Tools for digital twins in continuous downstream processing

Tallvod, Simon

2023

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

Tools for digital twins in continuous downstream processing

SIMON TALLVOD | CHEMICAL ENGINEERING | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

ISBN: 978-91-7422-996-7

Chemical Engineering
Faculty of Engineering
Lund University

Tools for digital twins in continuous downstream processing

Simon Tallvod

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 the 24th of November at 09.00, lecture hall DC:Stora hörsalen, Ingvar Kamprad Design Center, Sölvegatan 26, Lund

Faculty opponent
Priv.-Doz. Dipl.-Ing. Dr. Astrid Dürauer
University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation **Date of issue:** 2023-11-24

Author: Simon Tallvod

Title and subtitle: Tools for digital twins in continuous downstream processing

Sponsoring organization: Swedish Governmental Agency for Innovation Systems (VINNOVA)

Abstract:

Biopharmaceuticals are of ever-growing importance for human health. The development and production of biopharmaceuticals has the potential for efficiency improvements by switching from batch-wise to continuous production. This change demands more advanced process control in the form of automation and digital tools such as digital twins, which in turn need adequate mathematical models and accessible process data.

This thesis discusses the concept of digital twins as well as the automation of modelling and data acquisition in integrated continuous downstream processes for the development and production of biopharmaceuticals. Utilizing in-house supervisory control software for chromatographic systems enables automation of modeling, calibration, sampling, and analysis in integrated downstream processes, advancing the development of digital twins.

Key words: Digital twins, biopharmaceuticals, downstream processing, chromatography, mathematical modeling, ion exchange, automation, sampling, model calibration, python

Language: English

ISBN (print): 978-91-7422-996-7 **ISSN (pdf):** 978-91-7422-997-4

Number of pages: 130

Tools for digital twins in continuous downstream processing

Simon Tallvod

LUND
UNIVERSITY

Coverphoto by Simon Tallvod, using *Midjourney*

Faculty of Engineering

Department of Chemical Engineering

ISBN 978-91-7422-996-7 (print)

ISSN 978-91-7422-997-4 (pdf)

Printed in Sweden by Media-Tryck, Lund University

Lund 2023

Acknowledgements

I would like to express my gratitude to my supervisors Bernt Nilsson and Niklas Andersson for their diligent support during my time as a PhD student and especially during the writing of this thesis.

I would also like to thank my colleagues, both past and present, at the Department of Chemical Engineering for the hospitable work environment and the many interesting and enjoyable discussions around the tables in the breakroom.

I especially want to thank Mikael Yamanee-Nolin for all the support and guidance throughout my time at the department and for always having a positive outlook on things. Madelène Isaksson for her valuable input and for always making time for discussions and proofreading. Daniel Espinoza for always being there and for the great collaborations, as well as always taking the time for a round of table tennis.

Lastly, I wish to express my heartfelt thanks to my friends and family for their support and understanding. I am especially grateful to Josefin for her constant support and for always having my back.

Abstract

Biopharmaceuticals are of ever-growing importance for human health. The development and production of biopharmaceuticals has the potential for efficiency improvements by switching from batch-wise to continuous production. This change demands more advanced process control in the form of automation and digital tools such as digital twins, which in turn need adequate mathematical models and accessible process data.

This thesis discusses the concept of digital twins as well as the automation of modelling and data acquisition in integrated continuous downstream processes for the development and production of biopharmaceuticals. Utilizing in-house supervisory control software for chromatographic systems enables automation of modeling, calibration, sampling, and analysis in integrated downstream processes, advancing the development of digital twins.

Populärvetenskaplig sammanfattning

Biologiska läkemedel ger nya möjligheter

Biologiska läkemedel, alltså läkemedel som har bildats i en levande organism, är en väldigt utbredd kategori av läkemedel. Biologiska läkemedel har länge använts för att förebygga och behandla sjukdomar. Som exempel kan nämnas att 1923 års Nobelpris i medicin tilldelades Frederick Banting och James Macleod för deras upptäckt av insulin, som på den tiden tillverkades genom utvinning från bukspottskörtlar från grisar och kor. Utvecklingen har dock gått framåt och idag framställs insulin, och de flesta andra biologiska läkemedel, genom att odla genmodifierade jäst- bakterie- eller däggdjursceller i bioreaktorer Denna del av läkemedelstillverkningen brukar kallas uppströms. Den här avhandlingen behandlar vad som händer nedströms, det vill säga efter att läkemedlet har odlats i produktionsstegen.

Ett framväxande område inom biomedicin är de så kallade avancerade terapiläkemedlen, på engelska *Advanced Therapy Medicinal Products* eller ATMP. Till dessa hör till exempel cell- och genterapi och de är läkemedel som på sikt kan komma att användas för att behandla eller till och med bota sjukdomar som cancer, diabetes och Parkinsons. Dessa läkemedel måste ofta skräddarsys för varje patient och tillverkningen blir därmed mycket resurskrävande och komplicerad i förhållande till mer traditionella biologiska läkemedel.

Kontinuerliga nedströmsprocesser

Nedströmsprocesser är de processteg som sker efter att lösningen med ett läkemedel har lämnat tanken där det odlats. I detta skede innehåller vätskan mycket annat som måste tas bort för att få ett rent läkemedel som kan användas till patienter. Dessa föroreningar kan till exempel vara rester av celler, som DNA eller endotoxiner, odlingsmedium och proteiner. För att ta bort föroreningarna används ofta en metod som kallas kromatografi och det är den metoden som har använts i den här avhandlingen. Kromatografi går ut på att separera ämnen efter fysikaliska och kemiska egenskaper som till exempel storlek eller laddning. Principen är att en vätska med de olika ämnena pumpas genom en kolonn som är packad med ett poröst material. Kolonnens packning har en viss egenskap som gör att olika ämnen stannar olika länge i kolonnen och de olika ämnena kan på så vis samlas upp vid olika tidpunkter.

Traditionellt så har dessa separationssteg körts för hand och satsvis. Detta är inte effektivt och det kräver mycket manuellt arbete. Läkemedelsindustrin håller nu på att röra sig mot att sätta ihop flera separationssteg och att automatisera hela processer. Detta brukar kallas för kontinuerliga nedströmsprocesser.

Smartare tillverkning med automation

Genom att använda mer automation, det vill säga att låta processer sköta sig och kunna ta vissa beslut själva, så kan forskningen, utvecklingen och tillverkningen av biologiska läkemedel effektiviseras. Detta innebär att resurser kan sparas och att tiden för tillverkningen kan minskas. Alltså skulle vi kunna få ut rätt läkemedel till patienter på kortare tid, för mindre pengar och med potentiellt lägre påverkan på miljön. Med mer automation skulle man också kunna minska behovet av expertkunskap för att hantera slutproduktionen av skräddarsydda läkemedel vilket betyder att tillverkningen skulle kunna ske närmare patienten, till exempel på ett lokalt sjukhus.

Verktyg för digitala tvillingar

För att uppnå visionen om automatiserade processer för biologiska läkemedel behövs ett antal verktyg och i den här avhandlingen har jag specifikt tittat på någonting som kallas för *digitala tvillingar*.

En digital tvilling kan ses som en digital representation av ett fysiskt objekt, en process eller ett förlopp. Det vill säga en datorversion av någonting. Det finns olika nivåer av digitala tvillingar där de enklaste bara beskriver någon egenskap hos ett objekt och där egenskapen inte automatiskt uppdateras när det verkliga objektet förändras. Ett väldigt enkelt exempel är om man hade mätt temperaturen i ett badkar då och då och uppdaterat temperaturen i sin digitala representation av badkaret manuellt. Denna digitala tvilling hade så klart haft ett ganska begränsat användningsområde men man hade haft ett hum om vad temperaturen ungefär var förutsatt att man mätte ofta eller om temperaturen inte förändrade sig så mycket. Om man dessutom sparar mätpunkterna hade man eventuellt kunnat lära sig något om hur badkarets temperatur förändras med tiden.

Nästa nivå av digitala tvillingar använder automatisk datainsamling. För exemplet med badkaret hade det inneburit att vi hade mätt temperaturen med en temperaturgivare som automatiskt uppdaterade värdet i den digitala tvillingen. Nu hade man kanske kunnat upptäcka trender i temperaturen automatiskt och till exempel larma om temperaturen understiger ett visst värde.

I tredje och sista nivån av digitala tvillingar går data åt bägge hållen. Det innebär att den digitala tvillingen också skickar signaler till det fysiska systemet. I exemplet ovan hade det kunnat innebära att när temperaturen i badkaret sjunker under ett visst värde så hade varmvattenkranen öppnats en stund för att fylla på varmare vatten.

Gemensamt för alla digitala tvillingar är att de grundar sig på någon typ av modell. En modell är helt enkelt en beskrivning av ett objekt eller förlopp och hur beskrivningen ser ut beror på vad modellen ska användas till. Inom tillämpningarna i den här avhandlingen sker beskrivningen med hjälp av matematik och modellerna

kallas följaktligen för matematiska modeller. De fungerar genom att beskriva fysikaliska fenomen och tillstånd med ekvationer och samband. Matematiska modeller innehåller någonting som kallas *parametrar* och dessa bestämmer hur modellen uppför sig när man simulerar, det vill säga beräknar, den. För att få bra modeller som beskriver system på ett tillfredställande sätt måste dessa parametrar hittas och det görs med en så kallad kalibrering.

Kalibrering kan göras på flera sätt och det enklaste är att för hand beräkna modellen och justera parametrarna tills resultatet överensstämmer tillräckligt bra med verkliga data. Detta är mycket tidskrävande och kräver en stor arbetsinsats om den ens är möjligt för avancerade processer. Därför har jag i den här avhandlingen tagit fram verktyg och metoder för att automatiskt kunna kalibrera matematiska modeller och på så vis enkelt kunna ta fram avancerade matematiska beskrivningar av kontinuerliga nedströmsprocesser.

List of Papers

Paper I

Tallvod, S., Andersson, N. and Nilsson, B. Automation of Modeling and Calibration of Integrated Preparative Protein Chromatography Systems. *Processes.* 2022, 10, 945.

Paper II

Espinoza, D*., Tallvod, S*., J., Andersson, N. and Nilsson, B. Automatic procedure for modelling, calibration, and optimization of a three-component chromatographic separation. *Manuscript.*

Paper III

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

Paper IV

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 proceedings of 32 *European Symposium on Computer Aided Process Engineering*. 2022, 51, 619–624.

Paper V

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

* shared first authorship.

