002/btpr.1548.

[41] “Guidance for Industry Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products,” 1999. Accessed: May 08, 2021. [Online]. Available: http://www.fda.gov/cder/guidance/index.htm.

[42] C. J. Roberts, “Therapeutic protein aggregation: Mechanisms, design, and control,” *Trends in Biotechnology*, vol. 32, no. 7. Elsevier Ltd, pp. 372–380, Jul. 01, 2014, doi: 10.1016/j.tibtech.2014.05.005.

[43] W. Wang and C. J. Roberts, “Protein aggregation – Mechanisms, detection, and control,” *International Journal of Pharmaceutics*, vol. 550, no. 1–2. Elsevier B.V., pp. 251–268, Oct. 25, 2018, doi: 10.1016/j.ijpharm.2018.08.043.

[44] W. Wang, “Protein aggregation and its inhibition in biopharmaceutics,” *International Journal of Pharmaceutics*, vol. 289, no. 1–2. Elsevier, pp. 1–30, Jan. 31, 2005, doi: 10.1016/j.ijpharm.2004.11.014.

[45] J. Philo and T. Arakawa, “Mechanisms of Protein Aggregation,” *Curr. Pharm. Biotechnol.*, vol. 10, no. 4, pp. 348–351, Jun. 2009, doi: 10.2174/138920109788488932.

[46] A. S. Rosenberg, “Effects of protein aggregates: An Immunologic perspective,” *AAPS Journal*, vol. 8, no. 3. Springer, pp. E501–E507, Aug. 04, 2006, doi: 10.1208/aapsj080359.

[47] J. Vlasak and R. Ionescu, “Fragmentation of monoclonal antibodies,” *mAbs*, vol. 3, no. 3. Taylor & Francis, pp. 253–263, May 2011, doi: 10.4161/mabs.3.3.15608.

[48] H. Liu, G. Caza-Bulseco, D. Faldu, C. Chumsae, and J. Sun, “Heterogeneity of monoclonal

antibodies," *Journal of Pharmaceutical Sciences*, vol. 97, no. 7. John Wiley and Sons Inc., pp. 2426–2447, Jul. 01, 2008, doi: 10.1002/jps.21180.

[49] L. W. Dick, D. Qiu, D. Mahon, M. Adamo, and K. C. Cheng, "C-terminal lysine variants in fully human monoclonal antibodies: Investigation of test methods and possible causes," *Biotechnol. Bioeng.*, vol. 100, no. 6, pp. 1132–1143, Aug. 2008, doi: 10.1002/bit.21855.

[50] W. Zhang, S. Xiao, and D. U. Ahn, "Protein Oxidation: Basic Principles and Implications for Meat Quality," *Critical Reviews in Food Science and Nutrition*, vol. 53, no. 11. Crit Rev Food Sci Nutr, pp. 1191–1201, Jan. 2013, doi: 10.1080/10408398.2011.577540.

[51] I. J. Del Val, C. Kontoravdi, and J. M. Nagy, "Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns," *Biotechnol. Prog.*, vol. 26, no. 6, pp. 1505–1527, Nov. 2010, doi: 10.1002/btpr.470.

[52] D. A. Cumming, "Glycosylation of recombinant protein therapeutics: Control and functional implications," *Glycobiology*, vol. 1, no. 2. Oxford Academic, pp. 115–130, Mar. 01, 1991, doi: 10.1093/glycob/1.2.115.

[53] V. Kayser, N. Chennamsetty, V. Voynov, K. Forrer, B. Helk, and B. L. Trout, "Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies," *Biotechnol. J.*, vol. 6, no. 1, pp. 38–44, Jan. 2011, doi: 10.1002/biot.201000091.

[54] T. S. Raju and B. Scallon, "Fc glycans terminated with N-acetylglucosamine residues increase antibody resistance to papain," *Biotechnol. Prog.*, vol. 23, no. 4, pp. 964–971, Jul. 2007, doi: 10.1021/bp070118k.

[55] M. H. Tao and S. L. Morrison, "Studies of aglycosylated chimeric mouse-human IgG. Role of carbohydrate in the structure and effector functions mediated by the human IgG constant region.," *J. Immunol.*, vol. 143, no. 8, 1989.

[56] H. Liu, G. G. Bulseco, and J. Sun, "Effect of posttranslational modifications on the thermal stability of a recombinant monoclonal antibody," *Immunol. Lett.*, vol. 106, no. 2, pp. 144–153, Aug. 2006, doi: 10.1016/j.imlet.2006.05.011.

[57] S. T. Jung, T. H. Kang, W. Kelton, and G. Georgiou, "Bypassing glycosylation: Engineering aglycosylated full-length IgG antibodies for human therapy," *Current Opinion in Biotechnology*, vol. 22, no. 6. Elsevier Current Trends, pp. 858–867, Dec. 01, 2011, doi: 10.1016/j.copbio.2011.03.002.

[58] T. S. Raju, "Terminal sugars of Fc glycans influence antibody effector functions of IgGs," *Current Opinion in Immunology*, vol. 20, no. 4. Elsevier Current Trends, pp. 471–478, Aug. 01, 2008, doi: 10.1016/j.coi.2008.06.007.

[59] L. Dissing-Olesen, M. Thaysen-Andersen, M. Meldgaard, P. Højrup, and B. Finsen, "The function of the human interferon-β1a glycan determined in vivo," *J. Pharmacol. Exp. Ther.*, vol. 326, no. 1, pp. 338–347, Jul. 2008, doi: 10.1124/jpet.108.138263.

[60] C. T. Yuen *et al.*, “Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures,” *Br. J. Haematol.*, vol. 121, no. 3, pp. 511–526, May 2003, doi: 10.1046/j.1365-2141.2003.04307.x.

[61] A. G. Morell, G. Gregoriadis, I. H. Scheinberg, J. Hickman, and G. Ashwell, “The role of sialic acid in determining the survival of glycoproteins in the circulation.,” *J. Biol. Chem.*, vol. 246, no. 5, pp. 1461–1467, Mar. 1971, doi: 10.1016/s0021-9258(19)76994-4.

