# Digitalization in biopharmaceutical downstream processes

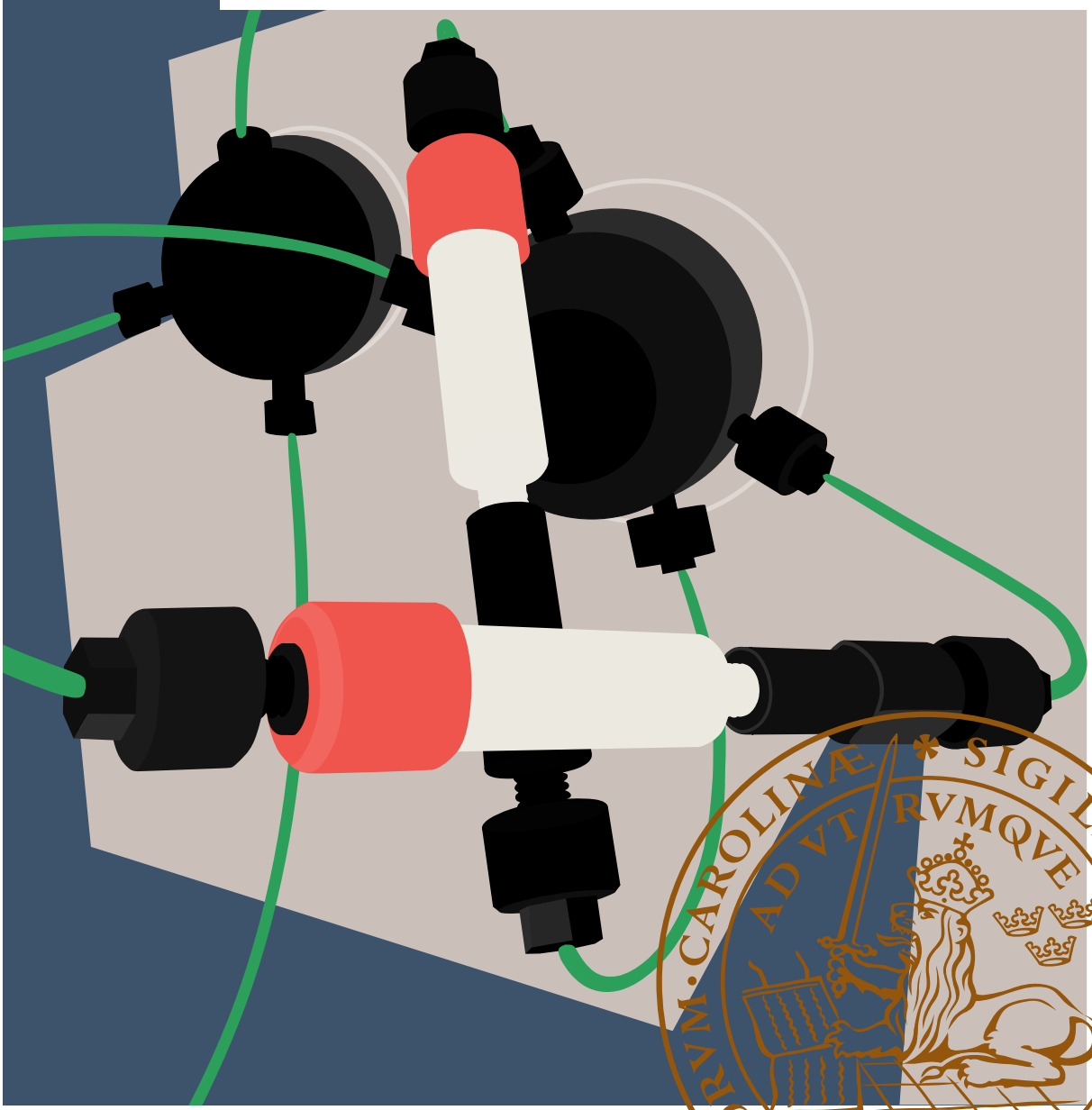
Espinoza, Daniel

2024

*Document Version:*
Publisher's PDF, also known as Version of record

# Digitalization in biopharmaceutical downstream processes

DANIEL ESPINOZA | CHEMICAL ENGINEERING | LUND UNIVERSITY



# Digitalization in biopharmaceutical downstream processes

Biopharmaceuticals are highly valued for their ability to specifically and efficiently treat diseases. However, they require rigorous purification that involves complex unit operations to reach the levels of purity necessitated for use in patients. Maintaining this purity while recovering as much of the valuable biopharmaceutical as possible is the core challenge in biopharmaceutical downstream processing.

Digitalization is a growing trend that has proven to increase efficiency in other manufacturing industries by implementing advanced automation based on high-quality data, automatic control tools and digital twin technology. This thesis examines the application of digitalization to the biopharmaceutical downstream process through the lens of six case studies. The use of digital twins, automated data acquisition and Internet-of-Things connectivity is showcased, shedding light on how digitalization can be used to achieve more efficient, robust production of biopharmaceuticals.

**DANIEL JONATHAN ESPINOZA** was born in San Salvador, El Salvador, and raised in Örebro, Sweden. In 2014, he moved to Lund to obtain his degree in chemical engineering. During this time, he found a growing interest in computer modelling of chemical engineering systems, and pursued further knowledge in the subject, leading to a PhD focused on applications of modelling and simulations in purification processes for biopharmaceuticals. He has applied his interests in modelling and programming to other subjects during his time as a PhD student, such as storm water management and purification of sustainable platform chemicals.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

ISBN: 978-91-8096-082-3

Chemical Engineering
Faculty of Engineering
Lund University

9 789180 960823

# Digitalization in biopharmaceutical downstream processes

Daniel Espinoza


LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Engineering at Lund University to be publicly defended on 13th of December at 10.00 in lecture hall K:C, Kemicentrum, Naturvetarvägen 14, Lund

*Faculty opponent*
Dr. Maria Papathanasiou

Imperial College London, UK

**Organization:** LUND UNIVERSITY

**Document name:** Doctoral Dissertation **Date of issue:** 2024-11-19

**Author(s):** Daniel Espinoza **Sponsoring organization:** Swedish Governmental Agency for Innovation Systems (VINNOVA)

**Title and subtitle:** Digitalization in biopharmaceutical downstream processes

**Abstract:**

Biopharmaceuticals are highly valued for their ability to specifically and efficiently treat diseases. However, they require rigorous purification that involves complex unit operations to reach the levels of purity necessitated for use in patients. Maintaining this purity while recovering as much of the valuable biopharmaceutical as possible is the core challenge in biopharmaceutical downstream processing. Digitalization is a growing trend that has proven to increase efficiency in other manufacturing industries by implementing advanced automation based on high-quality data, automatic control tools and digital twin technology. This thesis examines the application of digitalization to the biopharmaceutical downstream process through the lens of six case studies. The use of digital twins, automated data acquisition and Internet-of-Things connectivity is showcased, shedding light on how digitalization can be used to achieve more efficient, robust production of biopharmaceuticals.

**Key words:** biopharmaceuticals, monoclonal antibodies, downstream processing, digitalization, digital twin, data analysis, internet of things, mathematical modelling, simulation, optimization

**Language:** English **Number of pages**: 166

**ISSN:** 978-91-8096-083-0

**ISBN:** 978-91-8096-082-3

# Digitalization in biopharmaceutical downstream processes

Daniel Espinoza


LUND
UNIVERSITY

Faculty of Engineering
Department of Process and Life Science Engineering

ISBN 978-91-8096-082-3 (print)
ISSN 978-91-8096-083-0 (digital)

Printed in Sweden by Media-Tryck, Lund University
Lund 2024

*I’m very tired.*

# Acknowledgements

I have been privileged enough to have been surrounded by excellent people throughout my entire life. It is only thanks to them that I have been able to achieve the things that I have, despite my best efforts.

First off, my family. My mother, Roxana, for having given me the environment I needed to thrive and grow; my younger brother, Carlos, for always challenging me to be better; my youngest brother, Axel, for being the best kid in the world; Martin, for bringing my family the stability we have always needed; and my cousin, Claudia, for being a super cool person and friend; Ella, for being my best friend, partner, role model and supporter.

Secondly, my friends. Simon F and Fabi, for still being my friends after all these years. Lovisa, for broadening my horizons. Manfred and Erik, for giving me a place to belong when I knew no one. Frida, for being a good friend and a role model for me to strive towards. Gustav, for saving me from failing my classes multiple times. Elin, for being the coolest person I know. Malle, for helping me stay sane in Delft. Caroline, Emma M, Emma K, Linnéa, Max, the list could be made so much longer, but I think I'll stop here before I write another thesis' worth of text.

Thirdly, my colleagues. There are so many of you as well, so I'll try to keep it short. Simon T, for becoming one of my closest friends, for being my closest collaborator, and for making my time at the department so much more enjoyable. Madelène and Joaquín, for teaching me so much and for more awesome collaborations. Mikael Y-N for being a fantastic coach, a good friend and running partner. Anton for being a great office mate during my first time at the department. Maja and Julius, for being super cool and making me feel smart when they ask me questions. Melanie, for the nice collaboration and for the visit to Munich. Kensuke, for his patience in letting me practice my Japanese with him. Gusten, for hosting us in Portland and for being part of probably the most popular work I've co-authored so far. Salar for the nice collaborations, both past and future. Alex, for all the table tennis sessions. Mariona, Miguel and Ruben for giving me the opportunity to speak Spanish at work. Ashley, Medya, Edvin, Sofia, Sara, Bori, Simon G, Maria T, Maria M, Gity, Hilde, Mahala, Mikael S, Basel, Ahmet, and many more, for all the good talks we've had.

Finally, my teachers and supervisors. Bernt, for giving me this opportunity in the first place, for being patient with me, for his vast knowledge, and for being the best teacher I've ever had. Niklas, for his constant support during research and teaching and for solving crossword puzzles with me in the coffee room. Mats for teaching me a lot and for also solving crossword puzzles with me. Herje for his expertise in chromatography and for always being available for questions. Gunnar, Helena and Ola, for showing me how fun chemical engineering can be. Michael, for giving me feedback on my growth as a chemical engineer.

# Abstract

Biopharmaceuticals are highly valued for their ability to specifically and efficiently treat diseases. However, they require rigorous purification that involves complex unit operations to reach the levels of purity necessitated for use in patients. Maintaining this purity while recovering as much of the valuable biopharmaceutical as possible is the core challenge in biopharmaceutical downstream processing.

Digitalization is a growing trend that has proven to increase efficiency in other manufacturing industries by implementing advanced automation based on high-quality data, automatic control tools and digital twin technology. This thesis examines the application of digitalization to the biopharmaceutical downstream process through the lens of six case studies. The use of digital twins, automated data acquisition and Internet-of-Things connectivity is showcased, shedding light on how digitalization can be used to achieve more efficient, robust production of biopharmaceuticals.

# Populärvetenskaplig sammanfattning

## Bioläkemedel

Modern kemi ger mänskligheten möjligheten att tillverka specifika och komplexa läkemedelsmolekyler, men det finns gränser på hur stora och hur specifika dessa syntetiska molekyler kan bli. Därför har vi använt oss av naturens egna molekylfabriker, dvs levande celler, för att tillverka de molekyler som annars inte hade varit möjliga att syntetisera. Bioläkemedel syftar på just sådana medicinska molekyler som tillverkas i levande celler. Till exempel så har insulin tillverkats för att behandla diabetes sedan tidigt 1900-tal, då det utvanns ur kor och grisar. Framsteg inom biologi har gjort att vi inte behöver odla dem i levande djur, utan i isolerade jäst-, bakterie- eller animalieceller vars genetiska material modifierats till att tillverka molekylen i fråga.

Dessa läkemedelstillverkande celler odlas i så kallade bioreaktorer i ett tillväxtmedium, en vätska som innehåller cellerna, läkemedlet de tillverkar, samt eventuella näringsämnen som de behöver för tillverkningen. Både cellerna och näringsämnena behöver separeras från läkemedlet innan detta kan tas i bruk. Utöver läkemedlet så tillverkar cellen också andra ämnen som till exempel endotoxiner, vilka är skadliga för en patient och således också behöver separeras. Därför finns ett behov för extremt effektiva reningsprocesser vid bioläkemedelstillverkning. Dessa reningsprocesser kallas för *nedströmsprocessen*, medan själva odlingen av cellerna samt det som sker innan det kallas för *uppströmsprocessen*.

Ökande behov av nya bioläkemedel på grund av oväntade pandemier och upptäckten av nya kandidater till medicinska molekyler har orsakat ett behov av att accelerera utvecklingen av nya tillverkningsprocesser. Samtidigt så har långt ifrån alla människor på planeten tillgång till dessa läkemedel, på grund av både geografiska och ekonomiska skäl. Således finns det också en drivkraft till att göra den faktiska produktionen av bioläkemedlen mer effektiv.

## Digitalisering och Industri 4.0

Modern teknik har gett upphov till kraftfullare datorer och snabbare nätverkskommunikation. Som följd finns det en trend hos tillverkande företag att tillvarata denna teknik genom att använda sig av små sensorer med förmågan att strömma data till centrala databaser, avancerade simuleringar av tillverkningsprocesser och smart, automatiserad drift av processteknik. Den ökade datorkraften möjliggör även tillämpningar av avancerade kommunikations- och beräkningslösningar som kan ske i realtid. Detta tillämpande av nya, digitala verktyg för att effektivisera tillverkning kallas för *digitalisering,* eller *Industri 4.0.* Det senare begreppet syftar på en fjärde industriell revolution, då denna nya våg av digitalisering anses vara en utveckling på samma nivå som ångmotorn, elkraften och datoriseringen en gång varit. Inom kemi- och bioteknik så har tre huvudsakliga

koncept identifierats som väsentliga till digitalisering. De är som följer: *digitala tvillingar*, dvs digitala representationer av fysiska objekt som kan användas för att designa, tolka, förutsäga och styra det fysiska föremålet; *analys av stora datamängder* med hjälp av moderna metoder så som maskininlärning, vilket möjliggör konkretiserandet av sagda data och drar ner på tiden människor behöver lägga på att göra detsamma; och *Internet of Things*, vilket syftar på användandet av maskiner som kan kommunicera med varandra över nätverk och utbyta information som samlas från sensorer, vilket gör det möjligt att samla enorma mängder värdefulla data.

Digitalisering har potential att bidra till smartare automation i processindustrin, vilket reducerar produktionskostnader och effektiviserar tillverkningen. Genom att samla många data från processen via Internet of Things-integrerad utrustning och sedan analysera den med hjälp av metoder för hantering av stora datamängder, så kan underlag för att fatta beslut under processens gång skapas automatiskt. I kombination med digitala tvillingar och automatiska regleringstekniker så kan näst intill fullständig automation uppnås.