Other related publications

Löfgren, A., Yamanee-Nolin, M., Tallvod, S., Gomis-Fons, J., Andersson, N. and Nilsson, B. Optimization of integrated chromatography sequences for purification of biopharmaceuticals. Biotechnology Progress. 2019. 35, 6, e2871.

Sjölin, M., Sayed, M., Espinoza, D., Tallvod, S., 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*

Author's contribution to the papers

Paper I

I developed, planned, and performed all the experiments, did conceptualization, wrote code for experiments and model calibration, and wrote the paper with input from my co-authors.

Paper II

I developed, planned, and performed the experiments, did conceptualization, wrote code for calibration and optimization, and wrote the paper, all together with Daniel Espinoza.

Paper III

I designed the system, planned, and performed the experiments, wrote code, and wrote the paper, all together with Daniel Espinoza. Joaquín Gomis Fons contributed with code and design for communication between the systems, as well as expertise on PCC systems.

Paper IV

I did the conceptualization, design, and development and planned and performed all laboratory work. I wrote the paper together with Mikael Yamanee-Nolin.

Paper V

I planned and performed parts of the experiments, did conceptualization, design, and development for parts of the work described in the paper.

Table of Contents

1. Introduction

Biopharmaceuticals are becoming increasingly vital in the treatment of a multitude of medical conditions ranging from diabetes to cancer, as well as for preventative healthcare, for example in the form of vaccines. Biopharmaceuticals are defined as pharmaceuticals derived from living organisms and they are commonly produced by cells grown in bioreactors. The market share of biopharmaceuticals is growing every year as is the approval rate of new biopharmaceutical drugs (Hong et al., 2018; Moorkens et al., 2017; Walsh & Walsh, 2022).

Traditionally, the production of biopharmaceuticals has been carried out batch-wise, which is time and labour intensive, and is increasingly seen as an ineffective means of production. The industry is now moving towards more continuous production to increase efficiency, process robustness, and reproducibility. This shift is being accomplished by increasing the level of automation, which, in turn, enables better process control (Andersson et al., 2017; Arnold et al., 2019; Godawat et al., 2012, 2015; Gomis-Fons, Andersson, et al., 2020; Gomis-Fons, Schwarz, et al., 2020; Jungbauer, 2013; Konstantinov & Cooney, 2015; Sellberg et al., 2017).

To achieve better process automation, concepts such as digital twins and the tools for their implementation are introduced. Digital twins are digital representations of physical systems that are used for everything from process design and monitoring to visualization of production systems. The development of digital twins demand accurate and reliable mathematical models of processes as well as ample and high-quality data (Chen et al., 2020; Grossmann et al., 2010; Narayanan et al., 2022; Patnaik, 1999; Rathore, Mishra, et al., 2021; Rathore, Nikita, et al., 2021).

In our research group at the Department of Chemical Engineering at Lund University we have developed digital and hardware solutions for smart production of biopharmaceuticals which we cover in further detail in Paper V. For example we have developed a platform for the production of monoclonal antibodies in both lab and pilot scale (Scheffel et al., 2022; Schwarz et al., 2022), an automated buffer preparation system (Isaksson et al., 2023), and an automated quality analysis system (Tallvod et al., 2023). These solutions all contribute to the development of better process control strategies such as digital twins.

Aim

In this thesis I will present tools for the facilitation of digital twins in the production and development of biopharmaceuticals. I have developed methods for automatic modelling and calibration of integrated downstream processes, opening the door for better process automation and control (Papers I and II). I have also investigated the possibilities for automated sampling and analysis of downstream processes, enabling advanced process monitoring and control as well as widening the possibilities for automated product quality control (Papers III and IV).

Outline

This thesis consists of six chapters: Chapter 1 introducing the research topic, and Chapter 2 presenting the concept of digital twins. In Chapter 3, I discuss our in-house developed supervisory control software Orbit. Chapter 4 discusses the main findings of Papers III and IV and it covers Automated advanced quality analysis. Chapter 5 addresses automatic model generation and calibration and it includes the main findings of Papers I and II. Finally, Chapter 6 concludes the work presented in this thesis.

2. Digital twins

With digital twins we take the concept of mathematical modelling further by using mathematical models of some system and using them in tandem with the physical system. The term *digital twin* was introduced by Dr. Michael Grieves in 2003 and is defined as having three parts: a physical or real-world part, a digital representation of the physical part, and a link between them (Grieves, 2015). This link can have different forms depending on the application and how the digital twin is implemented, but it always consists of some kind of data, be it manually transferred experimental results used for model calibration or digitally sent signals for control actuation.

Three levels of digital twins

Kritzinger et al. (2018) defines the following taxonomy of digital twins where they identify three different levels of digital twins: digital models, digital shadows, and digital twins, see Figure 2.

A **digital model** is a representation of a physical object that has no automatic data flow between the digital object and the physical object. If digital data is used it is transferred manually and not in real-time. The digital model remains unchanged and is not automatically updated if the physical object changes. Applications of digital models could include process development, and simulation and optimization of process systems.

In a **digital shadow**, there is a one-way flow of data between the physical object and the digital representation, and this flow is automated. For example, data is automatically transferred to the digital counterpart and the model automatically updates when there is a change of state in the physical object. In this case, there is still no automatic data transfer from the digital object to the physical. Here, possible uses may involve process monitoring and automatically updating simulators.

Using this framework, in a **digital twin** the flow of data from the digital representation to the physical object is also made automatic. This means that the digital twin is able to change the physical object in some way by some form of control action. An application might be model based control.

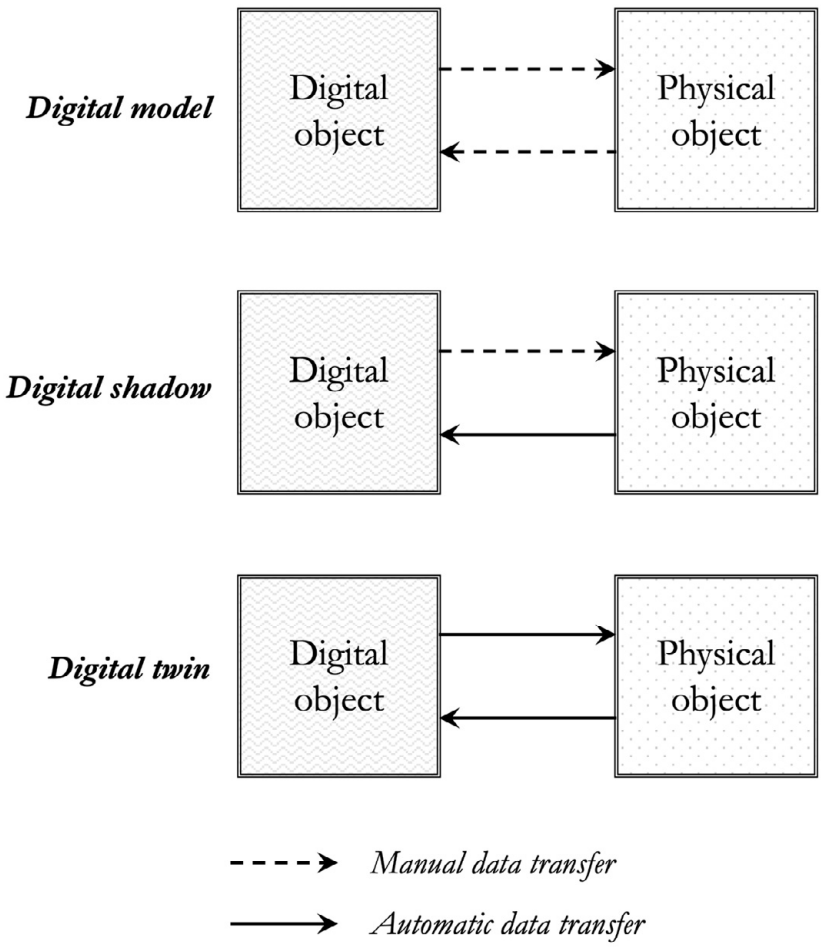

Figure 2
Illustration of the three levels of digital twins. At the top: Digital model. Data is only transferred manually in one or both directions. In the middle: Digital shadow. Data flows automatically in one of the directions. At the bottom: Digital twin. Data is flowing in both directions automatically.

Digital twins as a concept

What differentiates models from digital twins? In my opinion, it is what the intention of the representation is. The goal of a full digital twin is to be a near perfect representation of a system or a process where we want to acquire some knowledge beyond only having a representation. Grieves (2015) talks about how if we had a good digital twin of a factory, we could visualize data and the process itself so that for example we as humans could see trends and if products meet specifications:

> *"By using this information, we can change digital factory simulation, which attempts to predict how the product is to be manufactured, into a digital factory replication, which shows how the product is actually being manufactured. We can then compare it against the design specifications. This can occur in real time or near real-time. This provides a window onto the factory floor for anyone at any time from any place."*

He also emphasizes three concepts: conceptualization, comparison, and collaboration. He means that these concepts are powerful tools in the human tool kit that the concept of digital twins supports. Maybe a bit metaphysical but I interpret this as the more we could know about a process or system the more insights we could gain.

Digital twins in the biopharmaceutical field

The development and adaptation of digital twins in the biopharmaceutical industry is currently somewhat lagging in comparison to other fields and it is still very much an emerging technology. The current state of digital twins in the biopharmaceutical field is well covered in a review paper by Chen et al. (2020) and the two books *Digital Twins: Tools and Concepts for Smart Biomanufacturing* and *Digital Twins: Applications to the Design and Optimization of Bioprocesses* (Herwig et al., 2021b, 2021a). There is also promising research being done; for example, Gomis-Fons et al. (2020) developed a digital twin of a continuous monoclonal antibody platform, and more recently Helgers et al. (2022) demonstrated a digital twin of a complete production line of single-chain variable fragment in *E. coli* and Rischawy et al. (2023) showed a model of a fully integrated downstream purification process, as well as Isaksson et al. (2023) who demonstrated a digital twin of a buffer management system.

Challenges

There currently are some challenges when it comes to implementing digital twins in biopharmaceutical processes. One challenge is the heterogeneity in data formats and hardware interfaces among manufacturers, which complicates data handling and hardware integration. However, several manufacturers are moving towards supporting open platform communications (OPC) interfaces for their hardware.