[62] R. J. Solá and K. Griebenow, “Glycosylation of therapeutic proteins: An effective strategy to optimize efficacy,” *BioDrugs*, vol. 24, no. 1. Springer, pp. 9–21, Aug. 16, 2010, doi: 10.2165/11530550-000000000-00000.

[63] R. Schauer, “Achievements and challenges of sialic acid research,” *Glycoconjugate Journal*, vol. 17, no. 7–9. Springer, pp. 485–499, 2000, doi: 10.1023/A:1011062223612.

[64] B. Byrne, G. G. Donohoe, and R. O’Kennedy, “Sialic acids: carbohydrate moieties that influence the biological and physical properties of biopharmaceutical proteins and living cells,” *Drug Discovery Today*, vol. 12, no. 7–8. Elsevier Current Trends, pp. 319–326, Apr. 01, 2007, doi: 10.1016/j.drudis.2007.02.010.

[65] Y. Miwa *et al.*, “Are N-glycolylneuraminic acid (Hanganutziu-Deicher) antigens important in pig-to-human xenotransplantation?,” *Xenotransplantation*, vol. 11, no. 3, pp. 247–253, May 2004, doi: 10.1111/j.1399-3089.2004.00126.x.

[66] N. Jenkins, R. B. Parekh, and D. C. James, “Getting the glycosylation right: Implications for the biotechnology industry,” *Nat. Biotechnol.*, vol. 14, no. 8, pp. 975–981, 1996, doi: 10.1038/nbt0896-975.

[67] A. Noguchi, C. J. Mukuria, E. Suzuki, and M. Naiki, “Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in chinese hamster ovary cells,” *J. Biochem.*, vol. 117, no. 1, pp. 59–62, Jan. 1995, doi: 10.1093/oxfordjournals.jbchem.a124721.

[68] A. R. Flesher, J. A. Marzowski, W. -C Wang, and H. V. Raff, “Fluorophore-labeled carbohydrate analysis of immunoglobulin fusion proteins: Correlation of oligosaccharide content with in vivo clearance profile,” *Biotechnol. Bioeng.*, vol. 46, no. 5, pp. 399–407, 1995, doi: 10.1002/bit.260460502.

[69] H. LEIBIGER, D. WÜSTNER, R.-D. STIGLER, and U. MARX, “Variable domain-linked oligosaccharides of a human monoclonal IgG: structure and influence on antigen binding,” *Biochem. J.*, vol. 338, no. 2, pp. 529–538, Mar. 1999, doi: 10.1042/bj3380529.

[70] C. Quan *et al.*, “A study in glycation of a therapeutic recombinant humanized monoclonal antibody: Where it is, how it got there, and how it affects charge-based behavior,” *Anal. Biochem.*, vol. 373, no. 2, pp. 179–191, Feb. 2008, doi: 10.1016/j.ab.2007.09.027.

[71] K. Nakajou, H. Watanabe, U. Kragh-Hansen, T. Maruyama, and M. Otagiri, “The effect of glycation on the structure, function and biological fate of human serum albumin as revealed by recombinant mutants,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1623, no.

2–3, pp. 88–97, Oct. 2003, doi: 10.1016/j.bbagen.2003.08.001.

[72] L. E. M. Fernández, D. E. Kalume, L. Calvo, M. Fernández Mallo, A. Vallin, and P. Roepstorff, "Characterization of a recombinant monoclonal antibody by mass spectrometry combined with liquid chromatography," *J. Chromatogr. B Biomed. Sci. Appl.*, vol. 752, no. 2, pp. 247–261, Mar. 2001, doi: 10.1016/S0378-4347(00)00503-X.

[73] Z. Wei, E. Shacter, M. Schenerman, J. Dougherty, and L. McLeod, "The Role of Higher-Order Structure in Defining Biopharmaceutical Quality," 2011.

[74] C. Maas, S. Hermeling, B. Bouma, W. Jiskoot, and M. F. B. G. Gebbink, "A role for protein misfolding in immunogenicity of biopharmaceuticals," *J. Biol. Chem.*, vol. 282, no. 4, pp. 2229–2236, Jan. 2007, doi: 10.1074/jbc.M605984200.

[75] W. Zhang and M. J. Czupryn, "Free sulfhydryl in recombinant monoclonal antibodies," *Biotechnol. Prog.*, vol. 18, no. 3, pp. 509–513, Jan. 2002, doi: 10.1021/bp025511z.

[76] J. Schuurman, G. J. Perdok, A. D. Gorter, and R. C. Aalberse, "The inter-heavy chain disulfide bonds of IgG4 are in equilibrium with intra-chain disulfide bonds," *Mol. Immunol.*, vol. 38, no. 1, pp. 1–8, Jan. 2001, doi: 10.1016/S0161-5890(01)00050-5.

[77] F. R. Taylor, H. L. Prentice, E. A. Garber, H. A. Fajardo, E. Vasilyeva, and R. Blake Pepinsky, "Suppression of sodium dodecyl sulfate-polyacrylamide gel electrophoresis sample preparation artifacts for analysis of IgG4 half-antibody," *Anal. Biochem.*, vol. 353, no. 2, pp. 204–208, Jun. 2006, doi: 10.1016/j.ab.2006.02.022.

[78] S. Angal *et al.*, "A single amino acid substitution abolishes the heterogeneity of chimeric mouse/human (IgG4) antibody," *Mol. Immunol.*, vol. 30, no. 1, pp. 105–108, Jan. 1993, doi: 10.1016/0161-5890(93)90432-B.

[79] "Annex 4 Requirements for the use of animal cells as in vitro substrates for the production of biologicals (Addendum 2003)."

[80] D. E. Wierenga, J. Cogan, and J. C. Petricciani, "Administration of tumor cell chromatin to immunosuppressed and non-immunosuppressed non-human primates," *Biologicals*, vol. 23, no. 3, pp. 221–224, Sep. 1995, doi: 10.1006/biol.1995.0036.

[81] K. Buttenschoen, P. Radermacher, and H. Bracht, "Endotoxin elimination in sepsis: physiology and therapeutic application," *Langenbeck's Arch. Surg.*, vol. 395, no. 6, pp. 597–605, 2010, doi: 10.1007/s00423-010-0658-6.