**Digitalisering av bioläkemedelsproduktion**

Många av de tekniker som utgör grundpelare i digitalisering har använts i många år i processindustrin. Mekanistisk modellering, reglerteknik och automatisk analys och insamling av data är exempel på sådana. Inom bioläkemedelstillverkning är nedströmsprocessen en särskild utmaning, eftersom en av de huvudsakliga enhetsoperationerna, kromatografi, är svår att samla data från och kräver särskilt tunga beräkningar för att simulera. Detta gör realtidsobservationer svåra, modellutveckling långsam och därav experiment dyra. Därför behövs metoder för att snabbare och automatiskt utveckla nya kromatografimodeller, automatiskt reglera kromatografiprocesser och bättre utnyttja de data som genereras under sådana processer. Historiskt har både datainsamling och modellering av kromatografi gjorts manuellt, och reglering av kromatografiprocesser är föga utforskat i litteraturen. Därför har jag, i denna avhandling, utforskat olika digitaliseringsförlopp för just kromatografiprocesser, med målet att både klargöra betydelsen av de olika digitaliseringskoncepten för produktionen av biologiska läkemedel, samt driva arbetet med att implementera sagda koncept framåt.

# List of papers

*Paper I*

Espinoza, D., Andersson, N. and Nilsson, B. In-silico formulation of iterative learning control for chromatographic purification of biopharmaceuticals, Computer Aided Chemical Engineering, 2022, 51, 1183-1188.

*Paper II*

Espinoza, D., Andersson, N. and Nilsson, B. Binary separation control in preparative gradient chromatography using iterative learning control, Journal of chromatography A, 2022, 1673, 463078.

*Paper III*

Espinoza, D*., Tallvod, S*., Andersson, N. and Nilsson, B. Automatic procedure for modelling, calibration, and optimization of a three-component chromatographic separation, Journal of chromatography A, 2024, 1720, 464805.

*Paper IV*

Zandler-Andersson, G*., Espinoza, D*., Tallvod, S*., Andersson, N. and Nilsson, B. Real-time monitoring of gradient chromatography using dual Kalman-filters, Journal of chromatography A, 2024, 1731, 465161.

*Paper V*

Espinoza, D*., Tallvod, S*., Andersson, N. and Nilsson, B. Automated quality analysis in continuous downstream processes for small-scale applications, Journal of chromatography A, 2023, 1702, 464085.

*Paper VI*

Isaksson, M., Espinoza, D., Lorek, J. K., Sondell, M., Andersson, N., Nilsson, B., Autonomous operation and quality monitoring of a continuous antibody downstream process. *Manuscript*.

* shared first authorship.

# Other related publications

Andersson, N., Gomis-Fons, J., Isaksson, M., Tallvod, S., Espinoza, D., Sjökvist, L., Zandler Andersson, G., and Nilsson, B., Methodology for fast development of digital solutions in integrated continuous downstream processing, Biotechnology and Bioengineering, 2023.

Sjölin, M., Sayed, M., Espinoza, D., Tallvod, S., and Al-Rudainy, B., Regeneration of dimethyl carbonate and purification of 5-hydroxymethylfurfural used in a biphasic dehydration process through activated carbon adsorption and evaporation, *Manuscript*.

Gerigk, M., Espinoza, D., Peng, D., Nilsson, B., and Minceva, M., Nonlinear liquid-liquid chromatography: a comprehensive modeling approach, *Manuscript.*

Haghighatafshar, S., Hallinger, E., Espinoza, D., and Al-Rudainy, B., An innovative method for estimating settling velocity of particles in stormwater using absorbance measurements and modelling, Water Practice and Technology, 2024, 19, 1810-1821.

# Author's contributions to the papers

*Paper I*

I implemented the simulations of ion-exchange chromatography, conceptualized the control algorithm, performed the case studies and wrote the paper with input from my co-authors.

*Paper II*

I conceptualized, planned, wrote the code for, and performed the experiments, compiled the data, and wrote the paper with input from my co-authors.

*Paper III*

I developed, planned, and performed the experiments, wrote code for calibration and optimization, conceptualized the study, and wrote the paper, all together with Simon Tallvod.

*Paper IV*

I planned, wrote code for, and performed the experiments as well as implemented the mechanistic model and developed the Kalman filter implementation together with Gusten Zandler-Andersson.

*Paper V*

I designed the sample collection and preparation system, planned, wrote code for, and performed the experiments together with Simon Tallvod. Joaquín Gomis-Fons provided expertise on PCC systems, and helped develop the architecture for systems communication.

*Paper VI*

I implemented the automated sampling and analysis system on the full downstream process, performed design and analysis of the virus inactivation column, performed data analysis on the generated data and implemented the database functionality.

# Table of Contents

# Introduction

## Biopharmaceuticals

Biopharmaceuticals are molecules with pharmaceutical properties that are produced in living organisms, such as bacterial, mammal or fungal cells. This is in contrast to other pharmaceuticals that may be manufactured using, for example, organic synthesis. Biological systems have the advantage over other pharmaceutical manufacturing processes in that they are able to produce larger, more complex molecules, namely proteins, which in turn can have very specific effects in a patient. The advent of genetic manipulation of cells has broken further ground in the viability of biologics since they make it possible to tailor the molecules to specific needs. Indeed, the market for biologics has grown massively since the 1980s, when the use of monoclonal antibodies for cancer treatment was advancing [1]. Other examples of such molecules include insulin and adeno-associated viruses.

One major challenge when working with biological systems is that the very same complexity that enables the production of such useful molecules, also results in many byproducts that either contribute nothing or may be harmful to the patient. Microorganisms produce other proteins referred to as host cell proteins (HCPs), that may cause harmful responses in patients. The desired pharmaceutical proteins may also form complexes with each other, such as polymers and oligomers, which can have similar effects. In addition, all bioreactors for biopharmaceutical production require the addition of nutrients and other substances that the cell requires to produce the desired molecule. Some biologics remain within the cells when they are produced, in which case the cells themselves need to be removed from the bioreactor and undergo lysis for the molecules to be accessed. This informs the need for thorough purification processes before biopharmaceuticals can be applied in treatment of diseases.

The purification processes of biopharmaceuticals are usually referred to as downstream processes, as opposed to what happens prior to the first purification step, which is called the upstream process and includes the bioreactor. The downstream process constitutes the largest cost involved in biopharmaceutical manufacturing (henceforth referred to as biomanufacturing), standing for up to 80% of the total costs [2]. These costs stem from the expensive purification techniques that are required to ensure a degree of purification that is acceptable for patient use. The main workhorse of the downstream process is chromatography, which allows

for very high selectivity but requires the use of highly specialized and expensive resins, and is difficult to operate in a fast and optimized manner [3]. Another challenge to overcome in the downstream process is to keep pace with the developments occurring in the upstream. In order to increase productivity and drive down costs, much work has gone into optimizing and intensifying the upstream process, be it by improving operational conditions, engineering more efficient cell cultures or changing the mode of operation from batch-wise to continuous production [4], [5]. In order to handle the output from these improved upstream processes, many efforts have been made to adapt the downstream process, such as modifying it to take a continuous input and integrating process steps to decrease hold-up times. Finally, as humanity is afflicted by new diseases and pandemics, the biopharmaceuticals that are produced need to change and thus there is a need for rapid development of novel biomanufacturing processes, both up- and downstream.

# Digitalization

In this section, the term digitalization will be studied in closer detail, with the end goal of specifying and limiting future discussions of the term. Digitalization can mean vastly different things in different fields, and variations of the phrase such as *digitization* and *digital transformation* are used interchangeably by some authors, while they are considered separate by others. One common distinction is the following: *digitization* refers to the act of converting analogue information into digital form, whereas *digitalization* refers to the adoption, application and utilization of digital technology [6], [7]. To give concrete examples, scanning physical books into a computer and adding them to a database would constitute digitization, while introducing a formal workflow involving said database to an organization would be digitalization. For a chemical engineering example, making use of digital tools such as logistics software or modelling and simulation would constitute examples of digitalization [8].

In chemical engineering, digitalization is often discussed in conjunction with the term *Industry 4.0*, or the fourth industrial revolution. This term refers to the application of digitalization tools to chemical manufacturing processes, which involves advanced automation by means of network communication between process units, generation and exploitation of process data, and autonomous decision-making based on these data to minimize human intervention. Differs from Industry 3.0, which refers to the unification of computers and equipment for automation, in that the automation is taken to a further extent by implementing advanced data analytics in real time by implementation of modern technologies and concepts [9], [10], [11]. In one study, Udugama et al. (2022) highlight three main technologies that differentiate Industry 4.0, which are all closely tied to the idea of digitalization. These are *digital representations* or *digital twins, data-based*

*methodologies* and *Internet of Things* [12]. It is through the lens of these concepts that I will continue to discuss digitalization in this thesis, and thus it is imperative that we understand what they are before we continue to talk about digitalization in the biopharmaceutical manufacturing process.

## Digital twins

Digital representations in chemical engineering are, according to Udugama et al. [12], closely related to the term *digital twins*. Digital twins in a manufacturing context can be simply defined as digital representations of physical objects, but a very thorough discussion on the definition of a digital twin can be found in Kritzinger et al. (2018) [13]. In said study, the use of the term digital twins is examined in literature and an attempt is made to create a more unified definition. The result is a classification system for digital twins based on the level of integration between the digital and the physical worlds in terms how automated the flow of information is between the two. The following three classes were proposed:

- A *digital model* is a digital representation of a physical object with no automated flow of information between the two. A manual flow of information may have occurred from the physical to the digital in order to obtain model behavior that replicates the physical object (i.e., model calibration). Conversely, the digital model may inform decisions and actions in the physical object based on simulations, and thus a manual transfer of information may occur form the digital to the physical world. However, changes in either the digital or physical object have no direct effect on the other.
- A *digital shadow* takes the integration a step further by implementation of a one-way, automated flow of information between the physical and the digital. For instance, if a change occurs in the physical object, the digital object is updated to reflect that.
- A *digital twin* is, naturally, when data flows automatically from the physical object to the digital and vice versa, meaning that a change in either will affect the state of the other. As an example, data may flow from the physical to the digital object, and the resulting change of states in the digital object automatically informs a decision to be made about the physical object, which is executed automatically. In this example, the digital object acts as a controller for the physical.

A reader with experience from chemical engineering or other kinds of manufacturing may realize that the definitions of these classes are reminiscent of existing concepts in the field. In particular, the digital twin sounds very similar to any type of automatic controller, such as PID controllers. Indeed, while digital twins

may be novel in many other disciplines, the term is applicable to technologies that have been present in chemical engineering for decades. Still, it is worthwhile to examine these technologies from this broader perspective. Examples of different applications of digital twins in all different classes will constitute the larger part of this thesis.

## Data-based methodologies

The collection and utilization of data in manufacturing has been common practice for a long time, for example to perform linear, nonlinear, and/or multivariate regression to find correlations between operating conditions and process outputs. However, when the term "data-based methodologies" is used in the context of digitalization and Industry 4.0, it refers to management of large datasets that necessitate the use of more modern, automated data analysis methods [12]. Examples of such methods include clustering, classification, and modern regression techniques based on principal component analysis (PCA) or artificial neural networks (ANNs). Common for these methods is that they are, in the best case, capable of analyzing extremely large datasets that would take a substantial time to go through using classic data analysis techniques, potentially even finding correlations that are difficult to discern for the human eye. Another differentiating property of modern data analysis techniques is that they often function in an unsupervised manner, where the point of human interaction lies in the tuning of hyperparameters of the analytical method used, rather than wrangling with the data itself.

## Internet of Things

Internet of Things (IoT) is another term that is very broad in its application. In an everyday life example, a so-called "smart home", where appliances in a home are connected to a network and can be accessed and controlled via a singular device, such as a smartphone, is an example of IoT. From an information technology standpoint, much focus lies in the specific technology used to achieve network communication between such appliances. One more practical definition of IoT is the deployment of relatively small devices equipped with sensors that are capable of streaming data obtained from said sensors, as well as being controlled via network communication. From an industrial standpoint and based on the above definition, IoT can refer to the deployment of small sensors with network connectivity to a manufacturing plant that has no supervisory control and data acquisition (SCADA) system in place, making it a low-cost alternative that is made more viable by the availability of cheap, accessible devices such as Arduino micro controllers or Raspberry Pi computers. [14]

A natural question to ask is: what differentiates IoT from existing SCADA systems in manufacturing? One study by Wan et el. (2016) [15] discusses the application of IoT from an Industry 4.0 perspective and highlight one roadblock: the heterogeneity of hard- and software prevents efficient communication between different pieces of equipment over a network. From this identified issue, we can determine that an industrial IoT application refers to the unification of different pieces of equipment (sensors, industrial robots, etc.) and the data they provide into a common architecture that can be used to improve process operation, by means of network communication. Simply put, the interconnectivity of different process units is a key factor in industrial IoT.

# Thesis Aim and Outline

The remainder of this thesis will be dedicated to discussing how these digitalization concepts can be applied to the downstream processing of biologics. The papers included correspond to different studies performed both on individual unit operations in the pharmaceutical downstream processes, as well as on a full downstream processing train. During these studies, many techniques have been applied that can be connected to the above-mentioned three main pillars of digitalization, and the aim of this thesis is to bring the terminology of digitalization as a broad concept into the field of biopharmaceutical downstream process research. Many of the techniques that I have used have a long history in the chemical engineering field as a whole, and I aim to highlight the connection between existing techniques and new terminology to paint a clearer picture of what digitalization and Industry 4.0 mean for the biopharmaceutical industry. Automation is a major theme in most of my papers, and discussion on the role of automation in the digitalization workflow will also be included.

The outline for the remainder of this thesis is as follows: in Chapter 2, I will give an example of a downstream process that aligns with the trends of continuous manufacture and digitalization, as well as showcase the methods that can be used to operate such a system in an autonomous fashion. In Chapters 3, I will talk about how digital twins can be applied in many different steps of downstream process development, from design of both unit operations and controllers, to real-time applications. In Chapter 4, I will go over data-based methodologies that were applied in this work. Mainly, these will cover data acquisition, labelling and storage, which are essential for the advanced data analytics covered by this term to be used. In Chapter 5, the relationship between the methods I have used, and the Internet of Things will be explored. Finally, Chapter 6 concludes the thesis and sets up a future perspective on what work remains to be done for these concepts to go further and satisfy an Industry 4.0 viewpoint.