Another challenge is the collection of data. More advanced models need high quality data, and they may need it relatively often in order to update and accurately mirror the modelled system. Therefore, a big hurdle in the implementation of digital twins is how to collect reliable data at a sufficient rate. This requires reliable process analytical technologies (PAT). Paper III suggests a solution for automatic and potentially autonomous sampling of a downstream process which enables several analyses and a variety of data to be collected and used in a digital twin. In Paper IV, we show the integration of a sensor from an analytical instrument into a preparative process, and we use it to collect UV-VIS spectral data in real-time. This data could be used in a digital twin's model in a multitude of ways, for example to see separate components in a co-eluting mixture as Brestrich et al. (2014) has demonstrated.

In addition to these challenges, there is also the question of computational cost of simulation. Large or complex mechanistical models can take considerable time to compute and may be too slow to be used in real or near-real-time scenarios in a practical way. This requires strategies to circumvent computationally costly models by reducing their complexity, use hybrid models, use data-driven or statistical models, or a combination of several modeling paradigms.

In this thesis I will show tools for use in the development of digital twins in integrated downstream processes for biopharmaceutical production and development.

3. Supervisory control software

In order to better control and monitor preparative chromatography systems, a software tool named *Orbit* was developed at the Department of Chemical Engineering at Lund University (Andersson, 2018; Andersson et al., 2017). It was originally designed to control ÄKTA™ chromatography systems from the manufacturer Cytiva (Uppsala, Sweden), but has since been extended to the control of other pieces of hardware as well, e.g., HPLC-systems from Agilent (California, USA). Orbit enables advanced script-based control of these systems and allows for the collection and logging of the systems' sensor data, as well as communication between multiple systems. In addition to its control and monitor abilities, Orbit has a built-in simulator that can simulate a process running on a physical system. Lastly, Orbit is equipped with a database where collected run data can be stored and later retrieved for analysis and evaluation. An overview of the Orbit architecture is shown in Figure 3.1 below.

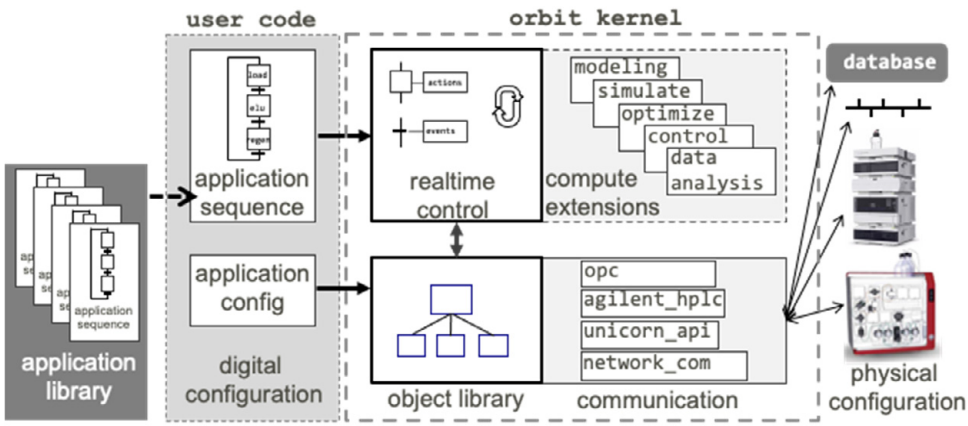

Figure 3.1.
Illustration showing working principle of the Orbit controller.

Implementation

Orbit is implemented using the Python programming language (Python Software Foundation, 2023; Van Rossum & Drake, 2009). Python is well-suited for this application for several reasons. One reason is that is a high-level language, which means that it is easy to use and fast to develop in. Another reason is that Python is open source, which means that there is a large user community that contributes to the code base and that there is a myriad of user-made software packages, as well as that it is easy to share solutions with anyone. Lastly, Python's user-friendliness and readable code makes it easy for anyone with minimal programming knowledge to get involved in development and to use the software. It should be noted that these benefits only hold true for applications in research and development. For commercial uses and applications in manufacturing, other aspects, e.g., reliability and safety weigh more heavily.

Orbit represents all units (such as valves, sensors, and columns) in the chromatography system as objects. The user defines which units are in the system, and if the user wishes to run the simulator, all the tubes and their connections need to be defined as well. A more in-depth discussion of the simulator is found in the sections below and in Paper I.

ÄKTA systems communicate and are controlled via the proprietary software UNICORN which means that Orbit must communicate with UNICORN. This is achieved via two protocols: Open platform communications (OPC) and more recently an application programming interface (API). OPC is a standard for communication and data exchange in industry (OPC Foundation, n.d.). Since the industry is making strides for a wider use of OPC (Chen et al., 2020), this opens the potential for Orbit to be expanded and be used with more and more hardware in the future.

Agilent systems are interfaced using an API called Instrument Control Framework (ICF) from Agilent and the implementation is further described in the sections below and in Paper IV.

Orbit can also use serial and Ethernet to communicate with a multitude of auxiliary equipment, such as laboratory scales, stirrers, pumps, and even Raspberry Pi microcomputers. Given the user-friendliness of Python and the modular design of Orbit, it's easy to further integrate hardware.

Sequential control

The core functionality of the Orbit controller is the ability to write scripts to control processes. These scripts are written in Python and are composed of multiple Orbit *Phase* objects. That is, user-defined phases such as loading, gradient, or elution. The phases are put in sequence and executed one after another. Usually, the phases have a fixed time in which they are performed, but it is also possible to exit a phase and continue to the next one on another condition than time, e.g., if a UV signal gets below a threshold. A visualisation of a simple Orbit script is shown in Figure 3.2 below. The illustration also shows two branches of phases running in parallel.

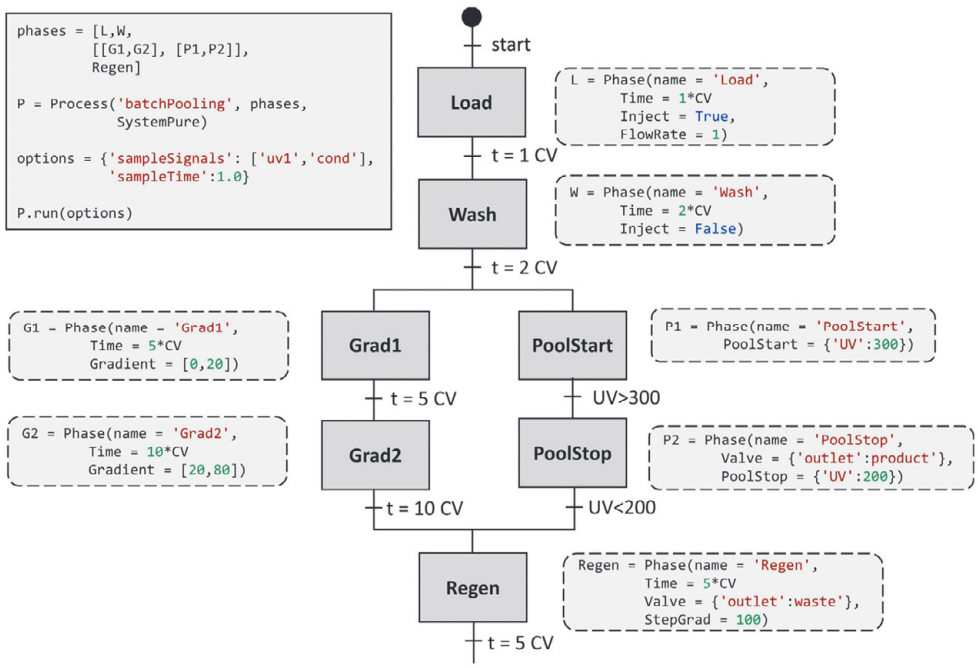

Figure 3.2
Example of a simple Orbit script with pooling and multiple gradient phases run in parallel (Andersson et al., 2022).

Communication

Multiple instances of Orbit running on different computers and controlling different systems can communicate over a local area network (LAN) making it possible for processes to be interconnected in various ways. This ability was heavily used in Paper III where three different chromatographic systems and processes were integrated and communicated over a LAN to achieve synchronization of sampling and analysis. In this case the different Orbit instances used shared Boolean variables, and all systems could read and write to.

Database

Orbit has a database which is implemented using MongoDB which is a non-relational database, or NoSQL. The advantage of using a NoSQL database is that it is scalable which is important when collecting large amounts of data such as absorption spectra (Chen et al., 2020). The Orbit data base can be used to collect all available data from process runs and the data can then be accessed for analysis, visualization, or model calibration.

Simulator

Orbit is equipped with a simulator which can be run in tandem with any process run. The simulator makes use of a user-defined system file which contains information on what units are in the system and how they are interconnected. Orbit contains mathematical representations of all the units in the system, as well as tube connections and valve settings and therefore all flow paths are known at any time point during a process. Each unit and every tube are included in the model and the virtual representation is automatically created from the user-defined system file. The simulator is discussed in further detail in the sections below.

4. Automated advanced quality analysis

With processes becoming more and more automated, the need arises for ways of testing for product quality and other key process parameters in an automated way. This chapter introduces methods for the automatic sampling and analysis of a downstream processes by using a separate chromatographic system for sampling and sample preparation as well as an analytical chromatography system for analysis. These systems are controlled by and communicates with each other via the supervisory software Orbit.

4.1. Integration of Agilent hardware into Orbit

There are several software applications available with the ability to control Agilent systems. These programs are sometimes called Chromatography Data Systems (CDS). Some examples are *OpenLab CDS*, *Empower*, and *Chromeleon*. However, to our knowledge these programs lack the capability to stream sensor data to third party programs. (Andersson et al., 2018) This is a requirement if the Agilent system is to be used in tandem with other systems or for real-time applications using Orbit. That is why we developed *Satellite*, an interface for Agilent systems in Orbit.

Satellite CDS

Satellite is implemented in *C#* using a framework from Agilent, called Agilent Instrument Control Framework (ICF), that allows for control of Agilent systems as well as systems from other vendors provided that the hardware uses the correct drivers. The ICF protocol and drivers were provided to us by Agilent.

Satellite works as a server that controls an Agilent system connected to the same computer as it is running on while communicating with an instance of Orbit over a network. The communication is done by sending commands to Satellite. Figure 4.1 below shows the Satellite framework.