[82] A. J. Shepherd, N. J. Wilson, and K. T. Smith, "Characterisation of endogenous retrovirus in rodent cell lines used for production of biologicals," *Biologicals*, vol. 31, no. 4, pp. 251–260, Dec. 2003, doi: 10.1016/S1045-1056(03)00065-4.

[83] "INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE PRECLINICAL SAFETY EVALUATION OF BIOTECHNOLOGY-DERIVED PHARMACEUTICALS S6(R1)."

[84] J. Maguire and D. Peng, "How to Identify Critical Quality Attributes and Critical Process Parameters," 2015.

[85] M. P. Ochoa, A. Deshpande, S. García-Muñoz, S. Stamatis, and I. E. Grossmann, “Flexibility Analysis For Design Space Definition,” in *Computer Aided Chemical Engineering*, vol. 47, Elsevier B.V., 2019, pp. 323–328.

[86] “Quality-by-Design as Applied to the Development and Manufacturing of a Lyophilized Protein Product | American Pharmaceutical Review - The Review of American Pharmaceutical Business & Technology.” https://www.americanpharmaceuticalreview.com/Featured-Articles/117782-Quality-by-Design-as-Applied-to-the-Development-and-Manufacturing-of-a-Lyophilized-Protein-Product/ (accessed Jun. 04, 2021).

[87] A. S. Myerson, M. Krumme, M. Nasr, H. Thomas, and R. D. Braatz, “Control Systems Engineering in Continuous Pharmaceutical Manufacturing Continuous Manufacturing Symposium,” doi: 10.1002/jps.24311.

[88] A. S. Rathore and H. Winkle, “Quality by design for biopharmaceuticals,” *Nature Biotechnology*, vol. 27, no. 1. Nature Publishing Group, pp. 26–34, Jan. 25, 2009, doi: 10.1038/nbt0109-26.

[89] G. Barringer, “Downstream Process Optimization Opportunities Using On-Line and At-Line PAT Instrumentation,” *BioPharm Int.*, vol. 2006 Supplement, no. 3, Feb. 2006, Accessed: Jun. 27, 2021. [Online]. Available: https://www.biopharminternational.com/view/downstream-process-optimization-opportunities-using-line-and-line-pat-instrumentation.

[90] V. Kumar, A. Bhalla, and A. S. Rathore, “Design of experiments applications in bioprocessing: Concepts and approach,” *Biotechnol. Prog.*, vol. 30, no. 1, pp. 86–99, Jan. 2014, doi: 10.1002/btpr.1821.

[91] L. Zhang and S. Mao, “Application of quality by design in the current drug development,” *Asian Journal of Pharmaceutical Sciences*, vol. 12, no. 1. Shenyang Pharmaceutical University, pp. 1–8, Jan. 01, 2017, doi: 10.1016/j.ajps.2016.07.006.

[92] S. J. Kalil, F. Maugeri, and M. I. Rodrigues, “Response surface analysis and simulation as a tool for bioprocess design and optimization,” *Process Biochem.*, vol. 35, no. 6, pp. 539–550, Jan. 2000, doi: 10.1016/S0032-9592(99)00101-6.

[93] S. Craven, J. Whelan, and B. Glennon, “Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller,” *J. Process Control*, vol. 24, no. 4, pp. 344–357, Apr. 2014, doi: 10.1016/j.jprocont.2014.02.007.

[94] J. Shlens, “A Tutorial on Principal Component Analysis,” Apr. 2014, Accessed: Jun. 27, 2021. [Online]. Available: http://arxiv.org/abs/1404.1100.

[95] P. Kadlec, B. Gabrys, and S. Strandt, “Data-driven Soft Sensors in the process industry,” *Computers and Chemical Engineering*, vol. 33, no. 4. pp. 795–814, Apr. 21, 2009, doi: 10.1016/j.compchemeng.2008.12.012.

[96] L. Mears, S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey, “Mechanistic Fermentation Models for Process Design, Monitoring, and Control,” *Trends in Biotechnology*, vol. 35, no. 10. Elsevier Ltd, pp. 914–924, Oct. 01, 2017, doi:

10.1016/j.tibtech.2017.07.002.

[97] C. Seng Yue and M. Ducharme, “Empirical Models, Mechanistic Models, Statistical Moments, and Noncompartmental Analysis,” 2015, pp. 817–849.

[98] P. Kotidis *et al.*, “Model-based optimization of antibody galactosylation in CHO cell culture,” *Biotechnol. Bioeng.*, vol. 116, no. 7, pp. 1612–1626, Jul. 2019, doi: 10.1002/bit.26960.

[99] Z. Amribt, H. Niu, and P. Bogaerts, “Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures,” *Biochem. Eng. J.*, vol. 70, no. 70, pp. 196–209, Jan. 2013, doi: 10.1016/j.bej.2012.11.005.

[100] H. Shirahata, S. Diab, H. Sugiyama, and D. I. Gerogiorgis, “Dynamic modelling, simulation and economic evaluation of two CHO cell-based production modes towards developing biopharmaceutical manufacturing processes,” *Chem. Eng. Res. Des.*, vol. 150, pp. 218–233, Oct. 2019, doi: 10.1016/j.cherd.2019.07.016.

[101] T. Müller-Späth *et al.*, “Model simulation and experimental verification of a cation-exchange IgG capture step in batch and continuous chromatography,” *J. Chromatogr. A*, vol. 1218, no. 31, pp. 5195–5204, Aug. 2011, doi: 10.1016/j.chroma.2011.05.103.

[102] R. Hahn, P. Bauerhansl, K. Shimahara, C. Wizniewski, A. Tscheliessnig, and A. Jungbauer, “Comparison of protein A affinity sorbents: II. Mass transfer properties,” *J. Chromatogr. A*, vol. 1093, no. 1–2, pp. 98–110, Nov. 2005, doi: 10.1016/j.chroma.2005.07.050.