# The Biopharmaceutical Downstream Process: Design and Operation

## Continuous process for production of monoclonal antibodies

In this chapter, I will present a downstream process for purification of a monoclonal antibody (mAb) that will be used as the context for the digitalization tools discussed in the rest of this thesis. The process in question, discussed in more detail in Paper VI, was designed to handle a continuous inflow of bioreactor harvest fluid (previously produced and frozen for storage). The mAb in question was trastuzumab, otherwise known as Herceptin, cultivated in Chinese hamster ovary cells [16], [17].

The process consisted of three primary unit operations: virus inactivation, product capture, and product polishing [18]. The virus inactivation took place in a packed bed column that was packed to such a height that it maintained a residence time of at least 30 minutes to ensure sufficient contact with the solvent and detergent system that was used to deactivate the viruses. The product capture was performed on three protein A chromatography columns operated in parallel using a configuration called periodic counter-current chromatography (PCC). Protein A chromatography is used due to its selectivity to the mAb: the mAb binds to the column while anything else passes through. The bound mAb is then released, or eluted, from the column by pumping an elution buffer through the column. In a PCC configuration, two protein A columns are connected in sequence and loaded with the bioreactor harvest (after virus inactivation) while the third column is being eluted. By altering which columns are being loaded and eluted, the PCC system can receive a continuous inflow of bioreactor harvest. The output from the PCC system is periodic, meaning that pulses of the product are received periodically from the eluted column.

The product obtained from the PCC is very close to pure due to the high selectivity of protein A chromatography. What remains to be separated from the product pool are variants of the mAb that are similar in structure, but different in effect. Size variants such as high molecular weight (HMW) species, or aggregated mAb molecules that have formed polymers, can have adverse effects in patients if they remain in the product and must be removed. The content of charge variants, i.e.,

mAbs that are structurally identical but possess different surface charges, also needs to be regulated. It is to this end that the polishing step is applied. In this process, the polishing step was performed using a multi-modal chromatography column operated in flowthrough mode, meaning that the product flows through the column without fully binding to it. Different size and charge variants interact differently with the column packing, leading to different residence times, which enables separation.

The virus inactivation and capture steps were implemented on an ÄKTA PCC chromatography system (Cytiva, Uppsala, Sweden). The polishing step was implemented on an ÄKTA Pure system (Cytiva, Uppsala, Sweden). The full downstream process is illustrated in Figure 1.

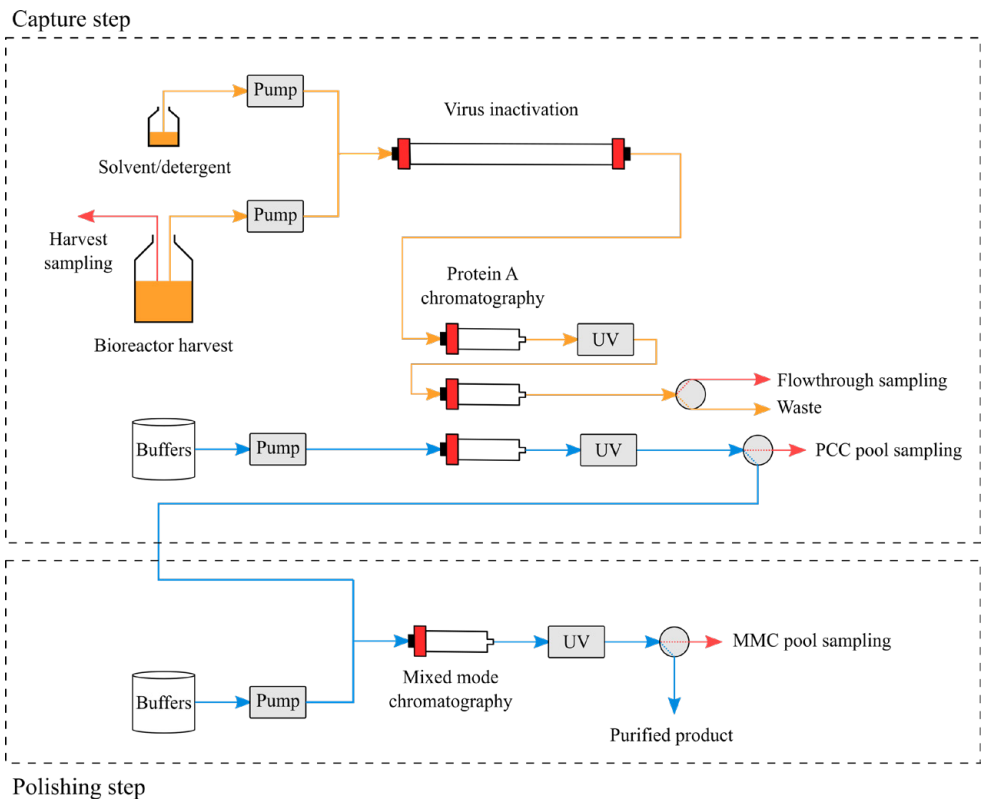



**Figure 1**
An example of a continuous downstream process for purification of monoclonal antibodies. A virus inactivation column receives a continuous flow of bioreactor harvest fluid. The flow is sent to a three-column PCC setup, in which two columns are loaded in sequence while one column is eluted. This means that the output of the PCC is periodic, while the input is continuous. The periodic product output is polished further in a multi-modal chromatography step in flowthrough mode. Three sampling points are indicated, where important process information may be obtained.

This process is an alternative to the continuous mAb platform applied by Scheffel et al. (2022) [16] and Schwarz et al. (2022) [17], which used cat- and anion exchange chromatography columns in sequence as the polishing step, the former in bind-and-elute mode and the latter in flowthrough mode. In other words, many different kinds of chromatography can be used in a biopharmaceutical downstream process, from Protein A, to ion exchange, to multimodal.

# Operation of downstream processes using Orbit

The ÄKTA chromatography systems used in the downstream process are typically controlled using a software interface called Unicorn (Cytiva, Uppsala, Sweden), which can be used to send instructions to the pumps, valves and sensors of a chromatography system. However, in the totality of my work, I have used a Python-based software interface called Orbit, which was developed at the Division of Chemical Engineering at Lund University, to operate the ÄKTA systems instead [19], [20]. Orbit acts as a bridge between Python and Unicorn, and enables users to define instructions as Python scripts. It is not limited to ÄKTA systems, but can be used with any piece of laboratory equipment with an open interface for programming via serial or Ethernet connections. For instance, Orbit has previously been used to control an analytical high-performance liquid chromatography (HPLC) system from Agilent (California, USA) via their own application programming interface (API) [21].

A schematic overview of Orbit is shown in Figure 2. The Orbit kernel consists primarily of a real time control engine, which executes commands in real time until certain conditions are met, and an object library, which contains digital representations of physical pieces of equipment. These digital objects contain methods that are specific to the physical object, such as setting valve positions or pump flowrates. They can also contain attributes that correspond to the physical characteristics of the object, such as tube volumes, column porosities, etc. While a process is being operated, information from the physical objects such as sensor signals and flowrates, is streamed to the real-time controller, which in turn sends instructions to the digital objects so that instructions may be executed in the physical configuration.

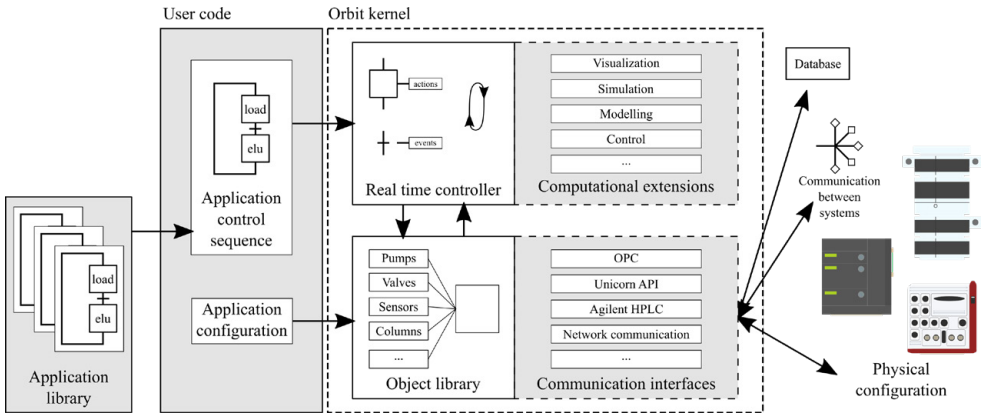



**Figure 2**
Schematic overview of Orbit and the way that users interact with it.

The Orbit kernel is highly flexible in that it can be extended with computational methods that function in tandem with the real time engine. For example, a real-time visualization program can be used to showcase important data as the process is being executed, or automatic control tools can be implemented thanks to the flexibility of the Orbit code. In addition, new communications interfaces can be implemented to make Orbit compatible with other types of equipment, as previously mentioned. Orbit can even be used to communicate with other instances of Orbit operating other pieces of equipment over a local area network.

The user interacts with Orbit via Python scripts in two ways. First, via a script that defines the configuration of the specific application to be used, i.e., the list of pumps, valves, sensors, columns, tubes and other units that make up the physical equipment configuration. Second, via a script that defines the specific sequence of instructions to be executed, as well as the conditions that need to be fulfilled for the individual items of the sequence to continue to the next step. In addition to these two points of interaction, the user may also make use of pre-existing applications when defining their application sequence. This is a core strength of the Orbit code: existing code may be reused and adapted to address new challenges.

# Digital Twins in Biopharmaceutical Downstream Processes

Digital twins have been applied in design of biopharmaceutical downstream processes for many years, although they may not have been referred to as such. The concept of the digital model has been central to the Quality-by-Design paradigm that was introduced by the U.S. Food and Drug Administration in the 2000s, which aimed to reduce the number of supplemental applications for approval of minor changes to a biomanufacturing process that occurred as responses to process variations, as well as to encourage the implementation of novel technologies and improvements to the process. The idea was to combine the quality assurance work into the design process for new biopharmaceuticals by improving understanding of the product and the process at an early stage. [22]

One key role of digital models in Quality-by-Design, particularly as it pertains to the downstream process, has been to aid in the determination of design spaces. When an optimal operating point has been selected for, e.g., a chromatographic separation step, its robustness can be tested by perturbing the operating parameters slightly and measuring their impact on critical quality attributes (CQA) of the process, such as product purity. Examples of operating parameters that may be relevant to product quality are the salinity and pH of buffers used in the chromatography process, or the concentration of the product in the feed to the process. By applying this methodology, a span of values of the operating parameters can be determined, within which the CQAs maintain acceptable values. While these design space characterizations can be performed by experimentation, it has been showcased that digital, mechanistic models posses the ability to predict process behavior in a wide region around an operating point, and enable the user to avoid extensive, exploratory laboratory work by substituting it with computer simulations [23], [24], [25]. The bulk of the experimental work is instead shifted to finding the parameters of the mechanistic model.

As was alluded to in the introduction of this thesis, the framework defined by Kritzinger et al. (2018) [13] allows for model-based controllers to be thought of as digital twins. Numerous examples of automatic control applied to a variety of biopharmaceutical DSPs, from control of specific unit operations to plant-wide control, as well as ranging from mechanistic to data-driven approaches [26], [27],

[28], [29], [30], [31]. In this chapter, different methodologies for creating digital twins will be covered, as well as some examples from the list of papers.

# Methodology: Mechanistic or Data-driven?

Modelling is central to the concept of digital twins, and the discussion on what constitutes a model is similar to that surrounding the definition of a digital twin. Is a 3D render of a chromatography column a model? Or is an equation that describes the chemical interactions in the column a model? Arguably, both are correct. Similarly, is the 3D render a digital twin of the column, or is the equation the digital twin? Again, arguably, both are true depending on the context: the purpose of the model or twin determines its applicability. In biopharmaceutical manufacturing and development, a narrow definition would be "a digital representation of the process with all the relevant specifications, user requirements and information sources, that supports associated stakeholders in their decision taking and enables direct control of the physical system" [32]. In the source of this citation, the authors stress the significance of the prediction and extrapolation capabilities of a digital twin. When regarding digital twins of individual unit operations, these are properties attainable through mathematical models, which will be the focus in the remainder of this section. Mathematical models can belong to different classes, two of which will be covered here: mechanistic and data driven. A combination of the two constitutes a third class, hybrid models, which are a very interesting avenue of research that is outside of the scope of this thesis.

## Mechanistic Models

Mechanistic models are based on the fundamental laws of physics and chemistry, representing a process by describing the underlying physical phenomena that govern system behavior. In the context of biopharmaceutical downstream processing, this involves modelling fluid flow, mass transfer, and adsorption kinetics in processes like chromatography.

In Paper III, we modelled the complete flow path of a chromatography system using a mechanistic approach. Three processes were of particular importance: continuous, ideally stirred tanks; fluid flow in cylindrical channels, i.e., tubing; and flow through porous, packed beds with descriptions of adsorption equilibria, i.e., chromatography columns. Stirred tanks can be modelled as follows:

$$\frac{dc}{dt} = \frac{F}{V}(c_{in} - c) \tag{3.1}$$

The concentration of a component, $c$, varies over time $t$ as function of itself, the flowrate $F$, the tank's volume $V$ and the concentration at the inlet, $c_{in}$.

Flow through tubing can be described by a one-dimensional convection-dispersion equation, as follows:

$$\frac{\partial c}{\partial t} = D_{ax}\frac{\partial^2 c}{\partial z^2} - u\frac{\partial c}{\partial z} \tag{3.2}$$

Here, the concentration $c$ of a component propagates along the spatial dimension $z$ by two mechanisms: axial dispersion, quantified by the dispersion coefficient $D_{ax}$, and convection via the superficial velocity, $u$.