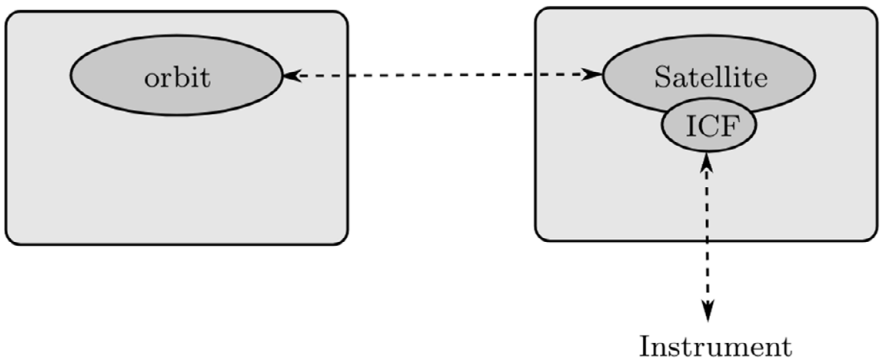

Figure 4.1
Illustration of how Satellite communicates. The Satellite instance uses Agilent ICF to communicate with and control the Agilent system connected to the same computer it is running on. An instance of Orbit on another computer sends commands to and receives data from Satellite via local area network.

Orbit implementation

In Orbit, the Agilent system, including all its units, are represented as an HPLC-object. This means that the system can be controlled in the same manner as an ÄKTA system in the Orbit paradigm, e.g., to set the system's flow rate you use the following notation: `HPLC.pump.setFlowrate()`. As of writing, the only Agilent system in our laboratory is an HPLC. However, given the modular implementation of systems in Orbit, it would be an easy task to implement any other Agilent system in the future.

4.2. Sampling techniques

In this thesis, two distinctive methods to sample a continuous downstream process were implemented. In the first method samples were taken from a process and handled by a separate system designed for sample preparation. When taking samples of column elution, the whole elution pool was sampled. This was used in Paper III and is called *pool sampling* below. In the second method the process flow was split into two streams where one smaller stream was diverted from the main process stream and run through detectors on an analysis system. This was done in Paper IV and is called *flow splitting* below. I will also briefly discuss a third sampling option which I call *fractionation*.

Pool sampling

In Paper III we designed a way to automatically sample and analyse a preparative downstream process. This system consisted of a separate chromatography unit that handled sampling and sample preparation, as well as an analytical HPLC system that handled sample analysis, see Figure 4.2 below.

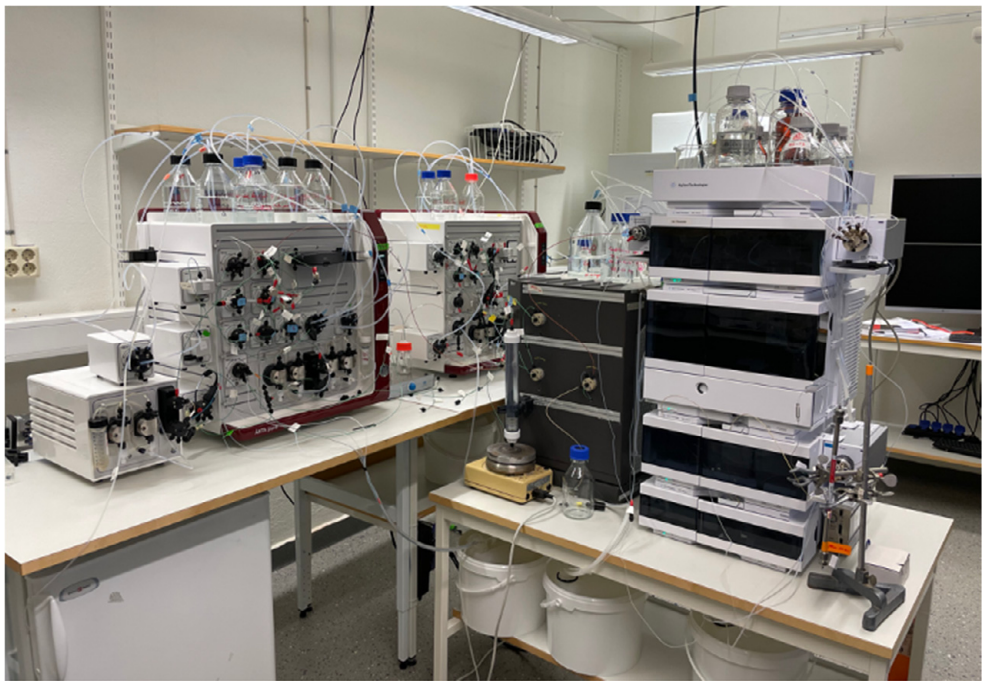

Figure 4.2
Picture of the laboratory setup from Paper III. Left to right: downstream process (PCC) on an ÄKTA Pure, preparative system implemented on an ÄKTA Explorer, and an Agilent 1260 HPLC system.

Pool sampling was done by pumping entire elution pools from a continuous process into a superloop on a separate sample preparation system. This was done in order to have a representative sample of the whole pool. The superloop was modified by the addition of a small magnetic stir bar which enabled the sample in the loop to be stirred and thusly homogenized. The sample preparation system was designed so that the sample in the loop could be diluted and conditioned, i.e., the pH could be adjusted. The sample was then stored in the loop until it was sent to analysis on an HPLC. In this setup the process being sampled transfers the sample to the sample preparation system which means that the sample does not pass through any pump. This is a positive aspect of this method as it is difficult to assure sample purity when passing it through a system pump. Of course, this issue could be sidestepped by using a pump where the flow path is uncompromised, such as a peristaltic pump or

a syringe pump. One could also sample a process by pumping the sample from a bottle or a surge tank, for example from a virus bottle. This is advantageous because the product is ideally already homogenized in the bottle, and only the amount needed for analysis have to be retrieved from the virus inactivation bottle.

One of the biggest problems of the superloop method is that since the whole pool is removed from the process, the process cannot continue with that pool; it has to start over. Alternatively, the sample could be returned once the amount required for analysis has been taken, but this is only practical if the sample has not been conditioned in a way that influences the downstream process.

It is also vital that not too much of the sample has been used for analysis and lost due to system dead volumes and pump washes. This is why the scale of the process plays a part. If the volumes in the process are too small, it may not be worth to wait for the sample to be returned. In this case it may be better to restart the previous process step and waste the remaining sample. In Paper III, the sample volumes were small and the collected sample was sent to the waste after analysis.

In Paper III we suggested multiple sampling points in the monoclonal antibody platform described by Scheffel et al. (2022) and Schwarz et al. (2022). These sampling points are shown in Figure 4.3 below. When increasing the number of sampling points and thus increasing the number of types of samples that need analysis in the system, the number of superloops should also be increased in the sample preparation system so that each sample point gets its own superloop.

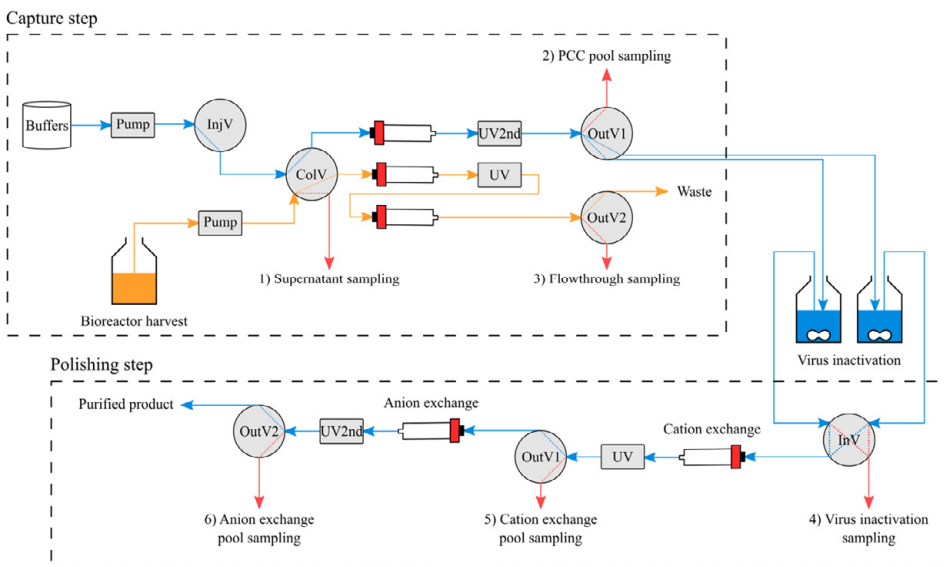

Figure 4.3
Schematic showing possible sampling locations in a monoclonal antibody platform.

Flow splitting

In the flow splitting method, we used a flow splitting valve to separate a small amount of the system flow and to run that separated stream through the detectors of an analytical HPLC system. This gave us the ability to capture the UV-VIS absorption spectra of the process continuously. The process diagram and a picture of the setup is shown in Figures 4.4 and 4.5 below. Figure 4.6 shows an absorption spectrum from Paper IV.

Of course, one does not need to split the flow; it is possible to run the process stream through the HPLC detectors directly. However, there were three major reasons for using a flow splitter. Firstly, the general goal of the solutions presented in this thesis is to be multipurpose and modular. This means that the analysis system should be able to run separately and in multiple ways if needed, maybe even during the same process. This was demonstrated in Paper IV by also being able to automatically load the sample loop of the HPLC from the process and running an analysis on an analytical column. This makes the analysis system flexible and enables a multitude of data acquisition possibilities. Secondly, in order to use the detectors from the HPLC in-line in the process stream one need to be wary of the flow rate and subsequently the pressure drop in the detectors. This means that the process flowrate now has to be constrained in order to protect the detectors. This is fully alleviated if the flow is split since the flow rate can be kept constant on the HPLC side. Thirdly, when splitting the flow, one could run destructive analyses such as mass spectrometry in the end of the chain of detectors.

One downside with the particular flow splitter used in this work is that it may mix the two flows, i.e., a small part of the analytical system's flow may be mixed into the process stream. This may not pose such a large problem depending on the process, but it is probably not considered a good manufacturing practice.