[103] F. Steinebach, M. Angarita, D. J. Karst, T. Müller-Späth, and M. Morbidelli, “Model based adaptive control of a continuous capture process for monoclonal antibodies production,” *J. Chromatogr. A*, vol. 1444, pp. 50–56, Apr. 2016, doi: 10.1016/j.chroma.2016.03.014.

[104] M. Grom, M. Kozorog, S. Caserman, A. Pohar, and B. Likozar, “Protein A affinity chromatography of Chinese hamster ovary (CHO) cell culture broths containing biopharmaceutical monoclonal antibody (mAb): Experiments and mechanistic transport, binding and equilibrium modeling,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 1083, pp. 44–56, Apr. 2018, doi: 10.1016/j.jchromb.2018.02.032.

[105] E. J. Close, J. R. Salm, D. G. Bracewell, and E. Sorensen, “Modelling of industrial biopharmaceutical multicomponent chromatography,” *Chem. Eng. Res. Des.*, vol. 92, no. 7, pp. 1304–1314, 2014, doi: 10.1016/j.cherd.2013.10.022.

[106] L. Aumann and M. Morbidelli, “A continuous multicolumn countercurrent solvent gradient purification (MCSGP) process,” *Biotechnol. Bioeng.*, vol. 98, no. 5, pp. 1043–1055, Dec. 2007, doi: 10.1002/bit.21527.

[107] F. Boukouvala, F. J. Muzzio, and M. G. Ierapetritou, “Dynamic data-driven modeling of pharmaceutical processes,” *Ind. Eng. Chem. Res.*, vol. 50, no. 11, pp. 6743–6754, Jun. 2011, doi: 10.1021/ie102305a.

[108] J. E. Tabora, “Data driven modelling and control of batch processes in the

pharmaceutical industry," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, Jan. 2012, vol. 8, no. PART 1, pp. 708–714, doi: 10.3182/20120710-4-SG-2026.00204.

[109] M. Rüdt, N. Brestrich, L. Rolinger, and J. Hubbuch, "Real-time monitoring and control of the load phase of a protein A capture step," *Biotechnol. Bioeng.*, vol. 114, no. 2, pp. 368–373, Feb. 2017, doi: 10.1002/bit.26078.

[110] N. Walch *et al.*, "Prediction of the Quantity and Purity of an Antibody Capture Process in Real Time," *Biotechnol. J.*, vol. 14, no. 7, Jul. 2019, doi: 10.1002/biot.201800521.

[111] V. S. Joshi, V. Kumar, and A. S. Rathore, "Optimization of ion exchange sigmoidal gradients using hybrid models: Implementation of quality by design in analytical method development," *J. Chromatogr. A*, vol. 1491, pp. 145–152, Mar. 2017, doi: 10.1016/j.chroma.2017.02.058.

[112] S. Liu and L. G. Papageorgiou, "Optimal Antibody Purification Strategies Using Data-Driven Models," *Engineering*, vol. 5, no. 6, pp. 1077–1092, Dec. 2019, doi: 10.1016/j.eng.2019.10.011.

[113] von M. Stosch, "Newcastle University ePrints Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry," *Biotechnol. J.*, 2014, doi: 10.1002/biot.201300385.

[114] B. D. Ripley, *Pattern recognition and neural networks*. Cambridge University Press, 2014.

[115] B. Bayer, B. Sissolak, M. Duerkop, M. von Stosch, and G. Striedner, "The shortcomings of accurate rate estimations in cultivation processes and a solution for precise and robust process modeling," *Bioprocess Biosyst. Eng.*, vol. 43, no. 2, pp. 169–178, Feb. 2020, doi: 10.1007/s00449-019-02214-6.

[116] B. Bayer, M. von Stosch, G. Striedner, and M. Duerkop, "Comparison of Modeling Methods for DoE-Based Holistic Upstream Process Characterization," *Biotechnol. J.*, vol. 15, no. 5, p. 1900551, May 2020, doi: 10.1002/biot.201900551.

[117] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, pp. 1–8, Feb. 2006, doi: 10.1186/1471-2105-7-91.

[118] H. Narayanan *et al.*, "Hybrid-EKF: Hybrid model coupled with extended Kalman filter for real-time monitoring and control of mammalian cell culture," *Biotechnol. Bioeng.*, vol. 117, no. 9, pp. 2703–2714, Sep. 2020, doi: 10.1002/bit.27437.

[119] P. Kotidis and C. Kontoravdi, "Harnessing the potential of artificial neural networks for predicting protein glycosylation," *Metab. Eng. Commun.*, vol. 10, p. e00131, Jun. 2020, doi: 10.1016/j.mec.2020.e00131.

[120] H. Narayanan, T. Seidler, M. F. Luna, M. Sokolov, M. Morbidelli, and A. Butté, "Hybrid Models for the simulation and prediction of chromatographic processes for protein capture," *J. Chromatogr. A*, vol. 1650, p. 462248, Aug. 2021, doi: 10.1016/j.chroma.2021.462248.

[121] X. Zhou and H. Lin, “Local Sensitivity Analysis,” in *Encyclopedia of GIS*, Springer US, 2008, pp. 616–616.

[122] X. Zhou, H. Lin, and H. Lin, “Global Sensitivity Analysis,” in *Encyclopedia of GIS*, Springer US, 2008, pp. 408–409.

[123] G. Li, S. W. Wang, and H. Rabitz, “Practical approaches to construct RS-HDMR component functions,” *J. Phys. Chem. A*, vol. 106, no. 37, pp. 8721–8733, Sep. 2002, doi: 10.1021/jp014567t.

[124] I. M. Sobol, “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates,” *Math. Comput. Simul.*, vol. 55, no. 1–3, pp. 271–280, Feb. 2001, doi: 10.1016/S0378-4754(00)00270-6.

[125] D. Solomatine, L. M. See, and R. J. Abrahart, “Data-Driven Modelling: Concepts, Approaches and Experiences,” in *Practical Hydroinformatics*, Springer Berlin Heidelberg, 2008, pp. 17–30.