Many approaches exist to modelling chromatography processes, and depending on the specific adsorption mechanism and application, different models may be viable. In Papers I, III and IV, ion-exchange chromatography was studied, and thus an appropriate model was selected. In general, chromatography can be described by an extension of the convection-dispersion equation that takes the porous nature of the bed into account, as well as adds an adsorption term:

$$\frac{\partial c}{\partial t} = D_{ax}\frac{\partial^2 c}{\partial z^2} - \frac{u}{\varepsilon}\frac{\partial c}{\partial z} - \frac{(1-\varepsilon_c)}{\varepsilon}\frac{\partial q}{\partial t} \tag{3.3}$$

$$\varepsilon = \varepsilon_p + (1-\varepsilon_c)\varepsilon_p \tag{3.4}$$

The velocity is adjusted with the total packed bed porosity, $\varepsilon$, which in turn is a function of the column void fraction, $\varepsilon_c$, and the particle porosity, $\varepsilon_c$. The added term describes adsorption taking place in the pore volume of the bed, which involves compensating with the column porosity, $\varepsilon_c$. The concentration of the component adsorbed to the column, $q$, is in turn described by an adsorption model specific to the application. In ion-exchange chromatography, the steric mass action model is often used. This model has the benefit of including a description of how a molecule that is adsorbed to the stationary phase of the column acts as a steric hindrance to other molecules binding to adjacent sites [33], [34]. While the base form of the model involves descriptions of ionic capacity, steric hindrance and characteristic charge of proteins, a simplified version of the model can be formulated which lumps several of the parameters together:

$$\frac{\partial q_i}{\partial t} = k_{kin,i}\left[H_{0,i}c_i\left(1-\sum_{j=1}^{n}\frac{q_j}{q_{max,j}}\right)^{v_i} - {c_s}^{v_i}q_i\right] \tag{3.5}$$

In this kinetic expression of the steric mass action model, the adsorption of species $i$ is dependent on the concentration of salt in the mobile phase, $c_s$, as well as on lumped parameter $H_{0,i}$, which combines the equilibrium partitioning coefficient between the mobile and stationary phases, the protein's characteristic charge, $v_i$, and the ligand density. The maximum adsorption capacity of one component, $q_{max}$, is a combination of the ligand density of the column, and the characteristic charge and steric shielding factor of the protein [35]. The benefit of this lumped parameter version of the model is the reduced dimensionality of the calibration problem: fewer parameters lead to better convergence.

While the models used in the papers included in this thesis are sufficiently covered in this section, other relevant mechanistic models for alternative unit operations exist. Membrane filtration, virus inactivation, ultrafiltration/diafiltration, and membrane chromatography constitute a few examples of important unit operations for which mechanistic models exist [36].

## Data-driven Models

While mechanistic models offer deeper insights into the process, data-driven models—which rely on empirical correlations between inputs and outputs—are often more practical for certain use cases. These models do not attempt to describe the underlying physics but instead use historical or real-time data to predict system behavior. Data-driven models excel at tasks such as pattern recognition and real-time optimization, especially when large datasets are available. They typically require less computational power than mechanistic models and can be rapidly deployed. However, their ability to extrapolate beyond the conditions in which they were trained is limited, meaning that they perform best within the dataset's bounds.

An example of a simple, data-driven model is shown in Papers I and II. A simple, linear, input-output model of linear gradient elution in chromatography was used as the basis for the iterative learning controller developed in that work:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.5}$$

The two inputs, $\mathbf{u} = [u_1 \quad u_2]$, are mapped to the outputs, $\mathbf{y} = [y_1 \quad y_2]$, by a linear function written on matrix form, $\mathbf{G}$. The resulting two-input, two-output model is fast to execute as it is a simple matrix multiplication, which makes it ideal for real-time applications. In this particular work, the inputs were the initial and final salt fractions of a linear salt gradient during the elution of two proteins, and the outputs were the residence times of the two proteins, expressed in volume. This is illustrated in Figure 3. Of course, the true elution behavior is non-linear, as can be seen in the steric mass action model. However, in a limited region, as in the acceptable design space for the process, the system behavior may be approximated by linearization.

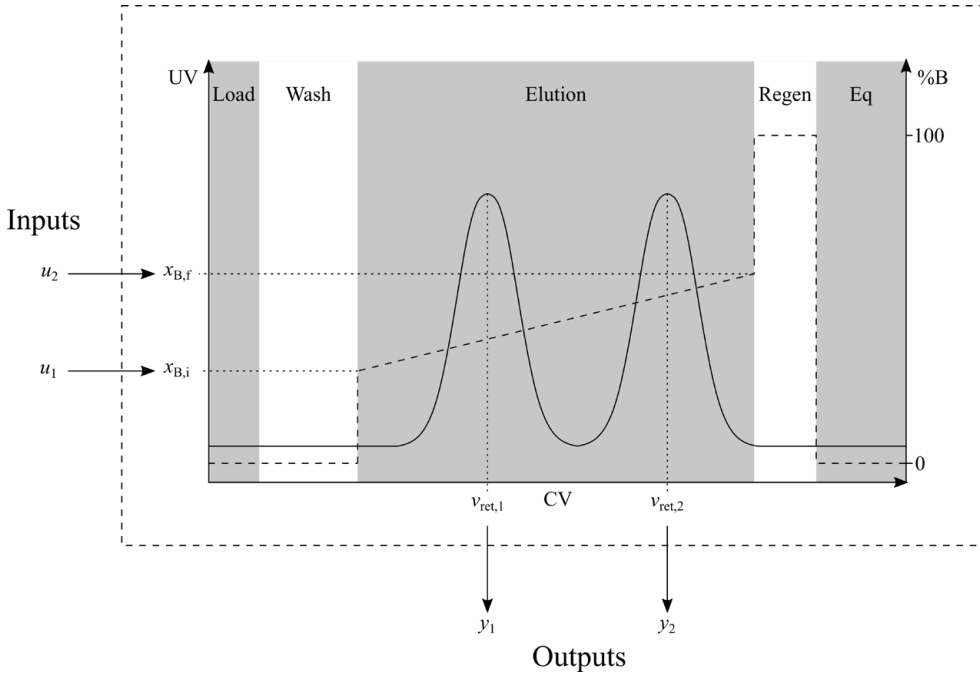



**Figure 3**
A description of how a data-driven model can be constructed for the elution of two components in a bind-and-elute chromatography process.

In the context of Industry 4.0, data-driven models typically refer to a more advanced approach to process modelling. One example of this is the artificial neural network (ANN). While not explored in the papers that constitute this thesis, ANNs possess several key benefits over simpler, linear models such as that detailed above, the main one being their ability to capture complex and non-linear correlations. This is also a benefit they possess over mechanistic models: when the process dynamics are too complex and not fully known, as tends to be the case in bioreactors [37] and membrane filtration [38], [39], [40] due to the complex composition of the fluid streams involved, an ANN may be capable of providing better predictions than a mechanistic model. Another advantage of ANNs over certain mechanistic models is their execution time. In chromatography, the mechanistic models are very complex sets of non-linear, partial differential equations. Their use in computation-heavy methods such as optimization, where the model needs to be evaluated several times to solve the problem, leads to long computation times. Using an ANN model as a surrogate for the mechanistic model can significantly hasten optimization problems. The main weaknesses of ANNs is the requirement of large volumes of data, which may be incredibly expensive to generate in biopharmaceutical contexts, and the difficulty of training such models, since they are prone to overfitting and thus may be less suitable for extrapolation beyond the range of the training data.

### Use Cases and Trade-offs

The choice between mechanistic and data-driven models depends on the specific needs of the digital twin. Mechanistic models are invaluable for situations where it is important to understand the system's internal behavior and to predict outcomes under a wide range of conditions. They are most useful when developing new processes or scaling up production, where the ability to extrapolate is critical. On the other hand, data-driven models are ideal for real-time process control and optimization tasks where the system is already well understood, and rapid response is more important than deep physical insight. Since operation of a process is typically performed within a range of operation conditions, the ability to extrapolate may not be as critical. For instance, a mechanistic model can be used initially to develop a digital twin and generate reliable predictions, while a data-driven model can take over in day-to-day operations to handle real-time adjustments based on incoming process data.

## Applying and Enabling Digital Twins

With the concept of digital twins in mind, we can now examine the case studies in this thesis and learn more about how different classes of digital twins, based on different types of models, can be applied to solve engineering problems in the biopharmaceutical downstream process. One important concept that will be highlighted is how enabling digital twins in one application can expedite novel digital twins in others.

### Digital Models

In Papers I and II, an automatic batch-to-batch controller for an ion exchange chromatography step was developed and showcased. The initial development of the controller, shown in Paper I, was based on a mechanistic model of the ion exchange chromatography process, i.e., a physically accurate and predictive description of the system. This allowed for a lot of freedom in experimentation on how the controller should be designed, as well as for some validation of the controller through simulated test runs with artificial disturbances, all while obeying the behavior of the physical system. Several experimental designs were tested until a functioning proof-of-concept controller was finalized, with no chemicals, proteins or other lab resources expended. This final controller design was then tested against real disturbances in a physical setup, as shown in Paper II. This shows how digital models can be used to expedite the development of novel technologies while minimizing trial-and-error experiments and operational costs of a research lab.

A major strength of mechanistic models is their ability to extrapolate beyond the conditions used to calibrate the model, making them highly valuable for designing new processes or optimizing existing ones under untested conditions. This is what made the in-silico iterative learning controller development possible. However, these models are often more complex and require extensive experimentation to obtain parameter values. When developing similar technology for different molecules, such as in novel drug development, significant work is needed to develop such a model. To mitigate the time and effort required for mechanistic modelling, a framework for modelling and calibration of chromatography was developed and presented in Paper III.

This automated framework was made possible by the flexibility of the Orbit software. Orbit has an embedded simulator, which makes use of the user-defined system configuration (i.e., the list of physical units in the system and their interconnectivity) to generate a mechanistic, digital model structure of the system. A detailed overview of the Orbit simulator is given by Tallvod et al. (2022) [35], who applied it to model the elution behavior of a single protein in ion exchange chromatography. In Paper III, we extended the framework to calibrate model parameters for multiple components and applied it to a ternary ion exchange separation. A short overview will be given here.

The Orbit simulator works by reading the system definition and sequential instructions given by the user when operating a physical piece of laboratory equipment, such as an ÄKTA chromatography system. The system configuration is read, and an adequate model equation is chosen for each physical unit, as shown in Figure 4. Proper boundary conditions for each unit are selected depending on how they are interconnected. For instance, in Figure 4, consider the chromatography column connected to the column valve (ColV): the outlet of the column is mapped to the inlet of the following piece of tubing as a boundary condition.

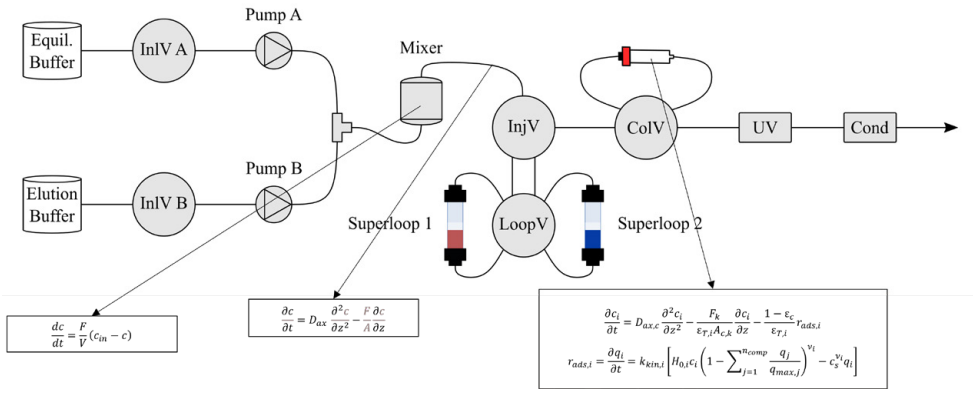



**Figure 4**
An example of a flowpath used in automatic model calibration. The Orbit simulator selects an appropriate model equation depending on each physical piece of equipment included in the system definition so that tubes, mixer chambers, valves and columns are properly represented.

Once the model system of equations has been set up, the calibration procedure is set into motion. In summary, several parameters of the model such as dead volumes, column porosities, void fractions, and adsorption parameters are each calibrated by corresponding sets of experiments. The experiments and their corresponding control sequences are pre-defined in Orbit code by the user, and so the calibration procedure pulls from this library of existing code to perform experiments, from which data are saved.

Once all experiments have been performed and the data have been obtained, the same code used to run the experiments is used to perform simulations of the chromatography system using initial guesses of the model parameter values. The simulations result in simulated data that correspond to the real data obtained from the experiments, be it UV absorbance or conductivity measurements. The difference between the simulated and measured data forms an *objective function*, and selecting parameter values that minimize this objective function is the goal of the calibration procedure. To this end, a mathematical optimization algorithm is used. Several of these optimization problems need to be solved in order to find all parameter values, and each optimization problem corresponds to one control sequence. An overview of the calibration procedure is given in Figure 5.

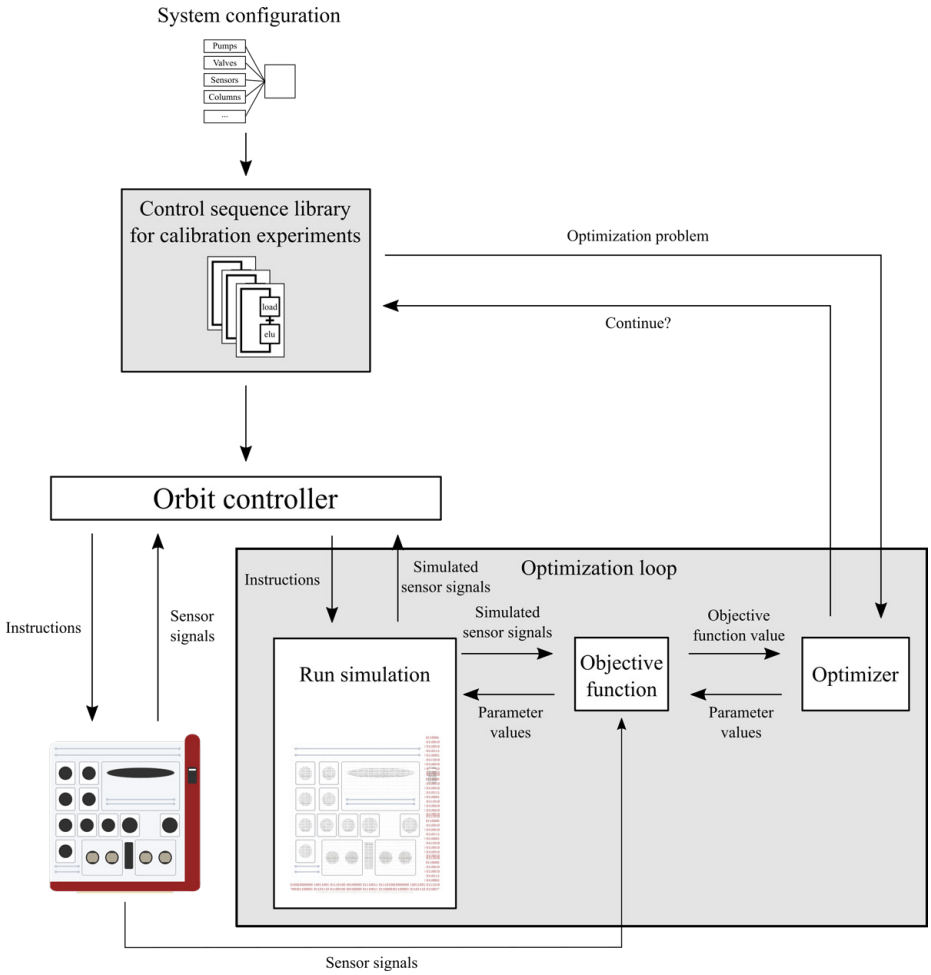



**Figure 5**
Schematic of the automatic calibration procedure workflow. The user interacts with both the physical system and simulated systems using the same code interface. The physical system works independently of the simulated one, and the recorded sensor signals are stored for use in the calibration procedure. An optimization problem is solved for each control sequence, each corresponding to a set of model parameters. Once an optimal set of parameter values is found, the next control sequence is used in the simulated system to solve the values of the next set of parameters, until all parameters have been determined.