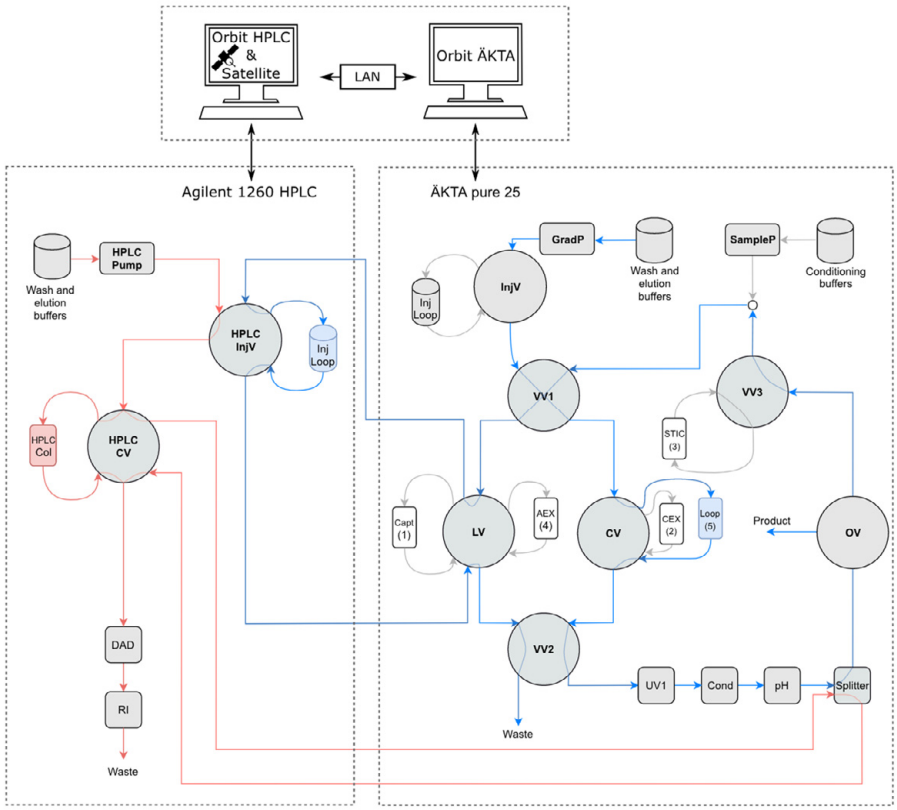

Figure 4.4
Process diagram of the setup from Paper IV where the flow was split and diverted through the detectors of an HPLC.

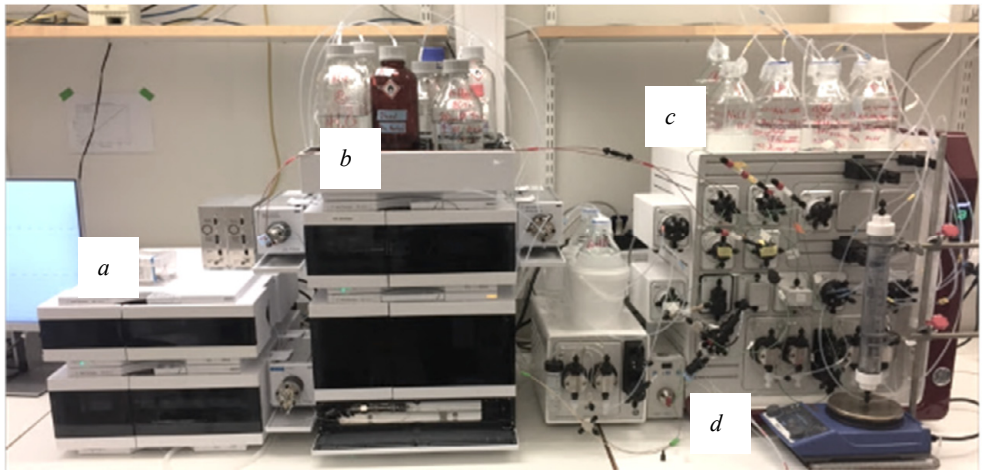

Figure 4.5
Picture of laboratory setup used in Paper IV. a) HPLC detectors, b) HPLC pump, autosampler, and column compartment, c) ÄKTA Pure with downstream process, and d) Splitter valve.

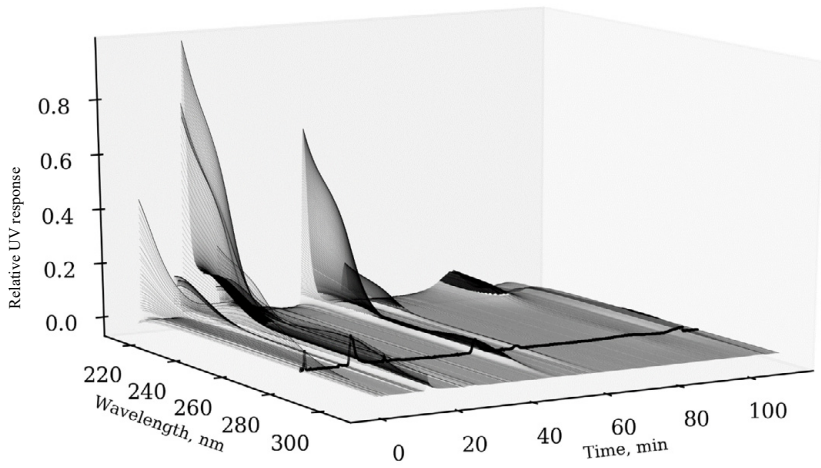

Figure 4.6
Continuous absorption spectrum from downstream process presented in Paper IV. The darker line depicts a standard chromatogram at 280 nm.

Fractionation

Another way of sampling that is different from pool sampling is fractionation. The principle is to take fractions of a pool instead of sampling the whole pool. In this way one could get insight into a pool's composition over time which could be considered an instant sample versus an integral sample in the case with a homogenic

pool. Used in this way this method is sometimes called 2D chromatography. This fractionation could be done in an automatic way by sampling the pool multiple times and storing the samples in ordinary capillary loops, as shown by Williams et al. (2017). If this were to be implemented in a similar setup as our sample preparation system one would have to be meticulous with how the samples were transferred through the system, i.e., all dead volumes must be known and Taylor dispersion in the tubes must be taken into consideration et cetera. This is primary as to not waste any sample and the reasoning is the same as with the pool size above; if the pool is too small, too much of the pool is taken out of the process. One solution to achieve better accuracy when doing this type of sampling is to use a digital twin of the preparative system to predict or track where in the tubes a particular sample is located at any given time as well as keeping track of dispersion.

4.3. Quality analysis system control strategy

To be able to control these complex processes together with the automatic quality analysis system, we utilized Orbit's communication capabilities: Each system had its own instance of Orbit running on the computer to which it was connected. That instance was responsible for controlling the system and running any processing scripts. The instances of Orbit communicated a LAN and used a shared variable space. This means that each Orbit can read and write to a set of variables at any time over the network. These Boolean variables are called flags since they can be lowered and raised in analogy with signal flags. This could be considered a simple implementation of what is called a semaphore in real-time systems engineering. Furthermore, each Orbit instance was set up to run as a sort of finite-state machine, which means that each system had to be in a well-defined state at any given time and for the system to transition into the next possible state, certain conditions had to be met. For example, if the sample preparation system wanted to send a sample to the analytical system, the flag signalling if the HPLC was ready needed to be set to *True*. An illustration of this can be seen in Figure 4.7 below.

The samples taken from the process are stored in superloops on the preparative system which means that they do not have to be analysed the moment they are collected. This creates some flexibility in the system since newly taken samples that may be considered to have a higher priority could be sent to the analysis system ahead of samples that have already been taken and are stored in the superloops. A simple way of determining priority would for example be to assign a priority number to all samples that are ordered and then selecting samples with the highest priority for analysis. To handle the possibility of some low-priority samples never being analysed, one could automatically increase the number over time.

In the implementation of this communication scheme in Paper III, it is the client, i.e., the downstream process, that decides whenever a sample is to be taken. This design choice could almost be considered arbitrary as it depends a lot on how the overall process is set up. For example, one could argue that the quality analysis system should decide when sampling is performed and order the client to prepare for sampling. It is also possible to imagine a setup where an additional instance of Orbit is acting as an external master that oversees the entire process. This may be advantageous if the downstream process consists of more than one system, e.g., there are polishing steps after the capture step in a monoclonal antibody process.

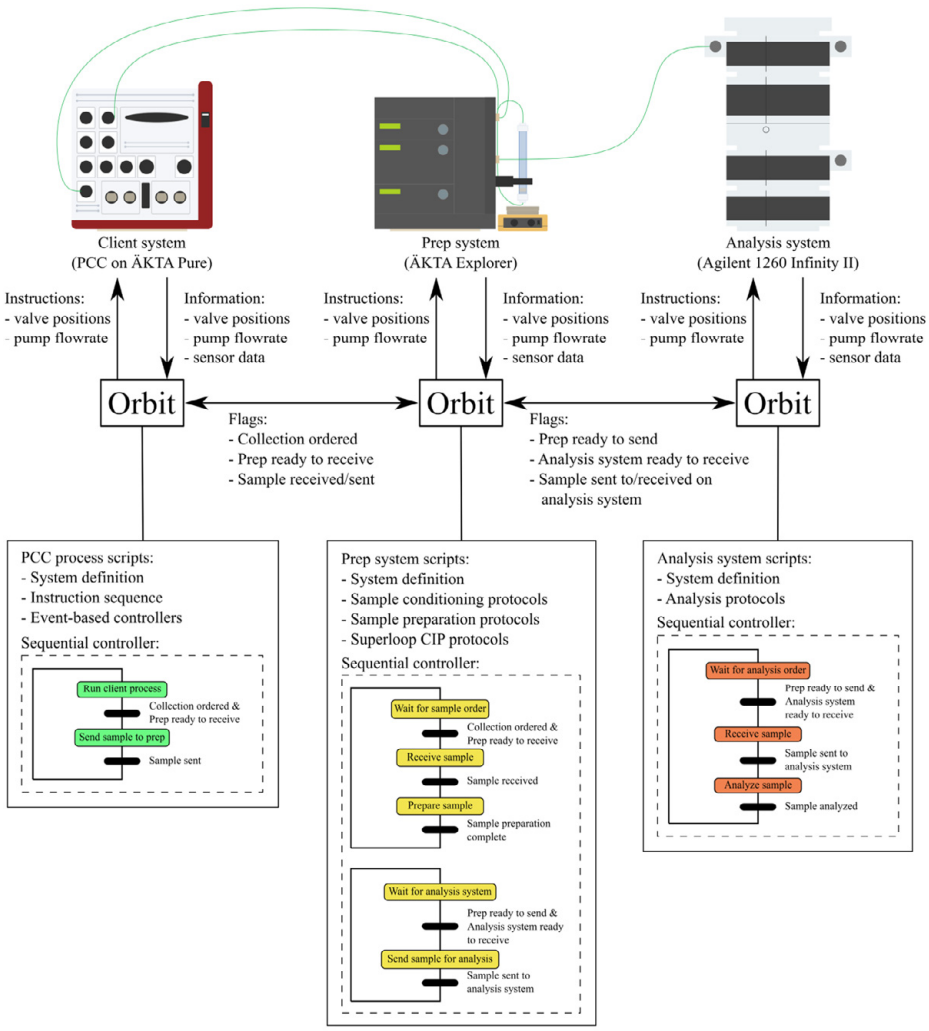

Figure 4.7
Schematic depiction of the quality analysis system's control scheme and architecture from Paper III.