[126] M. Krättli, T. Müller-Späth, and M. Morbidelli, “Multifraction separation in countercurrent chromatography (MCSGP),” *Biotechnol. Bioeng.*, vol. 110, no. 9, pp. 2436–2444, Sep. 2013, doi: 10.1002/bit.24901.

[127] M. Krättli, F. Steinebach, and M. Morbidelli, “Online control of the twin-column countercurrent solvent gradient process for biochromatography,” *J. Chromatogr. A*, vol. 1293, pp. 51–59, Jun. 2013, doi: 10.1016/j.chroma.2013.03.069.

[128] T. Müller-Späth, G. Ströhlein, O. Lyngberg, and D. Maclean, “Enabling high purities and yields in therapeutic peptide purification using multicolumn countercurrent solvent gradient purification.”

[129] F. Steinebach, N. Ulmer, L. Decker, L. Aumann, and M. Morbidelli, “Experimental design of a twin-column countercurrent gradient purification process,” *J. Chromatogr. A*, vol. 1492, pp. 19–26, Apr. 2017, doi: 10.1016/j.chroma.2017.02.049.

[130] M. M. Papathanasiou *et al.*, “Advanced control strategies for the multicolumn countercurrent solvent gradient purification process,” *AIChE J.*, vol. 62, no. 7, pp. 2341–2357, Jul. 2016, doi: 10.1002/aic.15203.

[131] T. Müller-Späth, L. Aumann, L. Melter, G. Ströhlein, and M. Morbidelli, “Chromatographic separation of three monoclonal antibody variants using multicolumn countercurrent solvent gradient purification (MCSGP),” *Biotechnol. Bioeng.*, vol. 100, no. 6, pp. 1166–1177, Aug. 2008, doi: 10.1002/bit.21843.

[132] O. J. Smith and A. W. Weeterberg, “Acceleration of Cyclic Steady State Convergence for Pressure Swing Adsorption Models,” *Ind. Eng. Chem. Res.*, vol. 31, no. 6, pp. 1569–1573, Jun. 1992, doi: 10.1021/ie00006a021.

Appendix A: First ANN Prediction and Validation Diagrams

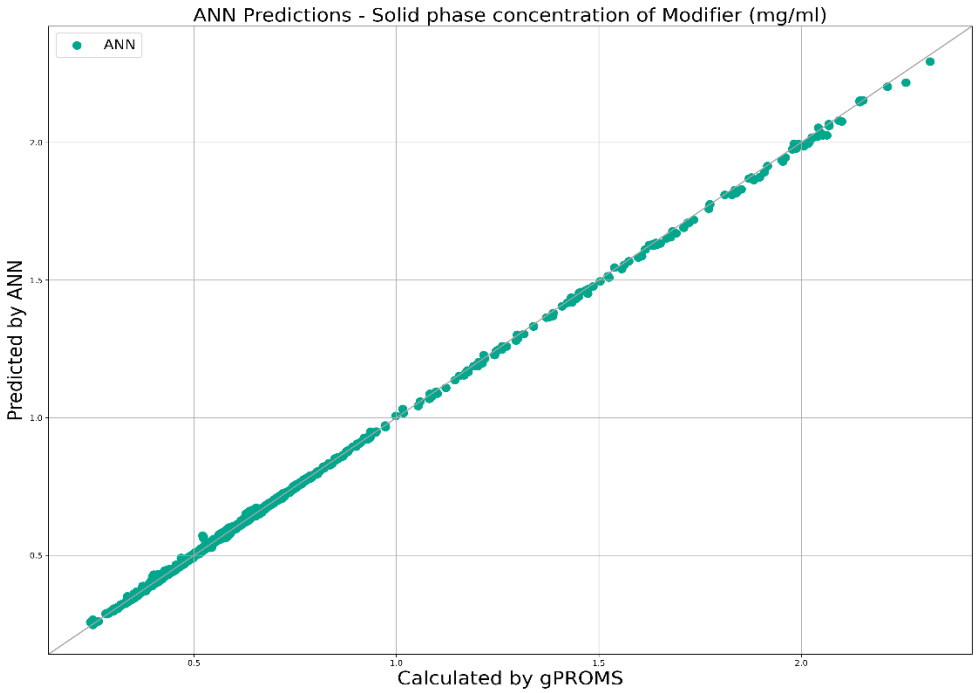

Figure 41: ANN predictions of the solid phase concentration of the modifier against the actual values calculated by gPROMS.

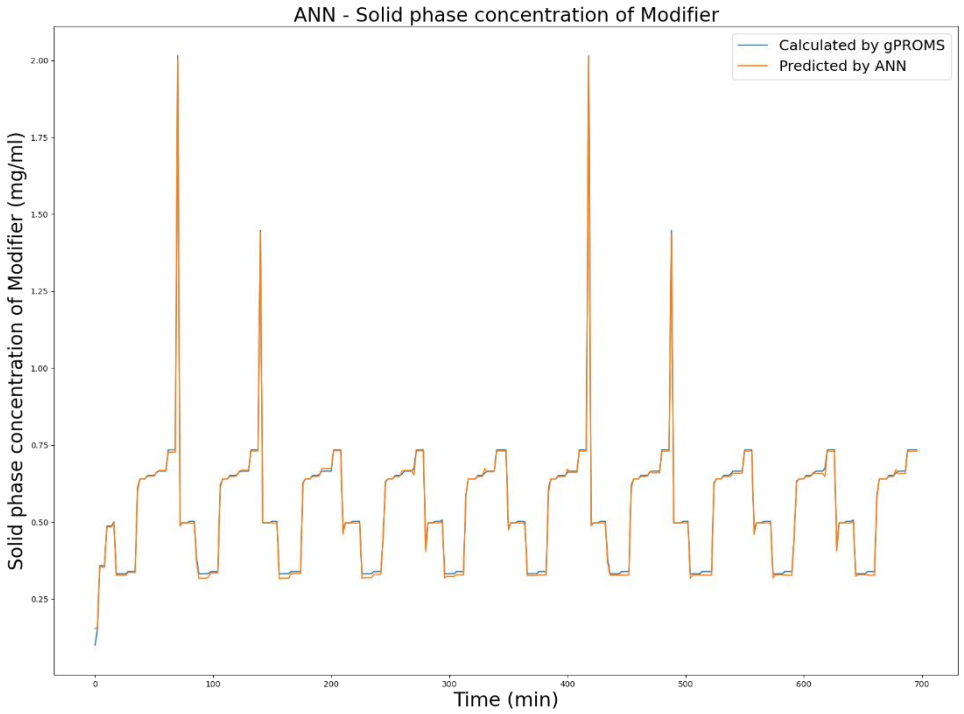

Figure 42: The solid phase concentration of the modifier throughout 10 process cycles calculated by the ANN and by gPROMS.