In Paper III, optimization of the yield and productivity of the ternary separation based on the calibrated model was performed as part of the automated procedure. This filled two purposes: to demonstrate the use case of a digital model for process design, and to generate a set of validation data to determine the model fit. Three different optima were obtained, obeying different constraints on product purity and placing a different weight on yield or productivity. One of the found optima was

selected, and a corresponding physical experiment was performed to validate the model at this optimum point. The results showed a surprisingly good fit to the data, although some discrepancies were found due to the optimum operating at conditions outside of the range of the model calibration data. These discrepancies could be assuaged through selection of a different model structure (i.e., a different adsorption model to that shown in Equation 3.5). Regardless, this further showcased how an automated framework can facilitate digital model development and application.

## Digital Shadows

In Paper IV, a real-time state estimation tool was implemented for model-based monitoring of ion exchange chromatography. The tool in question was based on Kalman filters, linear-quadratic state estimators that are used to combine real-time measurements with model predictions to obtain improved process understanding over what is possible with the two individually. Kalman filters have been applied in many industrial contexts for over 60 years, and is thus not a novel technology. However, their application to real-time control systems in industry has been limited due to the high level of complexity involved in performing the state estimation [41], [42]. From the perspective of the digital twin classification system, a Kalman filter can be regarded as a digital shadow due to the continuous nature of the flow of information from the physical system (via sensor measurements) to the digital.

A brief walkthrough of the Kalman filter algorithm is as follows: at a given point in time, a model of the system being monitored is evaluated to predict the system states at a later point in time, based on a system input. The state variance is also computed. This is called the *prediction step*. Next, as a measurement of the system output is obtained from the physical system, the difference between the model prediction and the real measurement is corrected based on the magnitude of the error and the state variance. The state variance is also updated. This is referred to as the *update step.* Every time a new measurement is obtained, one cycle of the algorithm is performed. A more in-depth explanation of the Kalman filter algorithm is provided in Paper IV.

In chromatography, the main benefits of Kalman filters are two-fold: from a model prediction perspective, one can obtain a better estimate of the true process states by integrating real-time data into the prediction. From a monitoring point of view, the detail provided by a model can give information that is unavailable when only using online measurements such as UV absorption or conductivity. To elaborate, in chromatography, typically the concentration profiles of eluting components inside the column are unavailable for measurement. Instead, measurement instruments are placed at the column outlet. UV detectors act as the primary measurement of elution profiles, and have the downside that components are difficult to distinguish from each other in the chromatogram if their chromatogram peaks overlap. A model of multiple components can distinguish between their respective elution peaks, but models are not perfect and will contain prediction errors in many cases. A Kalman

filter combines the strengths of real-time data and detailed modelling to create a best-of-both-worlds scenario. To illustrate this improvement, the open-loop, unfiltered process model prediction is compared to the Kalman filtered prediction in Figure 6.

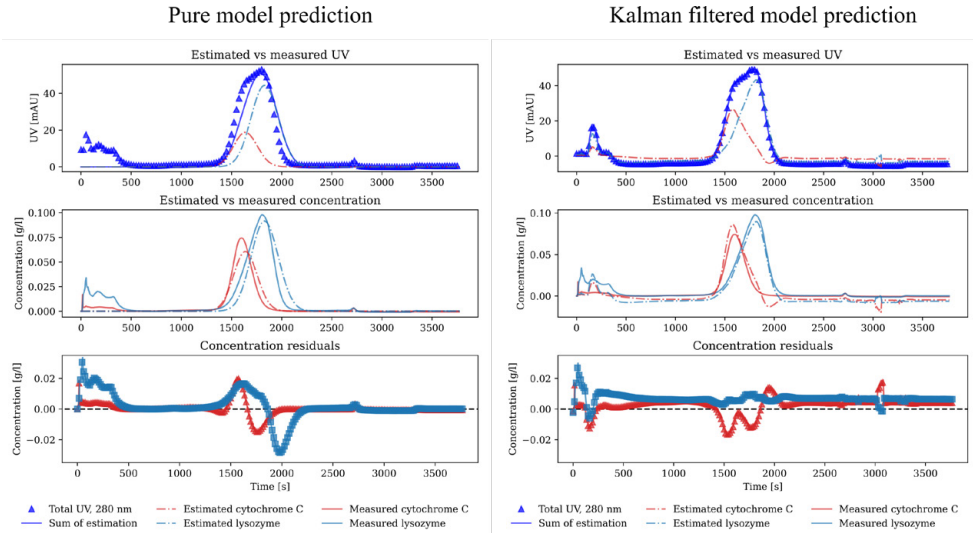



**Figure 6**
Comparison between unfiltered (left) and filtered (right) model predictions of two co-eluting components in ion exchange chromatography. The UV measurement (blue triangles) does not distinguish between the two components. The un-filtered model captures the elution profiles of the two separate components, but does not fit perfectly to the data. The Kalman filtered model response is able to correct for the model errors and matches the individual elution profiles better than the pure process model. A small, unphysical, negative prediction is seen in the filtered model prediction of one component (red).

The Kalman filter implementation in Paper IV was based on two filters working in tandem: one linear Kalman filter for the salt elution profile, which is crucial when describing bind-and-elute behavior; and one extended Kalman filter to describe the non-linear elution of binding components. The conductivity at the column outlet was used as the measurement for the former, and the UV absorbance for the latter. This split between two filters was essential due to the highly non-linear dynamics of the separation. By lifting the salt propagation in the column out of the binding model, the non-linearity was reduced, and the computational burden was lowered.

The execution time of the Kalman filter algorithm is a crucial factor to its implementation. The sampling frequency, i.e., the rate at which new measurements are obtained, needs to happen with enough time between each sample, that the Kalman filter algorithm is able to complete its execution before the next sample is

obtained. In Paper IV, we were able to achieve a robust Kalman filter with a sampling frequency of once every five seconds. This was accomplished through *parallel computing*, i.e., by distributing calculations of the algorithm across multiple computer processors to allow them to be performed in parallel, rather than in series. This was sufficient for the low loads on the chromatography column in the case study, but may not prove to be sufficient in a real case, where the load is much higher and may lead to higher degrees of non-linearity and thus longer computation times.

The Kalman filters were implemented as a computational extension to Orbit, along with a real-time visualization extension to monitor the Kalman filter performance in terms of model fit and execution time, as shown in Figure 7.

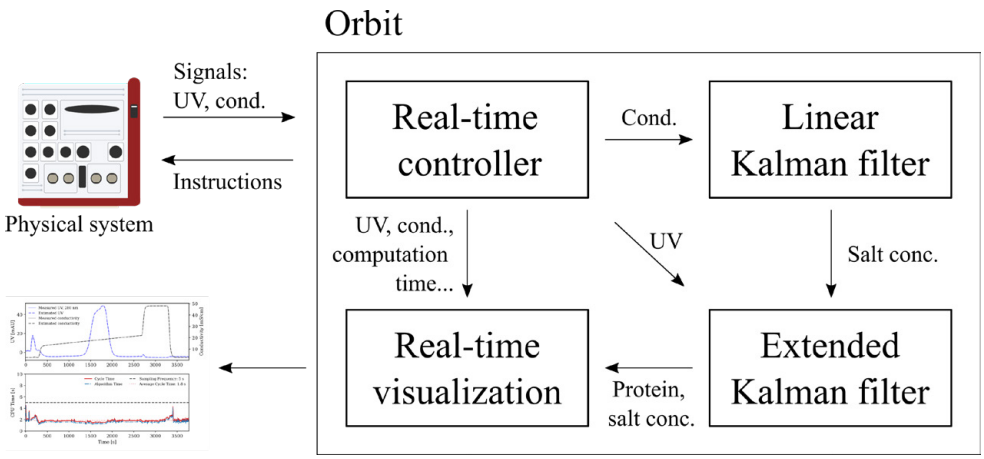



**Figure 7**
Graphic description of the Kalman filter implementation in Orbit. Measurements are obtained in real time from the physical system via Orbit, which are then used in the two Kalman filters to obtain an improved model prediction. The results are used in a real-time visualization extension, which can be used to monitor the Kalman filter performance.

The Kalman filter implementation in Paper IV falls under the category of digital shadow due to the one-way, continuous flow of information from the physical system to the digital. The improved model predictions are the main benefits of this application, but the possibilities do not end there. Thanks to the model predictions being accessible to Orbit, the step towards making use of them for model-based control of, e.g., product pooling, is a short one indeed. This use of a digital shadow for automatic control of the physical system introduces information flow from the digital to the physical, achieving two-way flow and thus reaching the digital twin stage.

## Digital Twins

The most straight-forward example of a two-way flow of information between a digital model and a physical system in a manufacturing process is an automatic controller. Such a controller is tasked with maintaining a process outcome at a desired set point value when exposed to process disturbances, for instance, maintaining a desired product yield or process productivity. To this end, the controller is able to manipulate control variables that affect the outcome, such as the slope of a linear elution gradient in chromatography or the elution flowrate. Paper II showcased such a controller application, where the residence times, or volumes, of two proteins in a bind-and-elute ion-exchange separation were controlled by the linear gradient settings during elution (see Figure 3). The real-time exchange of information between the physical and the digital was enabled by the Orbit implementation, which was able to both access the sensor data to determine the residence times of each component based on the chromatogram, and send instructions to the physical instrument as the control variables were determined.

The controller was based on iterative learning control, an automatic control concept originating from the field of robotics. The core idea behind iterative learning control is to improve the performance of processes that repeat over time by harnessing the data generated during each repetition. This makes it suitable for ion exchange separations in bind-and-elute mode, since the cycle of loading, washing, eluting, and regenerating the column is repeated during a continuous process run.

The controller algorithm is executed once for every process cycle, denoted $k$. The process inputs (linear gradient settings, denoted $\mathbf{u}$) in cycle $k-1$ are used in the chromatography process, $\mathbf{G}$, and results in two residence times, $\mathbf{y}$. Depending on a set of disturbances $\mathbf{d}$ which may affect the process, $\mathbf{y}$ may differ from the desired value, $\mathbf{y}_d$. The task of the controller, $\mathbf{KD}$, is to compensate for this error in the next cycle, by adjusting the process inputs. Three kinds of disturbances are considered: disturbances to the input (via, e.g., errors in the buffers used in the elution), disturbances to the process dynamics (due to column aging and capacity loss), and disturbances to the output (via errors in determining the residence times due to noisy data). The controller's effect on the former two was seen in Paper II, while the latter was not seen directly. In addition to these disturbances, a functionality was added to allow the user to change the desired value $\mathbf{y}_d$ and have the controller pre-emptively adjust the inputs to conform to the change. This was achieved with a feed-forward element to the controller, $\mathbf{FF}$. An overview of the controller structure is provided in Figure 8.

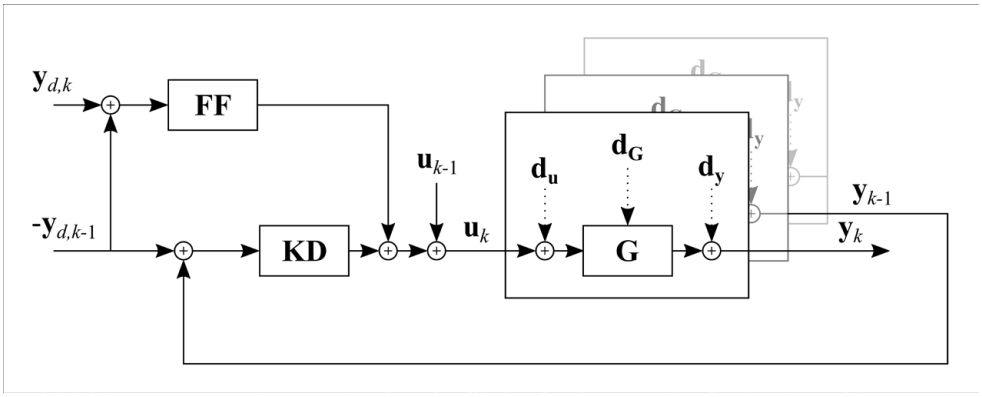



**Figure 8**
Block diagram of the controller structure from Paper II. At cycle number $k-1$, the chromatography process (denoted $\mathbf{G}$) takes an input $\mathbf{u}$ and results in an output $\mathbf{y}$. The process is subject to disturbances $\mathbf{d}$ to the input, output and the process dynamics, which leads to a deviation from the desired process outcome $\mathbf{y}_d$. On the next cycle, $k$, the error is corrected by the controller (denoted $\mathbf{KD}$). A feed-forward controller ($\mathbf{FF}$) was also implemented to preemptively adjust for changes to the set point by the user.

# Reflection and Conclusion

Different classes of digital twins have been applied in Papers I-IV. In this section, I will go over the ways in which the choices of mechanistic or data-driven models affected the different case studies in the papers and how the approach to each case may be improved upon.

In Papers I and III, digital models were applied, i.e., there was at most a one-time, manual transfer of information from the physical to the digital and vice-versa. This occurred in terms of model parameter calibration and process design. The mechanistic model's ability to, in detail, capture the behavior of the physical system proved essential in Paper I, where a new technology for automatic control (an iterative learning controller) was developed entirely in-silico, entirely foregoing lab trials during the development stage. This shows how digital models enable creative problem solving by giving the researcher freedom in pursuing many different avenues before going to the physical system for validation of the developed technology. In Paper III, this same, detail-oriented modelling approach proved very useful in the optimization performed at the end of the model calibration sequence, although its limits were tested by going outside of the range of the calibration data and some discrepancies between the model and the data were seen. In addition, the computationally expensive partial differential equations that constituted the model resulted in long optimization times. Previous studies have shown how mechanistic models can be used to generate simulated data to train artificial neural networks, which post-training are very fast to evaluate and can be used as a surrogate for the

mechanistic model in numerical optimization [43], [44], [45]. Still, it is important to consider the risk of overfitting when using ANNs: either the optimization must be constrained within the operating conditions used in the training data, or other constraints need to be placed on the ANN. Physics-informed neural networks (PINN) are an interesting avenue, since they embed the mechanistic model equations in the neural network without requiring expensive numeric integration [46], [47].