4.4. Conclusions

The methodology presented above enables a general and flexible system for automatic quality control in a continuous downstream platform. By using a sample preparation system, sampling can be performed in several ways to facilitate a multitude of analyses.

The ability to split the process flow means that detectors on an analytical chromatography system can be used alongside a process for advanced data acquisition, without compromising the flexibility of the analysis system. Furthermore, destructive detectors could be used in a continuous process without losing too much product.

The use of a supervisory control software such as Orbit allows for the implementation of advanced processes and for automatic quality analysis of these processes. The modularity and flexibility of Orbit makes it well-suited for applications where multiple systems are involved and they need to communicate.

Automatic sampling done in this way enables feedback to the modelled system in a manner that enables the implementation of full-scale digital twins.

5. Automatic generation of process models

The Orbit simulator is capable of simulating systems that run complete downstream processes. To be able to do this, the Orbit simulator needs adequate mathematical models that properly describe the chemical and physical processes taking place in the system. These models need to be calibrated in order to produce usable results. In the sections below I will try to describe the different aspects needed for the automatic generation and calibration of such models with the Orbit simulator.

5.1. Mathematical Modeling

The goal of modelling is to gain insight and understanding of an object, system, or phenomena. One wishes to capture a certain behaviour. A designer or architect might for example build a scale model to study the interactions of light and objects in space while a psychologist may use behavioural models to predict consumer patterns in marketing. Engineers use mathematical models to design and study physical systems and processes.

To model a downstream process, we need models that describe all the units in the process's flow path. In a preparative chromatography system these units are mixers, valves, detectors, tubes, and separation columns. In Papers I and II we use the modelling scheme described in the following sections.

Mass balances and isotherms

The mixers, valves, and detectors can in comparison to the thin capillary tubes of the chromatographic system be considered as having a relatively large dead volume. We can assume that mixing occurs in these units, and they can therefore be modelled as continuous stirred-tank reactors.

The rate of change in concentration of component c in the units can be modelled as

$$\frac{dc}{dt} = \frac{F}{V}(c_{in} - c) \tag{5.1}$$

where c is the concentration of the component, F the volumetric flow rate through the unit, V the volume of the unit, and c_{in} the inlet concentration.

Tubes can be modelled using the convective-dispersive model. That is, we assume that the propagation of a component in the tube is due to convective transport i.e., the component is transported by the flow through the tube, and that the only mixing that occurs is in the axial dimension and is due to dispersion. Equation 5.2 shows the convective-dispersive model.

$$\frac{\partial c}{\partial t} = D_{ax}\frac{\partial^2 c}{\partial z^2} - \frac{F}{A}\frac{\partial c}{\partial z} \tag{5.2}$$

where D_{ax} is the axial dispersion coefficient, z the length dimension, and A the cross-sectional area of the tube.

Chromatographic columns are modelled in a similar way to tubes. In order to capture the adsorption behaviour of components, a reaction term is added to the convective-dispersion model, see Equation 5.3. Since the column is assumed to be a porous packed bed, the column void and porosity must be taken into account. Equation 5.4 shows the relationship between the different porosities used.

$$\frac{\partial c}{\partial t} = D_a\frac{\partial^2 c}{\partial z^2} - \frac{u}{\varepsilon}\frac{\partial c}{\partial z} - \frac{(1-\varepsilon_c)}{\varepsilon}\frac{\partial q}{\partial t} \tag{5.3}$$

$$\varepsilon = \varepsilon_c + (1-\varepsilon_c)\varepsilon_p \tag{5.4}$$

where c is the mobile phase concentration, q the stationary phase concentration, D_a the apparent dispersion coefficient, u the superficial mobile phase velocity (defined as $u = F/A$), ε_c the column void, ε_p the particle void, and ε the total void of the column.

If we assume that the relationship between the mobile phase and the stationary phase concentrations are governed by a kinetic equation, this model can be called a reaction-dispersive model (Golshan-Shirazi & Guiochon, 1991, 1992). To reduce the complexity of the model and in this way make implementation easier, reduce computational cost, and simplify model calibration, all mass transfer phenomena and adsorption kinetics are lumped into one kinetic process.

A well-established and widely used adsorption isotherm for the separation of proteins using ion exchange chromatography is the steric mass-action (SMA) isotherm. It takes into consideration the salt concentration in the mobile phase and

the steric hindrance of stationary phase binding sites by the proteins (Brooks & Cramer, 1992; Jakobsson et al., 2007; Karlsson et al., 2004; Nilsson & Andersson, 2017; von Lieres & Andersson, 2010). Equation 5.4 below shows the SMA isotherm on kinetic form.

$$\frac{\partial q_i}{\partial t} = k_{kin,i}\left[K_{eq,i}c_i\left(\Lambda - \sum_{j=1}^{\mathrm{n}}(\nu_j + \sigma_j)q_j\right)^{\nu_i} - c_s^{\nu_i}q_i\right] \tag{5.4}$$

where q_i is the concentration of adsorbed protein i on the stationary phase, c_s, the mobile phase salt concentration, Λ the ligand density, n the number of components modelled, ν_i the characteristic charge of the protein, σ the shielding factor of the protein, K_{eq} the equilibrium constant, and k_{kin} the adsorption kinetic constant.

For the sake of simplicity, we can combine some parameters in Equation 5.4 and reparametrize it as

$$\frac{\partial q_i}{\partial t} = k_{kin,i}\left[H_{0,i}c_i\left(1 - \sum_{j=1}^{\mathrm{n}}\frac{q_j}{q_{max,j}}\right)^{\nu_i} - c_s^{\nu_i}q_i\right] \tag{5.5}$$

where

$$H_{0,i} = K_{eq,i}\Lambda^{\nu_i} \tag{5.6}$$

$$q_{max,i} = \frac{\Lambda}{\nu_i + \sigma_i} \tag{5.7}$$

Since the adsorption isotherm is salt dependent, the salt concentration needs to be modelled as well. Assuming monovalent counter-ions, the rate of change of adsorbed salt, q_s, can be modelled as

$$\frac{\partial q_s}{\partial t} = -\sum_{j=1}^{\mathrm{n}}\frac{\partial q_j}{\partial t}\nu_j \tag{5.8}$$

However, in order to decrease model complexity, in this work it is assumed that the salt does not adsorb to the stationary phase and therefore the mass balance equation of the salt can be written as

$$\frac{\partial c_s}{\partial t} = D_a\frac{\partial^2 c_s}{\partial z^2} - \frac{u}{\varepsilon}\frac{\partial c_s}{\partial z} \tag{5.9}$$

Discretization

In order to compute the propagation, or change in concentration, of target components through the chromatographic system over time, the partial differential equations (PDEs) in the previous section are discretized. The equations are discretized in the space dimension according to the method of lines paradigm using the finite volume method (FVM) (Eymard et al., 2000; Nilsson & Andersson, 2017). This method divides the modelled units into several finite volumes, or cells, which are balanced to their neighbouring cells. This makes the finite volume method quite robust although special care needs to be taken to handle the eventuality of propagating discontinuities stemming from saturation of the adsorption isotherms.

After discretization, the PDEs are reduced to a system of ordinary differential equations (ODEs) which are solved using an initial problem solver for ODEs in *Python* and using the backward differentiation formula (BDF) method (*Scipy.Integrate.Solve_ivp — SciPy v1.11.1 Manual*, n.d.).

The column used in this work has a volume of 1 ml and is discretized using 50 grid points which means that every finite volume in the column model is 20 µl. This is user-configurable and can be changed in the simulator on a case-by-case basis. The tubes are discretized using a sparse grid of one grid point for every five centimetres. This reduces the number of equations needed for solving the model compared to a denser grid. However, using a coarse grid like this could give rise to numerical dispersion (Wu & Forsyth, 2008).

5.2. Automatic process model

In the Orbit simulator, the units of the chromatographic system are automatically modelled using the user-defined system file. In this file the user states which units are present in the system and how they are connected to each other. Orbit then uses its library of models, comprised of the equations described in the section above, to create a system matrix containing all the discretized equations for all the states to be simulated.

If we look at the example system shown in Figure 5.1 below, it has six valves, one mixer, three detectors, and two superloops at one equation each; one column in which we have set the number of grid points to 50, i.e., 100 equations since both mobile phase and stationary phase concentrations are calculated; and let’s say two metres of tubing, with a sparse grid of one grid point per five centimetres. This gives us a total of approximately 150 equations. This number is multiplied with the number of components that is simulated so in the case of salt-dependent ion exchange separation of three proteins, i.e., four components including the salt ions, we need a total of 600 equations in the system matrix.

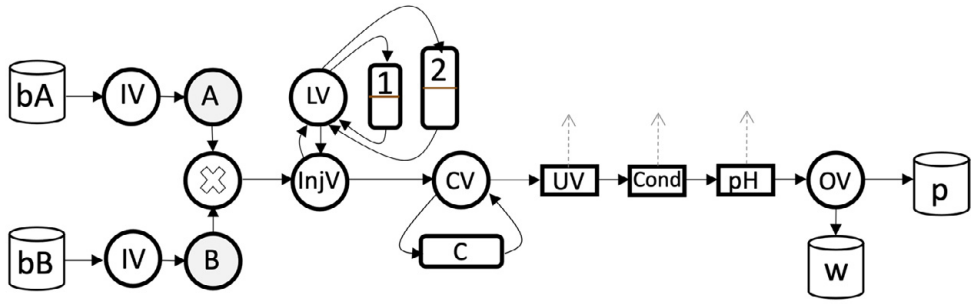

Figure 5.1
Illustration of an experimental setup configuration for system calibration. The system consists of buffer bottles connected to inlet valves (IVs) via pumps (A and B); an injection valve (InjV); a loop valve (LV) with superloops; a column valve (CV) with column (C); UV, conductivity, and pH sensors; and outlet valve (OV).

The relationships between the different units, i.e., which unit feeds into which unit etc., are governed by a structure matrix and this matrix is updated every time the flow path changes. The modelled units are connected and disconnected to each other depending on the flow path of the system. In the event of the flow rate or flow path changes, the model is updated to mirror this change and the simulations continues until the next such event. The flow rate in the system is governed by the system's pumps; each pump has a flow rate that depends on the overall system flow rate and the percentage of buffer B. The different flow rates from the pumps are subsequently added in the mixer.