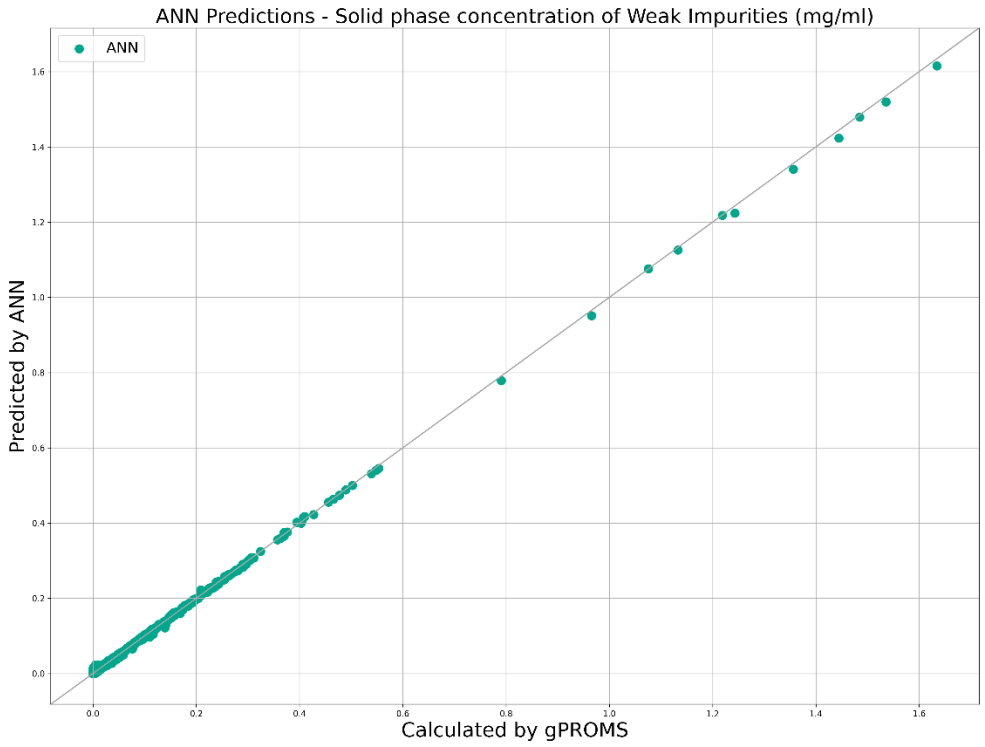

Figure 43: ANN predictions of the solid phase concentration of the weak impurities against the actual values calculated by gPROMS.

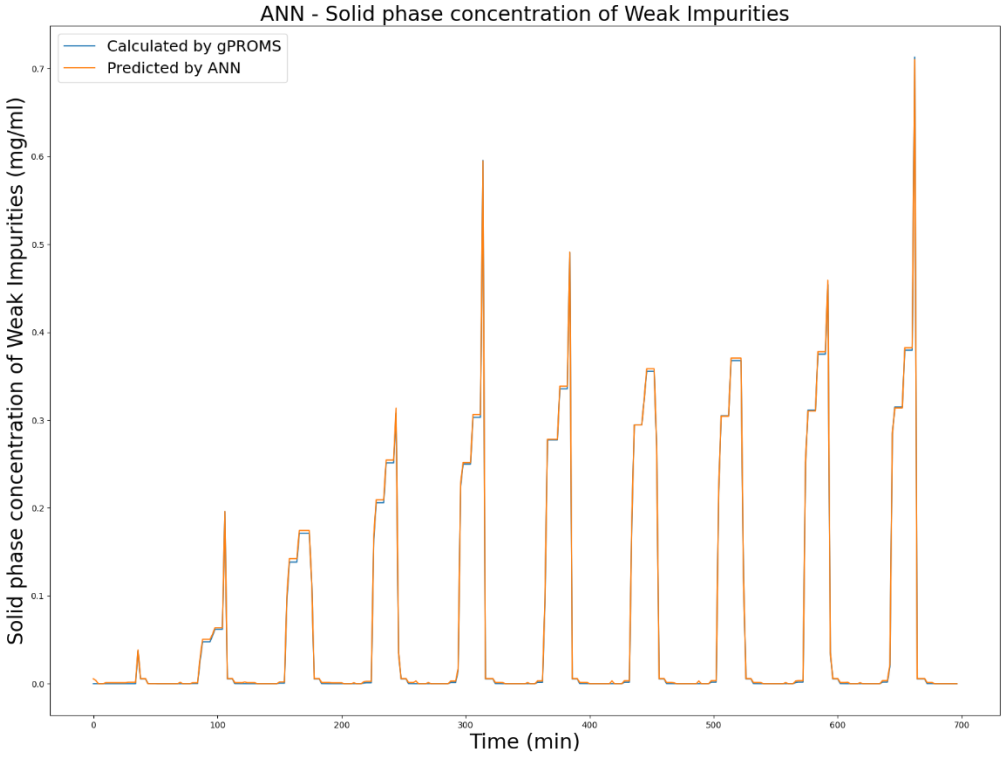

Figure 44: The solid phase concentration of the weak impurities throughout 10 process cycles calculated by the ANN and by gPROMS.