In Paper IV, the implementation of a digital shadow, with automated, real-time data flow from the physical to the digital system, was shown using existing engineering concepts (Kalman filters). The combination of online sensor data and mechanistic modelling provided an improved model prediction over what either could in isolation. However, the long simulation times of mechanistic chromatography models proved an even bigger hurdle in this study due to the sampling frequency of the online sensors: ideally, the higher the sampling frequency, the better the Kalman filter. Thus, data-driven surrogate models are also an interesting alternative in this case. Here, the overfitting is less of an issue since the model is not used to explore different configurations of operating parameters: it is enough if the fit is good to and around the operating point. This digital shadow application is very close to being a digital twin. By utilizing the improved predictions to, e.g., automatically control the pooling of the product, an automated flow of information in both directions is achieved.

Finally, in Paper II, a digital twin was implemented with a data-driven model as its foundation using an iterative learning controller approach. The consistent process performance even under disturbances is the main benefit of automatic controllers, in addition to how they reduce human intervention. Data-driven models are executed quickly and are thus suitable for real-time applications such as this. An overview of the types of models used in the papers, as well as their place in the digital twin classification system, is given in Figure 9.

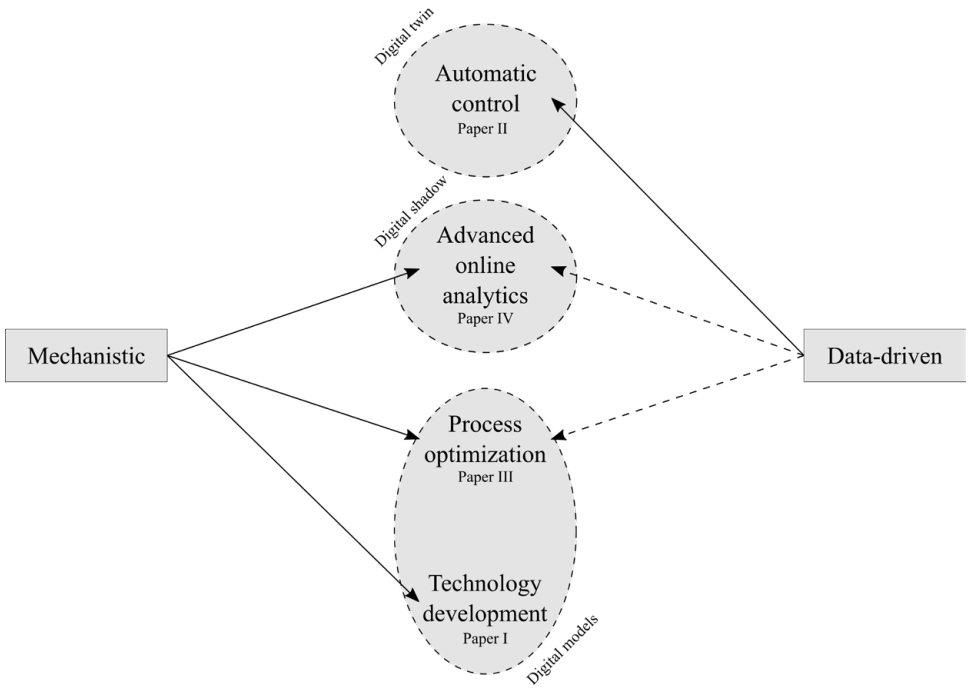



**Figure 9**
Graphical overview of the digital twin applications covered in this thesis. All three levels of the digital twin hierarchy have been applied, using both mechanistic and data-driven approaches. Two potential improvements have been identified, shown with dashed lines: the computation time of both the real-time monitoring tool (Paper IV) and the process optimization (Paper III) could be accelerated by applying appropriate, data-driven surrogate models, e.g., based on artificial neural networks.

In conclusion, both mechanistic and data-driven models have their place in the development and implementation of digital twins in biopharmaceutical downstream processes. Mechanistic models provide the depth and predictive power needed for extrapolation and the design of new processes and technologies, while data-driven models offer simplicity, ease of use, and faster computation for real-time applications. By leveraging automated modelling and calibration frameworks and implementing parallelized computation, it is possible to overcome the challenges associated with mechanistic models, making them more accessible for real-time use. Ultimately, the choice of model should align with the specific goals of the digital twin, whether it be long-term process optimization or immediate control and monitoring.

The development of digital twins for downstream processing in biopharmaceuticals is a significant milestone in the ongoing evolution toward Industry 4.0. By combining real-time simulations, advanced process modelling, and automation, digital twins offer powerful tools for optimizing biomanufacturing processes, improving both efficiency and product quality. Although challenges remain in terms

of optimization of simulation times and model accuracy, the groundwork laid by digital shadows, such as the Kalman filter applications, provides a stepping stone toward fully integrated digital twins that will redefine process control in biopharmaceutical manufacturing. Novel technologies such as ANNs give a promising outlook on the future of digital twins in biopharmaceutical downstream processes.

# Data-based Methodologies

## Introduction

As biopharmaceutical processes move toward higher degrees of automation, the ability to generate, process, and interpret large datasets has become crucial for ensuring process reliability, product quality, and operational efficiency. The complexity of the bioreactor harvest composition can make real-time mechanistic modelling applications difficult, since clear information about the actual composition of the fluid flow cannot be obtained from conventional sensors. Instead, more involved analytical methods need to be applied to obtain high-quality data. Such data can then be used to obtain data-driven models for digital twin applications, or to mine important information about process performance and product quality via the use of analytical tools for large data sets. However, both of these applications require vast amounts of data, and the quality of said data directly affects the performance of the application. A robust framework must be established to automate sample collection, preparation, and analysis, as well as to correctly label and store the data in a centralized manner for accessibility purposes.

This chapter focuses on the development of such frameworks, emphasizing the need for continuous and automated data acquisition systems that streamline the collection of high-quality data. In particular, we explore the automated quality analysis system (QAS) developed for continuous downstream processes. This framework forms the basis for the generation of process data, which can be further leveraged for applications such as digital twins in downstream bioprocess control. Additionally, this data allows for monitoring of key performance indicators (KPIs) that can be integrated into a plant-wide control strategy.

## Framework for Automated Data Acquisition: The Quality Analysis System

The core of this chapter revolves around the implementation of the Quality Analysis System, developed to enable at-line sampling, sample preparation, and analysis in continuous downstream processes. The QAS provides a solution to one of the major bottlenecks in data acquisition: manual sample handling, which is both time-

consuming and prone to human error. By automating the sampling process, the QAS ensures consistent and reliable data collection, reducing variability and allowing for high-frequency sampling over extended process runs.

Paper V details the first implementation of the QAS as a support system to periodic counter-current chromatography (PCC) purification setup for mAbs. It was designed to automatically collect samples from two critical points: the bioreactor supernatant and the product pool from the chromatography columns. These samples were then conditioned and analyzed using high-performance liquid chromatography (HPLC), providing data on key product quality attributes such as aggregate content and charge variant composition. The hardware setup is showcased in Figure 10.

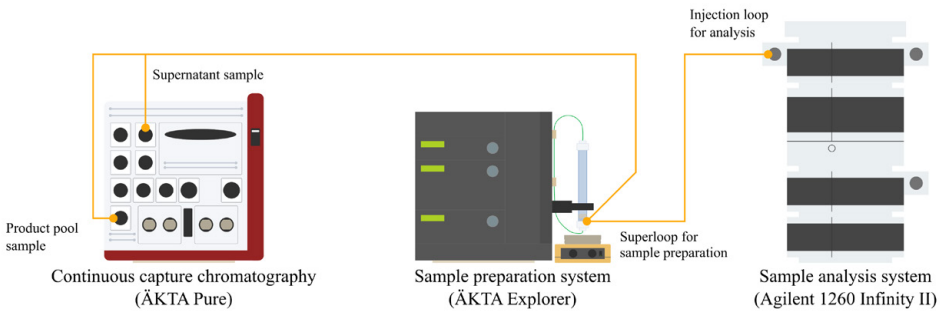



**Figure 10**
Hardware configuration of the automated quality analysis system. Continuous capture chromatography was performed in a periodic counter-current setup on an ÄKTA Pure system. A sample preparation system was configured on an ÄKTA Explorer system, which included a superloop for sample collection and preparation. The collected sample was sent to an Agilent analytical chromatography system for analysis.

The sampling functioned by diverting the fluid flow of the purification process to the sample preparation system by switching its valves. Regard, for example, sampling of the product pool from the PCC system. If a sample is ordered, the product collection valve on the ÄKTA Pure system would switch positions and divert the pump flow away from the product collection vessel, and instead to the sample preparation system. There, it would be stored in a superloop, which is a cylindrical vessel that is divided into two parts by a small, watertight piston. The sample would fill one side of the piston, where a small magnetic stirrer was placed to homogenize the concentration of the sample. Once collected, the sample could be prepared in many different ways, be it by adjusting pH or adding reagents required for whatever analytical method needs to be performed. The prepared sample would then be sent from the sample preparation system to an analytical system. In Paper V, this was an Agilent 1260 Infinity II HPLC system, where two different analysis

protocols were implemented: size exclusion chromatography for determining the contents of size variants of the mAb, and weak cation exchange chromatography for investigating the charge variant profile. For the latter, the pH needed to be higher than the elution pH of the Protein A chromatography on the PCC, a conditioning buffer was added to the sample on the preparation system prior to being sent for analysis.

## Data Storage and Labelling

The concept of the QAS was expanded in Paper VI in two ways. Firstly, it was applied to an end-to-end, integrated downstream process with additional sampling points (see Figure 1). Secondly, database connectivity for data collection, storage and labelling was implemented. The database was based on the non-SQL MongoDB framework and was hosted on a server in our laboratory. Orbit interfaces with the database via network communication and streams information such as sensor data, pump flowrates, accumulated volumes and valve positions to the database in real time, giving access to chromatograms, conductivity profiles etc. The data stream begins at the start of an Orbit run, and is tagged with a randomized ID number.

Collections of data stored in the database can be broadly divided into two different types: discrete and continuous data collections. Discrete data collections contain information such as the user-defined name of a process run, time stamps for the start and end of the run, the type of run, the physical system configuration and other information that pertains to the run as a whole. Essentially, any metadata about the run is stored in a discrete data collection. In contrast, continuous data collections gather time-varying information from the process during a run, most notably recorded sensor signals. This can be expanded by the user to include other information about the process that changes over time, like states in the process such as valve positions and pump flowrates. The continuous data is essentially the *result* of a run, and is linked to the run's metadata, i.e., to a discrete data collection. One way of viewing it is that discrete data collections are filled with information at the start and end of a run, i.e., at discrete points in time, while continuous data collections are constantly updated with new information during a run, i.e., they are continuously updated with new information. The continuous data is what the user would make use of to draw chromatograms, for example. The relationship between discrete and concrete data collections is shown in Figure 11.

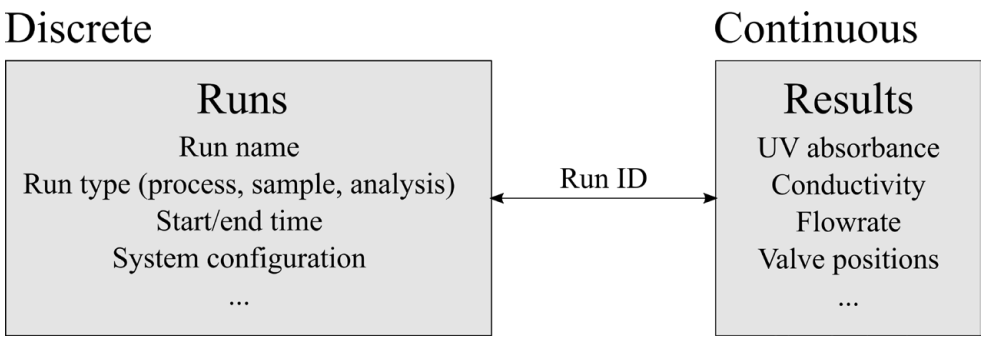



**Figure 11**
Relationship between discrete and continuous data collections in the database used in Paper VI. Discrete data collections correspond to a full run and contain metadata about the run in question. Continuous data collections correspond to a run, are linked to the run's discrete data collection by a run ID number, and contain continuously changing information from the run, such as sensor measurement.

When a process was initialized in Paper VI, individual discrete runs were created on the database for each of the involved systems. When a sample was ordered, a database run of type "sample" was created to correspond to the sample, which included a connection to its corresponding process run's ID number. As analyses were performed on each sample, a run of type "analysis" was created and linked to the corresponding sample run, also by ID number. The continuous data obtained from each analysis would then be saved as a result of that analysis run. The data obtained from the main process would, correspondingly, be saved as a result of the process run. Thus, all data is accessible via the main process run. This is illustrated in Figure 12.

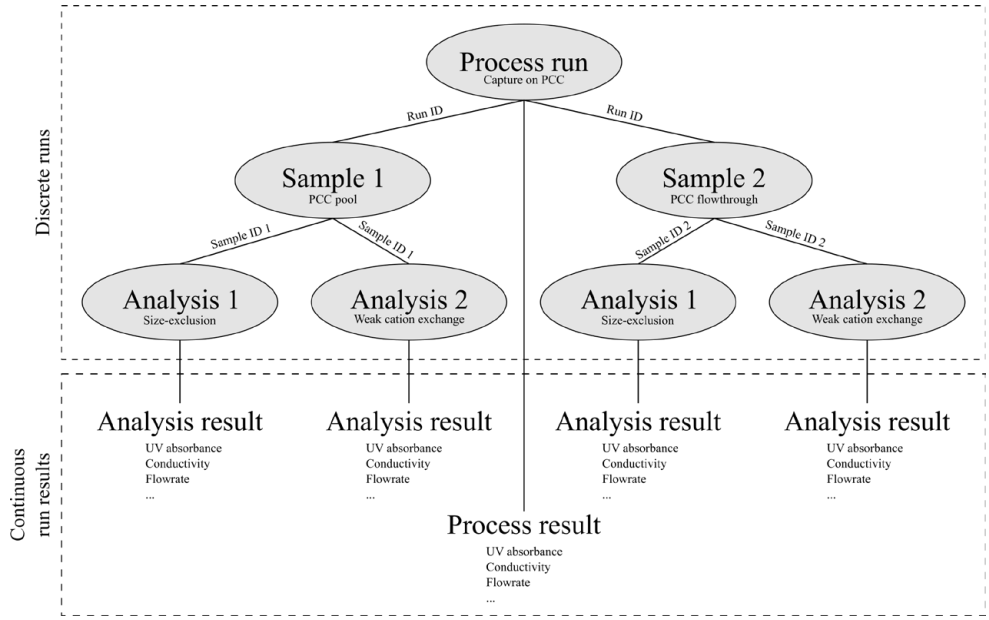



**Figure 12**
Example of the relationships between process, sample, and analysis runs, as well as their corresponding results, in the database. All information is accessible from the main process run via unique ID numbers.