The ability to track the system's flow path is a powerful tool for process development and it was the first intended use of the Orbit simulator.

All the simulated states in all units are known at any time which makes it possible to do advanced visualization and system monitoring.

It should also be noted that depending on what the simulator is used for you could choose from which unit to take the simulated results. When calibrating, it makes sense to use the simulated data from the UV sensor since it's there the data is recorded. When using the simulator for process optimization it would make more sense to use the result from the outlet valve or from the outlet tube since it's from there you would collect the product.

5.3. Integrated calibration procedures for ion exchange chromatography

To achieve well-fitting models in the Orbit simulator the models need to be calibrated; that is, to find what numerical value each parameter in the models has for the simulator to have a satisfactory response. This is done in a process called *model calibration*.

Calibration procedure

To determine all the model parameter values, a sequence of experiments with corresponding parameter estimations must be performed (Hahn et al., 2016; Huuk et al., 2014; Max-Hansen et al., 2015; Müller-Späth et al., 2011; Ojala et al., 2012; Osberghaus et al., 2012; Raje & Pinto, 1997; Saleh et al., 2020).

In this context, a calibration is roughly performed by running an experiment on a physical system, recording the data, and then running the exact same experiment in the simulator and adjusting parameters until the difference between the experimental and the simulated data is minimized. This means that the design of the experiments is crucial. It is of great importance that the experiment excites the correct parameter that is calibrated i.e., that when a parameter is changed the simulation result also changes in a way that it can approach the experimental data. This is why, to get a more comprehensible calibration procedure, one or a couple of model parameters are calibrated individually in sequence.

The calibration procedure is performed in several steps with each calibration consisting of two stages: First, the experiment is automatically performed on the system and the data is recorded. Second, the associated parameters are estimated using a suitable calibration method. The calibration method runs the simulation until sufficient model fit is achieved. An illustration of how the calibration is done is shown in Figure 5.2. These calibration cycles are performed in sequence, with each step refining the model.

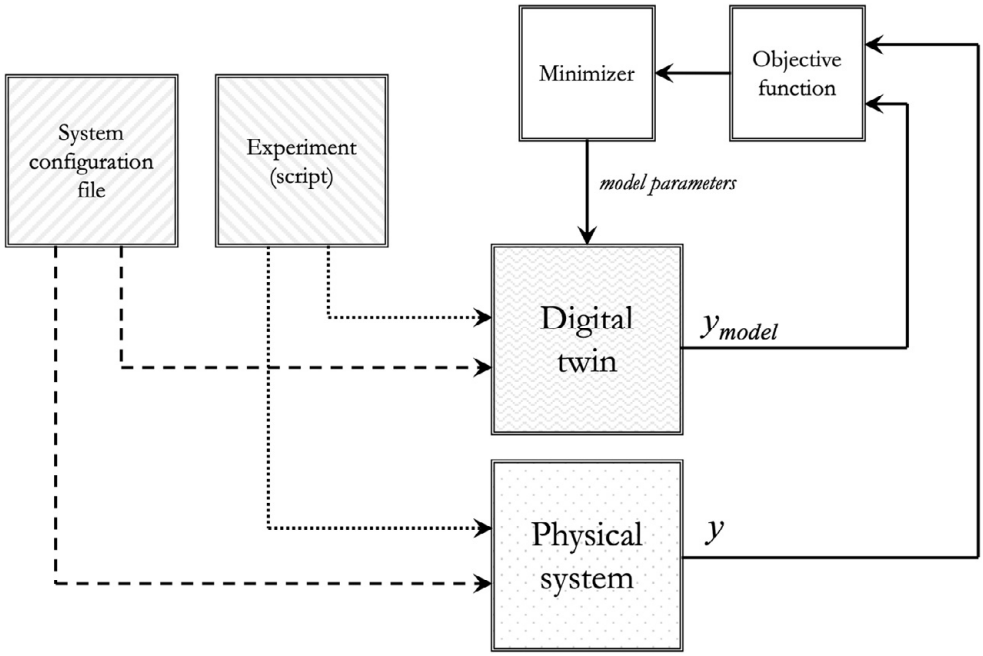

Figure 5.2
Schematic overview of the calibration procedure. The digital twin and the physical system both uses the configuration and script files in Orbit. After the experiment is run, the minimizer algorithm runs the digital twin simulation while adjusting model parameters until a satisfactory fit is achieved.

Objective function

To determine how well the simulations fit the experiment, an objective function, $Q(x)$, is created. An objective function is defined in such a way that the function value becomes smaller the more the simulation fits the experimental data. This can be achieved in several ways and the specific definition depends on what is considered a good fit.

In this work, we utilize two distinct methods to assess model fit concerning a model parameter, x: the difference in residence time between simulation and experiment,

$$Q(x) = \left(t_{residence,exp} - t_{residence,sim}(x)\right)^2 \tag{5.10}$$

and the sum of squared errors between a number of data points, i.e., the difference between a simulated and real chromatogram, or sum of squared errors (SSE),

$$Q(x) = \sum_{k=1}^{n} \left(y_{exp,k} - y_{sim,k}(x)\right)^2 \tag{5.11}$$

where n is the number of data points and y is typically the measured and simulated UV absorption signal.

The first case is used when the calibrated parameters mostly affect residence time, e.g., the length of tubes in the system or the volume of certain units. The second is used when the parameters influence the shape of the eluted peaks. Some parameters have an influence on both residence time and peak shape. In these instances, it can be useful to use a weighted sum of both the object functions above.

Finally, to minimize the objective function a minimizer is used. In this work we used the Python function *minimize* from the SciPy library (*Scipy.Optimize.Minimize — SciPy v1.11.2 Manual*, n.d.). The methods used in the minimization were derivative-free methods, such as the Nelder-Mead method (Gao & Han, 2012; Nelder & Mead, 1965). Direct minimization methods may not be the fastest methods, but they are robust and that is of great importance in an automated setting. That and the possibility to use constraints is why they were chosen.

The Yamamoto method

In Papers I and II we used a method called the Yamamoto method to estimate the model parameters ν and H_0 in Equation 5.4 above (Ishihara et al., 2005; Rüdt et al., 2015; Saleh et al., 2020) from three linear gradient experiments. An example of three gradient experiments and the calibrated simulator's response for three components is shown in Figure 5.3.

The method uses the salt concentration at the peak maximum of an eluted protein, $c_{s,i}$, initial and final salt concentrations of the gradient, $c_{G,initial}$ and $c_{G,final}$, and gradient volume, V_G, as well as column interstitial volume, V_{col}, and column void, ε_c

$$\log(GH) = (\nu_i + 1) \cdot \log(c_s) - \log\left(H_{0,i} \cdot (\nu_i + 1)\right) \tag{5.12}$$

$$g = \frac{c_{G,final} - c_{G,initial}}{V_G} \tag{5.13}$$

$$GH = g(V_{col} - \varepsilon_c V_{col}) \tag{5.14}$$

The parameters are found by means of linear regression and the expressions above.

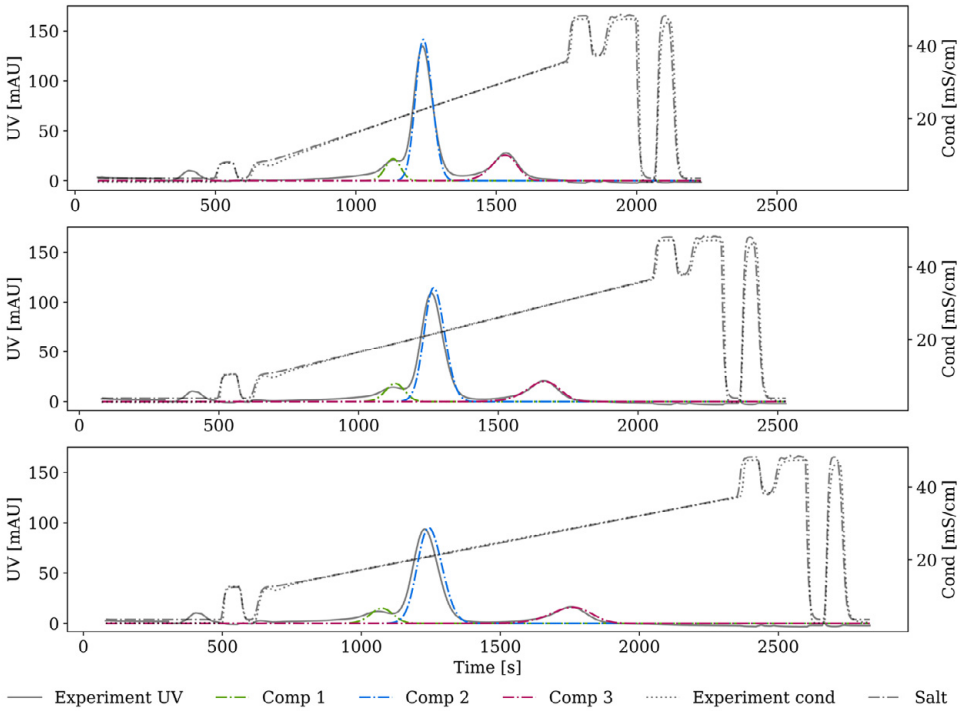

Figure 5.3
Three linear gradient experiments with results from the calibrated model.

Example calibration procedure

In Papers I and II, a chromatography system with an ion exchange column was calibrated using the following calibration procedure:

1) System and column calibrations

   a) System mixing and dead volume.

      Using a pulse experiment where some of the sample was injected into the system with the column in bypass, we calibrated the length of one of the system's tubes so that the residence times of the pulse was equal in the experiment and the simulation.

   b) Column void and porosity.

      By injecting the column with a mixture containing one tracer that did not go into the pores of the column, and one that entered all the column's void we estimated the void and porosity of the column.

2) Adsorption calibration, find salt dependency.

   Three linear gradient experiments were used to estimate the adsorption parameters of the model. The three gradients were each made successively less and less steep for each run by increasing the gradient length.

3) Find column capacity.

   One overloading experiment was used to calibrate the column's binding capacity.

Single versus multi component calibration

The biggest difference between Paper I and Paper II in terms of calibration, is that in Paper I we only calibrated one component with a known concentration, while in Paper II we calibrated three components with unknown relative concentrations. The increase in complexity between the two cases gave rise to an interesting insight: When choosing the linear gradient experiments, care had to be taken to ensure that all components eluted during the gradient phase. It was also a bit delicate how to change the gradient while ensuring that the Yamamoto method worked as intended. When calibrating only one component, the gradients could be chosen in a way as to spread the points in the logarithmic plot of the Yamamoto method; one could be a bit aggressive with the slope. When calibrating three components it was more difficult to find gradients that both ensured somewhat separated peaks that eluted during the gradient and that gave enough of an impact in the Yamamoto method. It was by no means impossible, but the approach had to be different.