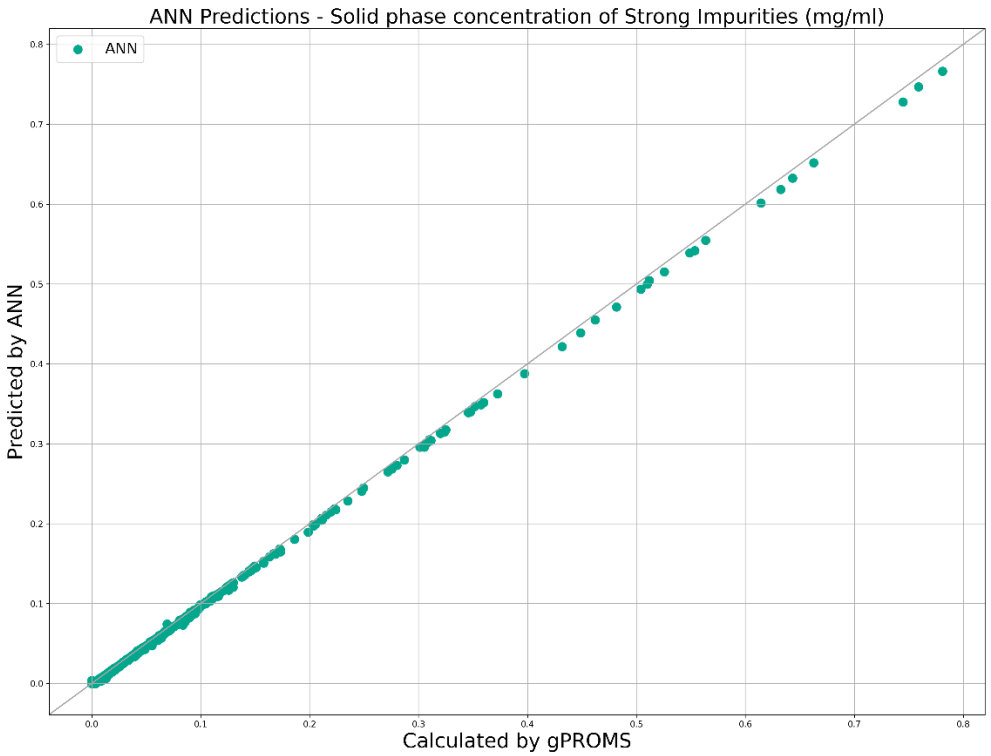

Figure 45: ANN predictions of the solid phase concentration of the strong impurities against the actual values calculated by gPROMS.

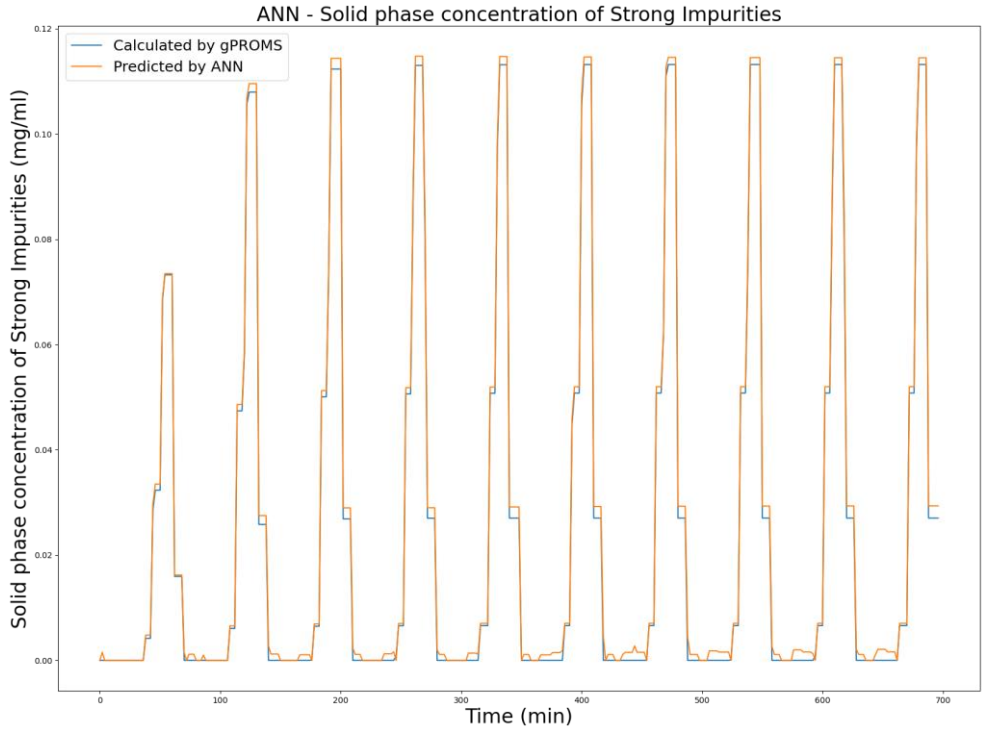

Figure 46: The solid phase concentration of the strong impurities throughout 10 process cycles calculated by the ANN and by gPROMS.