The structure of the database makes it possible to keep specific data that may be relevant to only a certain type of run in the discrete data collection, while keeping the continuous results congruent for ease of access. For example, a collection of the type "sample" may contain the volume of the sample as it was received, the volume post-dilution and pH adjustment, and a list of "analysis" runs that correspond to the analyses that were performed. The main process run would then contain a list of all runs of type "sample" that correspond to the samples taken. Thanks to the different types of data being connected in a logical fashion, the step towards e.g. automated data analysis in real time during a process run, via the database, becomes exceptionally shortened.

# Potential Applications of the QAS

## Leveraging Data for Digital Twins

The high-quality data provided by the QAS can be used to enhance the accuracy of digital twins by feeding real-time information into the model. For example, data on

the concentration of aggregates or charge variants from chromatography can be integrated into the digital twin of the downstream process, allowing the virtual model to make more accurate predictions about process performance.

Furthermore, this data can be used to continually update and calibrate the digital twin, ensuring that it remains reflective of the physical system. The ability to synchronize real-time data with the virtual model enhances process control by allowing for real-time adjustments based on the digital twin's predictions. This is particularly useful for processes with slow dynamics, where real-time feedback is essential for maintaining product quality and operational efficiency.

The Kalman filter we developed is very dependent on known information about the input to the chromatography step, i.e., the concentration of product in being loaded. In a continuous processing train that includes the bioreactor, changes to the product concentration due to the reactor need to be taken into consideration for the Kalman filter to continue operating robustly. Thus, it is important to sample and analyze the composition of column-binding components in the bioreactor harvest at regular intervals so that the Kalman filter has accurate, up-to-date information to work from. While the product concentration is expected to vary over time during a continuous manufacturing run, in a case such as the manufacturing platform discussed in Chapter 2, the dynamics are expected to be very slow. This means that the frequency of the sampling can be low, e.g., once every 24 hours. The capacity of the QAS is well-suited for this type of application, and is a potential solution to this issue.

## Real-time Decision-making

One key, potential application of the data obtained from the QAS is for real-time decision making in the process. For example, if a sudden increase in aggregate content or shift in the charge variant profile is detected during a process run, the user has a chance to intervene thanks to the anomaly being made visible by the QAS. One potential roadblock for this application is the capacity of the QAS. The analysis protocols were not optimized in any of the case studies, and were thus quite lengthy. In particular, the cation exchange protocol was around 40 minutes long. However, the PCC load cycle times in the case study in Paper VI were up to approximately 220 minutes long, and thus the analysis times fell within the time limit for intervention in the process on a cycle-to-cycle basis.

In order for this type of anomaly detection to be possible, the process of analyzing the data from the QAS needs to be automated. As previously mentioned, the data in the database is structured in such a way that data automation is a simple to implement since all relevant information is accessible, both on a specific sample level and an overall process level. Thanks to the database being accessible from Orbit, one could conceptualize an observer Orbit that interacts with the database in real time during operation of the integrated downstream process from Paper VI. This

instance of Orbit would perform data analysis as new data is available on the database, and make decisions about actions to be taken based on the outcome. Combined with Orbit's ability to communicate with other instances of Orbit over a local area network, such decisions could be executed automatically, since each individual process is controlled by an instance of Orbit. Figure 13 shows how the different instances Orbit involved in such an application would interact.

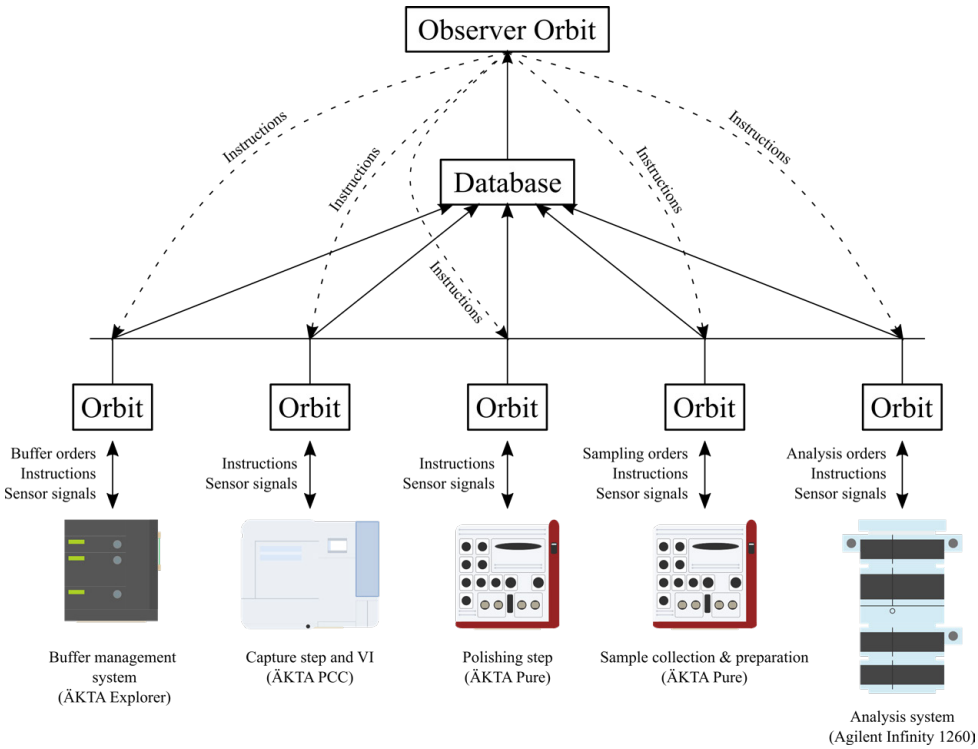



**Figure 13**
Example of how an observer Orbit instance could be implemented in the integrated mAb purification platform described in Paper VI. The observer would access the database and perform data analysis on new information as it is added to the database in real time. Decisions could then be made based on predefined protocols and be executed by sending the relevant information to the corresponding instance of Orbit via the local area network connection.

# Conclusion

The developed automated sampling and analysis platform offers several benefits. Firstly, it reduces manual labor. With samples being collected and analyzed automatically, the need for manual intervention is minimized, freeing up time for more critical process development tasks. Secondly, it achieves consistent data acquisition. The platform eliminates variability associated with manual sampling, ensuring that data collected throughout the process is both accurate and reliable. Finally, it enables the implementation of further automation in data analysis thanks to its database connectivity. Making use of such automated frameworks for data collection and storage is a quintessential aspect of digitalization. In this thesis, I have focused mainly on the process control and automation aspect of data collection, but in industry, this type of unified database provides benefits in generating reports on process and product performance, particularly during audits of the manufacturing plant. Of course, if one wishes to implement artificial neural networks or other data-driven analysis methods, consistent and frequent data generation is invaluable. By implementing an automated framework such as this, the workload of generating such data is also significantly reduced.

# Internet of Things in Biopharmaceutical Downstream Processes

The Internet of Things (IoT) has gained significant attention across various industries for its potential to interconnect devices and systems, enabling seamless data exchange and enhanced control over complex operations. In the realm of biopharmaceutical manufacturing, the application of IoT presents unique opportunities and challenges, particularly in downstream processes. However, the usefulness of IoT in biomanufacturing is not always clear due to the presence of established control systems such as Supervisory Control and Data Acquisition (SCADA) and proprietary software provided with commercial equipment.

This chapter explores the role and potential of IoT in biopharmaceutical downstream processes, focusing on its relevance to system integration, data acquisition, and process control. Drawing on my experience with the Orbit software framework, I argue that Orbit, which enables flexible access to sensor data and facilitates communication between independent systems over a local network, represents an novel application of IoT in biomanufacturing research. In particular, the Quality Analysis System in Papers V and VI, which links multiple ÄKTA chromatography systems, serves as a compelling example of IoT enabling a degree of system integration otherwise not possible with traditional software configurations.

## The Challenge of IoT in Biopharmaceutical Manufacturing

The downstream processes in biopharmaceutical manufacturing—such as protein purification through chromatography—are traditionally controlled by sophisticated software systems like SCADA, which provide centralized monitoring and control over various unit operations. These systems are integral to ensuring process stability, regulatory compliance, and product quality. However, SCADA systems are typically closed-loop environments, which limit flexibility in how data is accessed and shared across different platforms. They are designed to optimize

performance within the boundaries of a predefined process flow, making system modifications and custom integration challenging.

The rigid architecture of SCADA systems creates barriers for the widespread application of IoT in downstream processing. IoT typically requires open communication protocols that allow devices and systems to share data in real-time, enabling the creation of a connected, intelligent network of equipment that can respond dynamically to changes in the process. In biomanufacturing, such interconnectivity could allow for more modular and creative process designs, leading to improvements in efficiency, data acquisition, and process optimization. However, traditional SCADA-based environments are often incompatible with the decentralized, flexible approach required for true IoT implementation.

# Orbit: A Flexible IoT Framework for Biomanufacturing

Despite the challenges posed by traditional control systems, my work with Orbit demonstrates how an IoT-based approach can be successfully implemented in biopharmaceutical downstream processes. Orbit is a Python-based software framework that provides direct access to sensors and instruments, allowing users to customize control strategies and data acquisition in ways that SCADA systems do not typically allow. More importantly, it enables multisystem integration by facilitating communication between different unit operations, such as chromatography systems, over a local area network (LAN).

In our work on the Quality Analysis System (QAS), we utilized Orbit to connect multiple ÄKTA chromatography systems. Each system operated independently but communicated with the others through the LAN, exchanging data and synchronizing operations. An overview of this application is shown in the previous chapter, Figure 13. In the study in Paper VI, the transfer of the capture pools from the capture step to the polishing step was enabled by synchronizing the valves and pumps of the ÄKTA PCC and the ÄKTA Pure systems. The QAS acted as an external client to these two systems, gathering samples and performing analyses automatically. Finally, a buffer preparation system was implemented that provided the individual systems with buffers automatically as they were close to running out. The communication framework established by Orbit allowed the systems to work together in a coordinated manner, ensuring the timing of sample collection and preparation, as well as the buffer preparation, was synchronized with the ongoing purification process.

This ability to interconnect multiple systems and manage them through a unified communication network reflects the core principles of IoT. The Orbit framework provided the flexibility to work beyond the predefined configurations of individual ÄKTA systems, allowing for a more modular and adaptable approach to process

design. This modularity, which is often difficult to achieve with SCADA systems, enables a level of creativity in process development that is crucial for the advancement of digitalization in biomanufacturing.

## IoT as an Enabler of Digitalization: Digital Twins and Big Data

The flexibility offered by IoT-enabled systems like Orbit is critical for the broader implementation of digital twins and big data methodologies in biopharmaceutical downstream processes. For digital twins to function effectively, real-time data from various unit operations must be continuously fed into the virtual model to ensure it remains reflective of the physical process. In traditional setups, accessing such data across multiple systems can be a significant challenge, particularly when these systems are siloed under different control platforms. However, with an IoT framework, data from different sources can be seamlessly integrated, enabling the continuous flow of information needed for accurate and timely digital twin updates.

Moreover, IoT-based systems like Orbit allow for the direct capture of data from sensors, which can then be used to drive process optimization and advanced control strategies such as real-time control and monitoring. In my multisystem applications, the data obtained through Orbit's direct sensor access was used to develop and refine digital twins for chromatography processes. Additionally, this data was leveraged for real-time process monitoring using Kalman filters in Paper IV, providing accurate state estimates. Without the ability to directly access the sensor data in real time via Python, it would have been impossible to develop the iterative learning controller in Papers I and II and the Kalman filters at all.

The Quality Analysis System is a perfect example of this integration. By connecting multiple chromatography systems via Orbit, the QAS enabled real-time sampling, data acquisition, and analysis that fed directly into process models. This modularity and ability to directly access and share data is a key benefit of IoT, which can be extended to other biopharmaceutical manufacturing operations to enhance process control and data-driven decision-making.

## Modular Process Design Through IoT

One of the most significant advantages of IoT in biomanufacturing is the ability to create modular process designs. With traditional systems, processes are often constrained by the rigid architecture of the control system, making it difficult to modify workflows or integrate new technologies. However, by utilizing IoT

frameworks like Orbit, it becomes possible to introduce new unit operations, sensors, or control strategies into an existing process without the need for extensive system reconfiguration.

In my work, this modular approach has allowed for greater flexibility during the process design stage. For example, the Orbit software enabled me to connect different chromatography systems and analytical devices in a way that was not possible using standard configurations. This flexibility allowed me to experiment with different sampling strategies, analytical methods, and control schemes, ultimately leading to more optimized and innovative process designs.

The modularity introduced by IoT-based systems also has implications for scalability. By creating a flexible, interconnected network of devices, it becomes easier to scale processes up or down depending on production needs. New devices can be seamlessly integrated into the network, and process control can be adapted to account for changes in production scale or product complexity. This adaptability is crucial for biopharmaceutical manufacturers looking to transition from lab-scale production to pilot or commercial-scale manufacturing while maintaining process robustness and compliance.

# Conclusion

The application of IoT in biopharmaceutical downstream processes offers significant potential for improving system integration, data acquisition, and process control. While traditional control systems such as SCADA limit the flexibility of IoT implementation, the use of Orbit demonstrates how IoT principles can be applied to create a more modular and interconnected process environment. By enabling communication between independent systems, Orbit allows for greater creativity in process design and supports the development of advanced digitalization tools such as digital twins, big data analytics, and real-time process control.

As biopharmaceutical manufacturing continues to evolve, the ability to integrate IoT-based frameworks will be essential for achieving the level of digitalization required to meet the industry's growing demands for efficiency, scalability, and product quality. IoT, when combined with flexible software like Orbit, represents a key enabler of this transformation, paving the way for more modular, adaptive, and data-driven approaches to downstream processing.

# Conclusions

Digital twins can be applied in different ways for different purposes. In my work, I have shown all three versions of digital twins: a digital model was used to develop an iterative learning controller for linear gradient chromatography, which allowed for extensive trials without spending lab resources; a digital shadow was developed by applying state observer techniques called Kalman filters to improve monitoring and resolution of coeluting proteins, achieving a real-time flow of data from the physical to the digital; and a digital twin was used applied by means of an iterative learning controller, where two-way flow of information was achieved by computing the residence times of two proteins, and using them in a data-driven model to obtain an improved set of operating parameters.

In addition to showing the use of different kinds of digital twins, I have also showcased an automated, integrated framework for developing mechanistic models for use in digital twins. Integrated in this case refers to the way in which the same code used to operate the chromatography system is used to simulate it, as well as to the manner in which the full system is included in the simulation: tubes, valves, sensors, mixers and chromatography columns alike.

Big data analytics pose a major challenge in their application due to the cost of generating the large sets of data that are required. In this thesis, I have demonstrated how the collection, preparation and analysis of samples from a continuous downstream process can be automated using interconnected chromatography systems and network communications. By designing automated systems for obtaining high-quality data, a small step has been taken in the direction of big data applications in biopharmaceutical downstream processes.

All of the solutions covered in this thesis were made possible through the use of the Orbit software, which allows for control, data acquisition and network communication. Operation of complex, interconnected processes with support systems for sample collection and analysis is enabled, which shows how the application of Internet of Things principles and the modularity that follows makes innovation possible.

In conclusion, Industry 4.0 and digitalization can be realized by combining existing technology, such as mechanistic modelling and state observers such as Kalman filters, with the modern available tools such as network communication, machine learning algorithms and parallel computing. This thesis showcases several examples

of such applications. An important factor for future innovation is the availability of accessible APIs in process equipment, which will open the doors for more creative solutions to existing problems.

# References


[1] G. F. Schimizzi, 'Biosimilars from a practicing rheumatologist perspective: An overview', *Autoimmunity Reviews*, vol. 15, no. 9, pp. 911–916, Sep. 2016, doi: 10.1016/j.autrev.2016.07.001.

[2] V. Girard *et al.*, 'Large-scale monoclonal antibody purification by continuous chromatography, from process design to scale-up', *Journal of Biotechnology*, vol. 213, pp. 65–73, Nov. 2015, doi: 10.1016/j.jbiotec.2015.04.026.

[3] J. Pollock, G. Bolton, J. Coffman, S. V. Ho, D. G. Bracewell, and S. S. Farid, 'Optimising the design and operation of semi-continuous affinity chromatography for clinical and commercial manufacture', *Journal of Chromatography A*, vol. 1284, pp. 17–27, Apr. 2013, doi: 10.1016/j.chroma.2013.01.082.

[4] A. Jungbauer, 'Continuous downstream processing of biopharmaceuticals', *Trends in Biotechnology*, vol. 31, no. 8. pp. 479–492, 2013. doi: https://doi.org/10.1016/j.tibtech.2013.05.011.

[5] S. S. Farid, M. Baron, C. Stamatis, W. H. Nie, and J. Coffman, 'Benchmarking biopharmaceutical process development and manufacturing cost contributions to R&D', *Mabs*, vol. 12, no. 1. 2020. doi: https://doi.org/10.1080/19420862.2020.1754999.

[6] P. Parviainen, M. Tihinen, J. Kääriäinen, and S. Teppola, 'Tackling the digitalization challenge: how to benefit from digitalization in practice', *IJISPM*, vol. 5, no. 1, pp. 63–77, Feb. 2022, doi: 10.12821/ijispm050104.

[7] M. Gradillas and L. D. W. Thomas, 'Distinguishing digitization and digitalization: A systematic review and conceptual framework', *Journal of Product Innovation Management*, vol. n/a, no. n/a, Jul. 2023, doi: 10.1111/jpim.12690.

[8] H. J. Feise and E. Schaer, 'Mastering digitized chemical engineering', *Education for Chemical Engineers*, vol. 34, pp. 78–86, Jan. 2021, doi: 10.1016/j.ece.2020.11.011.

[9] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, 'About The Importance of Autonomy and Digital Twins for the Future of Manufacturing', *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, Jan. 2015, doi: 10.1016/j.ifacol.2015.06.141.

[10] M. Bisschops and L. Cameron, 'Process Intensification and Industry 4.0: Mutually Enabling Trends', in *Process Control, Intensification, and Digitalisation in Continuous Biomanufacturing*, 2022, pp. 209–229. doi: 10.1002/9783527827343.ch7.

[11] H. Narayanan *et al.*, 'Bioprocessing in the Digital Age: The Role of Process Models', *Biotechnology Journal*, vol. 15, no. 1, p. 1900172, Jan. 2020, doi: 10.1002/biot.201900172.

[12] I. A. Udugama, C. Bayer, S. Baroutian, K. V. Gernaey, W. Yu, and B. R. Young, 'Digitalisation in chemical engineering: Industrial needs, academic best practice, and curriculum limitations', *Education for Chemical Engineers*, vol. 39, pp. 94–107, Apr. 2022, doi: 10.1016/j.ece.2022.03.003.

[13] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, 'Digital Twin in manufacturing: A categorical literature review and classification', *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, Jan. 2018, doi: 10.1016/j.ifacol.2018.08.474.

[14] D. Gahan, 'Internet of Things and its Implications for the Process Industries', *The Chemical Engineer*, no. 988, Sep. 2023. [Online]. Available: https://www.thechemicalengineer.com/features/internet-of-things-and-its-implications-for-the-process-industries/

[15] J. Wan *et al.*, 'Software-Defined Industrial Internet of Things in the Context of Industry 4.0', *IEEE Sensors Journal*, vol. 16, pp. 7373–7380, 2016.

[16] J. Scheffel *et al.*, 'Design of an integrated continuous downstream process for acid-sensitive monoclonal antibodies based on a calcium-dependent Protein A ligand', *J Chromatogr A*, vol. 1664. p. 462806, 2022. doi: 10.1016/j.chroma.2022.462806.

[17] H. Schwarz *et al.*, 'Integrated continuous biomanufacturing on pilot scale for acid-sensitive monoclonal antibodies', *Biotech & Bioengineering*, vol. 119, no. 8, pp. 2152–2166, Aug. 2022, doi: 10.1002/bit.28120.

[18] Cytiva, 'MAb polishing step development using Capto$^{TM}$ adhere in flow-through mode', Uppsala, Sweden, Procedure 29-0192-56 AA, Jun. 2012.

[19] N. Andersson, A. Löfgren, M. Olofsson, A. Sellberg, and B. Nilsson, 'The Orbit Controller', Lund University, Report, 2016.

[20] N. Andersson, A. Löfgren, M. Olofsson, A. Sellberg, B. Nilsson, and P. Tiainen, 'Design and control of integrated chromatography column sequences', *Biotechnology Progress*, vol. 33, no. 4. pp. 923–930, 2017. doi: https://doi.org/10.1002/btpr.2434.

[21] S. Tallvod, M. Yamanee-Nolin, J. Gomis-Fons, N. Andersson, and B. Nilsson, 'A novel process design for automated quality analysis in an integrated biopharmaceutical platform', in *Computer Aided Chemical Engineering*, vol. 51, L. Montastruc and S. Negny, Eds., Elsevier, 2022, pp. 619–624. doi: 10.1016/B978-0-323-95879-0.50104-1.

[22] A. S. Rathore and H. Winkle, 'Quality by design for biopharmaceuticals', *Nature Biotechnology*, vol. 27, no. 1. pp. 26–34, 2009. doi: 10.1038/nbt0109-26.

[23] M. Degerman, K. Westerberg, and B. Nilsson, 'A Model-Based Approach to Determine the Design Space of Preparative Chromatography', *Chemical Engineering & Technology*, vol. 32, no. 8, pp. 1195–1202, Aug. 2009, doi: 10.1002/ceat.200900102.

[24] K. Westerberg, E. Broberg Hansen, M. Degerman, T. Budde Hansen, and B. Nilsson, 'Model-Based Process Challenge of an Industrial Ion-Exchange Chromatography Step', *Chemical Engineering & Technology*, vol. 35, no. 1, pp. 183–190, Jan. 2012, doi: 10.1002/ceat.201000560.

[25] K. Suzuki, T. Yajima, and Y. Kawajiri, ‘Process Robustness Evaluation for Various Operating Configurations of Simulated Moving Bed Chromatography’, in *Computer Aided Chemical Engineering*, vol. 53, F. Manenti and G. V. Reklaitis, Eds., Elsevier, 2024, pp. 229–234. doi: 10.1016/B978-0-443-28824-1.50039-9.

[26] G. Dünnerbier *et al.*, ‘Model-based Control of Batch Chromatography’, *AIChE Journal*, vol. 47, no. 11. pp. 2493–2502, 2001. doi: https://doi.org/10.1002/aic.690471112.

[27] C. Grossmann, G. Ströhlein, M. Morari, and M. Morbidelli, ‘Optimizing model predictive control of the chromatographic multi-column solvent gradient purification (MCSGP) process’, *Journal of Process Control*, vol. 20, no. 5, pp. 618–629, Jun. 2010, doi: 10.1016/j.jprocont.2010.02.013.

[28] J. Gomis-Fons *et al.*, ‘Model-based design and control of a small-scale integrated continuous end-to-end mAb platform’, *Biotechnology Progress*, vol. 36, no. 4. p. e2995, 2020. doi: https://doi.org/10.1002/btpr.2995.

[29] F. Feidl *et al.*, ‘Process-wide control and automation of an integrated continuous manufacturing platform for antibodies’, *Biotechnology and Bioengineering*, vol. 117, no. 5, pp. 1367–1380, May 2020, doi: 10.1002/bit.27296.

[30] A. Löfgren, J. Gomis-Fons, N. Andersson, B. Nilsson, L. Berghard, and C. Lagerquist Hägglund, ‘An integrated continuous downstream process with real-time control: A case study with periodic countercurrent chromatography and continuous virus inactivation’, *Biotechnology and Bioengineering*, vol. 118, no. 4, pp. 1645–1657, Apr. 2021, doi: 10.1002/bit.27681.

[31] A. Holmqvist and A. Sellberg, ‘A Generic PAT Software Interface for On-Line Monitoring and Control of Chromatographic Separation Systems’, *Computer Aided Chemical Engineering*, vol. 38. pp. 811–816, 2016. doi: https://doi.org/10.1016/B978-0-444-63428-3.50140-5.

[32] M. Sokolov, M. von Stosch, H. Narayanan, F. Feidl, and A. Butté, ‘Hybrid modeling — a key enabler towards realizing digital twins in biopharma?’, *Current Opinion in Chemical Engineering*, vol. 34, p. 100715, Dec. 2021, doi: 10.1016/j.coche.2021.100715.

[33] C. A. Brooks and S. M. Cramer, ‘Steric mass-action ion exchange: Displacement profiles and induced salt gradients’, *AIChE Journal*, vol. 38, no. 12, pp. 1969–1978, 1992, doi: 10.1002/aic.690381212.

[34] E. von Lieres and J. Andersson, ‘A fast and accurate solver for the general rate model of column liquid chromatography’, *Computers & Chemical Engineering*, vol. 34, no. 8, pp. 1180–1191, Aug. 2010, doi: 10.1016/j.compchemeng.2010.03.008.

[35] S. Tallvod, N. Andersson, and B. Nilsson, ‘Automation of Modeling and Calibration of Integrated Preparative Protein Chromatography Systems’, *Processes*, vol. 10, no. 5, 2022, doi: 10.3390/pr10050945.

[36] F. Rischawy *et al.*, ‘Integrated process model for the prediction of biopharmaceutical manufacturing chromatography and adjustment steps’, *Journal of Chromatography A*, vol. 1681, p. 463421, Oct. 2022, doi: 10.1016/j.chroma.2022.463421.

[37] S. Nargund, K. Guenther, and K. Mauch, ‘The Move toward Biopharma 4.0’, *Genetic Engineering & Biotechnology News*, vol. 39, no. 6, pp. 53–55, Jun. 2019, doi: 10.1089/gen.39.06.18.

[38] M. Krippl, A. Dürauer, and M. Duerkop, ‘Hybrid modeling of cross-flow filtration: Predicting the flux evolution and duration of ultrafiltration processes’, *Separation and Purification Technology*, vol. 248, p. 117064, Oct. 2020, doi: 10.1016/j.seppur.2020.117064.

[39] M. Krippl, I. Bofarull-Manzano, M. Duerkop, and A. Dürauer, ‘Hybrid modeling for simultaneous prediction of flux, rejection factor and concentration in two-component crossflow ultrafiltration’, *Processes*, vol. 8, no. 12, pp. 1–24, 2020, doi: 10.3390/pr8121625.

[40] M. Krippl, T. Kargl, M. Duerkop, and A. Dürauer, ‘Hybrid modeling reduces experimental effort to predict performance of serial and parallel single-pass tangential flow filtration’, *Separation and Purification Technology*, vol. 276, 2021, doi: 10.1016/j.seppur.2021.119277.

[41] R. Faragher, ‘Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation [Lecture Notes]’, *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 128–132, Sep. 2012, doi: 10.1109/MSP.2012.2203621.

[42] F. Auger, M. Hilairet, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura, ‘Industrial Applications of the Kalman Filter: A Review’, *IEEE Transactions on Industrial Electronics*, vol. 60, no. 12, pp. 5458–5471, Dec. 2013, doi: 10.1109/TIE.2012.2236994.

[43] R. Pishkari, M. Fechtner, T. Keßler, and A. Kienle, ‘Optimization of Simulated Moving Bed Chromatographic Processes using Surrogate Models’, in *Computer Aided Chemical Engineering*, vol. 52, A. C. Kokossis, M. C. Georgiadis, and E. Pistikopoulos, Eds., Elsevier, 2023, pp. 343–348. doi: 10.1016/B978-0-443-15274-0.50055-X.

[44] A. S. Rathore, S. Nikita, G. Thakur, and S. Mishra, ‘Artificial intelligence and machine learning applications in biopharmaceutical manufacturing’, *Trends in Biotechnology*, vol. 41, no. 4, pp. 497–510, Apr. 2023, doi: 10.1016/j.tibtech.2022.08.007.

[45] S. Sachio, B. Likozar, C. Kontoravdi, and M. M. Papathanasiou, ‘Computer-aided design space identification for screening of protein A affinity chromatography resins’, *Journal of Chromatography A*, vol. 1722, p. 464890, May 2024, doi: 10.1016/j.chroma.2024.464890.

[46] A. W. Rogers, I. O. S. Cardenas, E. A. Del Rio-Chanona, and D. Zhang, ‘Investigating physics-informed neural networks for bioprocess hybrid model construction’, in *Computer Aided Chemical Engineering*, vol. 52, A. C. Kokossis, M. C. Georgiadis, and E. Pistikopoulos, Eds., Elsevier, 2023, pp. 83–88. doi: 10.1016/B978-0-443-15274-0.50014-7.

[47] S.-Y. Tang, Y.-H. Yuan, Y.-C. Chen, S.-J. Yao, Y. Wang, and D.-Q. Lin, ‘Physics-informed neural networks to solve lumped kinetic model for chromatography process’, *Journal of Chromatography A*, vol. 1708, p. 464346, Oct. 2023, doi: 10.1016/j.chroma.2023.464346.