Appendix B: Second ANN Prediction Diagrams

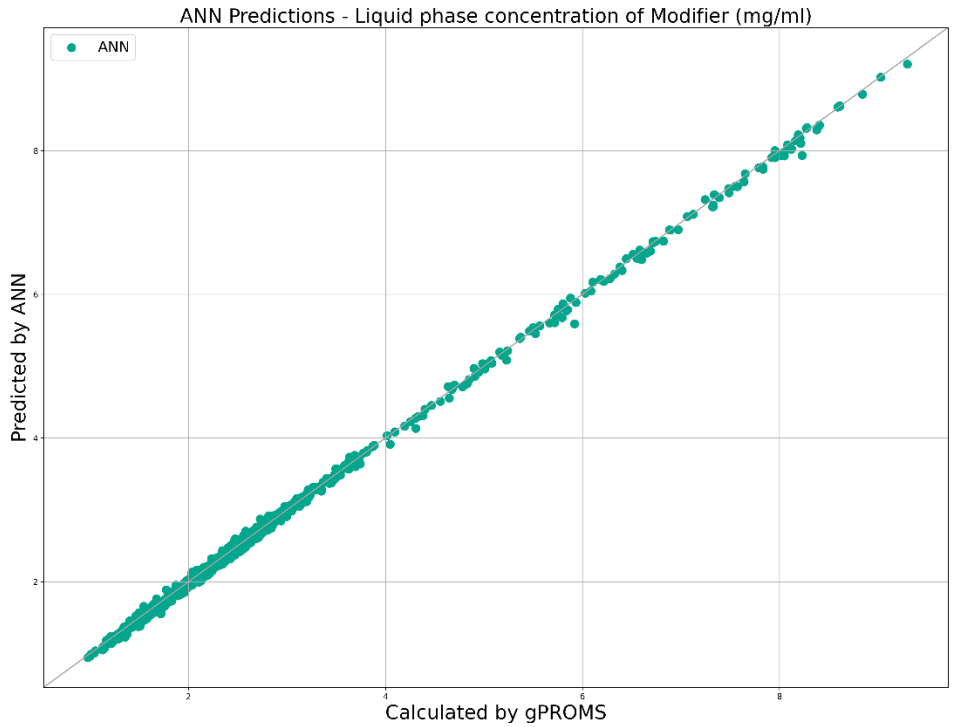

Figure 47: ANN predictions of the liquid phase concentration of the modifier against the actual values calculated by gPROMS.

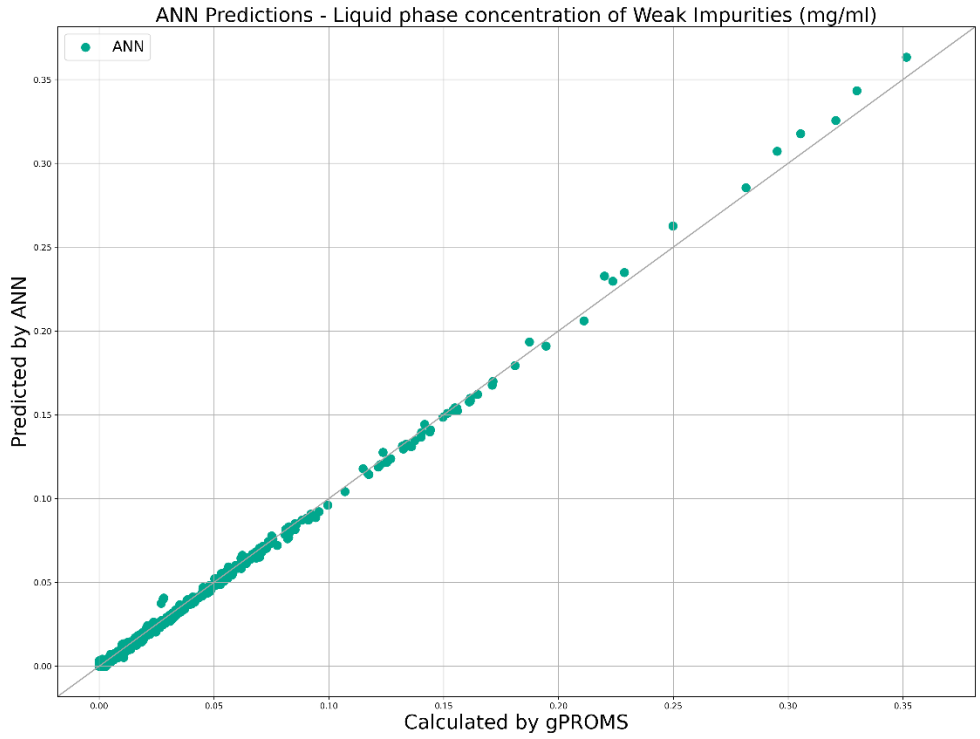

Figure 48: ANN predictions of the liquid phase concentration of the weak impurities against the actual values calculated by gPROMS.

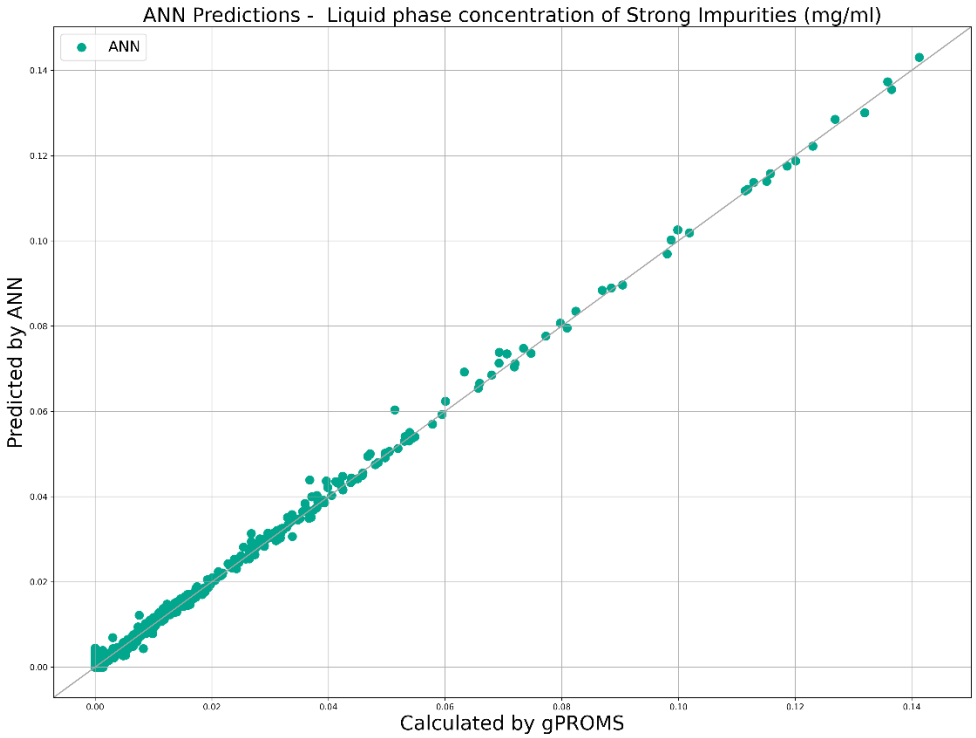

Figure 49: ANN predictions of the liquid phase concentration of the strong impurities against the actual values calculated by gPROMS.