In Paper I, we only used the Yamamoto method to estimate the adsorption parameters, but in Paper II we used minimization to find them. However, we used an estimation from the Yamamoto method as starting values in the calibration. This was done since the adsorption parameters are very sensitive and the simulation result must be somewhat close to the experiments. If one would choose starting values of these parameters arbitrarily, the simulation would probably return a chromatogram where the peaks were so far from the experiments that they would not overlap which would cause the minimization to not converge. This is because even when the parameters somewhat change, the value of objective function does not change. This may not be a problem if a global optimizer such as a stochastic optimizer is used. This was however not explored further in the work presented in this thesis.

5.4. Discussion

Model selection

In the work presented in this thesis, we only modelled ion exchange chromatography using the steric mass action (SMA) model. This is mainly because this model was capable of capturing the adsorption behaviour in the separations we ran and because of previous experience. However, there are no reasons why you could not implement other kinds of chromatography models or other models for ion exchange in the Orbit simulator. In fact, there has been work done to add models for hydrophobic interaction (HIC), affinity chromatography, and size exclusion chromatography (SEC) among others (Malmström, 2022). One could also imagine other types of column models, such as film or general rate models.

Another kind of model that could be implemented is the hybrid model. A hybrid model is a model which consists of a combination of a mechanistic model and a statistic model, typically a neural network (Krippl et al., 2020; Nagrath et al., 2004; Narayanan et al., 2021; Walch et al., 2019).

One thing to keep in mind when using different models is that the experiments for calibration need to be chosen in a way as to excite the model parameters in a controlled way. This means that for every type of model, a selection of model-specific experiments must be developed. In the case of hybrid models, one would expect a higher number of parameters for the neural network which would generally mean that more experiments need to be performed.

In the case of models that incorporate some kind of dependency on flow rate, e.g., mass transfer models such as general rate, experiments with different flow rates need to be performed as well. This is of course also true for models with pH dependencies.

Human interaction

In order for the simulator to properly work, the Orbit controller needs some user input, or *human interaction*, as it were: Firstly, the user needs to create a system configuration file which describes the units that the system consists of. If the system is only being controlled or monitored, the user only has to define which units are present in the system. However, if the Orbit simulator is used, the tubes that connect the units and their approximate lengths must be defined as well. Secondly, the user needs to make sure that the separation problem to be calibrated is possible. This may be self-evident but what this means is that a user could not take an arbitrary sample and expect Orbit to be able to automatically calibrate a model for its separation. Some knowledge of the separation problem is necessary. This can be seen in the

gradient elution steps of the calibrations described above, where the user must guarantee that the target protein elutes during the gradient.

It is worth noting that this is only *calibration*; the experimental behaviour must follow model theory. We only calibrate the parameters in the models. We do not develop new models or model structures in these procedures. Also worth noting is that this is model calibration and not parameter estimation. The actual values of the estimated parameters are inconsequential, and they may even vary between systems.

Another requirement of user input may be in the determination of what constitutes a good enough fit to determine if the calibration is successful. One solution could be to let the user define to what level they want the simulator to fit the experiments, using some measurement of goodness of fit. Orbit could then decide whether a calibration was successful or not and act accordingly. Maybe either by terminating the calibration, trying again with new starting parameter values, or by changing the model structure used.

In the case of multiple component calibration as in Paper II, the user also must decide on how many components are present and need to be calibrated. Each component should also have sufficient separation in the user-provided experiment to allow for automatic detection using peak detection algorithms.

Lastly, the column capacity needs to be approximately known in order to do an overloading experiment. In this work we used a loading factor of about 25% of the column's maximum capacity.

5.5. Conclusions

We have developed a general methodology for the calibration of a downstream separation process using the Orbit controller. A given model structure yields a specific set of experiments. This presupposes that a model structure exists and that the experiments chosen sufficiently sample the parameterized properties. A next step would be to find experiments automatically, perhaps by examining data from a high-throughput screening (HTS) robot if such data is available.

This methodology is a stepping stone in the development of digital twins for integrated downstream processes.

6. Conclusions

In this thesis I have presented several tools for use in the application of digital twins in downstream processes for the production and development of biopharmaceuticals.

I presented a system for general and flexible analysis in a downstream process by utilizing a sample preparation system in tandem with an analysis system. This setup enables numerous types of analyses and options for process control.

By using a flow splitter and diverting some of the process flow we facilitate the use of advanced flow-through analytical sensors for real-time data acquisition in downstream processes without a considerable loss of product. The implementation proposed in Paper IV keeps the analytical system operational, allowing for the possibility of conducting analyses on process products as well.

The use of a supervisory control software, exemplified by Orbit, allows for implementation of advanced downstream processes, for automatic quality analysis, and advanced control.

Orbit's modular and flexible design allows for applications with multiple systems from different manufacturers working together and communicating. Automatic sampling, as demonstrated in this work, provides feedback to the system models, enabling the creation of full-scale digital twins.

We have developed a general methodology for model calibration of a downstream separation process using the Orbit software. User-defined system configurations generate a model structure which yields a specific set of experiments. These experiments must properly sample the parameterized properties. To further enhance this approach, future work could explore automated experiment selection, possibly leveraging data from high-throughput screening (HTS) robots when available.

This methodology serves as a crucial step towards realizing digital twins for integrated downstream processes.

7. References

Andersson, N. (2018). *The Orbit Controller*. Department of Chemical Engineering, Lund University.

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

Andersson, N., Gomis Fons, J., & Nilsson, B. (2022). Chapter Five - Smart platform for development of small-scale integrated continuous downstream processes. In D. Moscatelli & M. Sponchioni (Eds.), *Advances in Chemical Engineering* (Vol. 59, pp. 131–158). Academic Press.

Andersson, N., Löfgren, A., Olofsson, M., Sellberg, A., Nilsson, B., & Tiainen, P. (2017). Design and control of integrated chromatography column sequences. *Biotechnology Progress*, *33*(4), 923–930.

Andersson, N., Tallvod, S., & Nilsson, B. (2018). *Satellite CDS*. Department of Chemical Engineering, Lund University.

Arnold, L., Lee, K., Rucker-Pezzini, J., & Lee, J. H. (2019). Implementation of Fully Integrated Continuous Antibody Processing: Effects on Productivity and COGm. *Biotechnology Journal*, *14*(2), 1800061.

Brestrich, N., Briskot, T., Osberghaus, A., & Hubbuch, J. (2014). A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. *Biotechnology and Bioengineering*, *111*(7), 1365–1373.

Brooks, C. A., & Cramer, S. M. (1992). Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *AIChE Journal*, *38*(12), 1969–1978.

Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., & Ierapetritou, M. (2020). Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review. *Processes*, *8*(9), Article 9.

Eymard, R., Gallouët, T., & Herbin, R. (2000). Finite volume methods. In *Handbook of Numerical Analysis* (Vol. 7, pp. 713–1018). Elsevier.

Gao, F., & Han, L. (2012). Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, *51*(1), 259–277.

Godawat, R., Brower, K., Jain, S., Konstantinov, K., Riske, F., & Warikoo, V. (2012). Periodic counter-current chromatography – design and operational considerations for integrated and continuous purification of proteins. *Biotechnology Journal*, *7*(12), 1496–1508.

Godawat, R., Konstantinov, K., Rohani, M., & Warikoo, V. (2015). End-to-end integrated fully continuous production of recombinant monoclonal antibodies. *Journal of Biotechnology*, *213*, 13–19.

Golshan-Shirazi, S., & Guiochon, G. (1991). Combined effects of finite axial dispersion and slow adsorption desorption kinetics on band profiles in nonlinear chromatography. *The Journal of Physical Chemistry*, *95*(16), 6390–6395.

Golshan-Shirazi, S., & Guiochon, G. (1992). Comparison of the various kinetic models of non-linear chromatography. *Journal of Chromatography A*, *603*(1), 1–11.

Gomis-Fons, J., Andersson, N., & Nilsson, B. (2020). Optimization study on periodic counter-current chromatography integrated in a monoclonal antibody downstream process. *Journal of Chromatography A*, *1621*, 461055.

Gomis-Fons, J., Schwarz, H., Zhang, L., Andersson, N., Nilsson, B., Castan, A., Solbrand, A., Stevenson, J., & Chotteau, V. (2020). Model-based design and control of a small-scale integrated continuous end-to-end mAb platform. *Biotechnology Progress*, *36*(4), e2995.

Grieves, M. (2015). *Digital Twin: Manufacturing Excellence through Virtual Factory Replication*.

Grossmann, C., Ströhlein, G., Morari, M., & Morbidelli, M. (2010). Optimizing model predictive control of the chromatographic multi-column solvent gradient purification (MCSGP) process. *Journal of Process Control*, *20*(5), 618–629.

Hahn, T., Baumann, P., Huuk, T., Heuveline, V., & Hubbuch, J. (2016). UV absorption-based inverse modeling of protein chromatography. *Engineering in Life Sciences*, *16*(2), 99–106.

Helgers, H., Hengelbrock, A., Schmidt, A., Vetter, F. L., Juckers, A., & Strube, J. (2022). Digital Twins for scFv Production in Escherichia coli. *Processes*, *10*(5), Article 5.

Herwig, C., Pörtner, R., & Möller, J. (Eds.). (2021a). *Digital Twins: Applications to the Design and Optimization of Bioprocesses* (Vol. 177). Springer International Publishing.

Herwig, C., Pörtner, R., & Möller, J. (Eds.). (2021b). *Digital Twins: Tools and Concepts for Smart Biomanufacturing* (Vol. 176). Springer International Publishing.

Hong, M. S., Severson, K. A., Jiang, M., Lu, A. E., Love, J. C., & Braatz, R. D. (2018). Challenges and opportunities in biopharmaceutical manufacturing control. *Computers & Chemical Engineering*, *110*, 106–114.

Huuk, T. C., Hahn, T., Osberghaus, A., & Hubbuch, J. (2014). Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography. *Separation and Purification Technology*, *136*, 207–222.

Isaksson, M., Gomis-Fons, J., Andersson, N., & Nilsson, B. (2023). An automated buffer management system for small-scale continuous downstream bioprocessing. *Journal of Chromatography A*, *1695*, 463942.

Ishihara, T., Kadoya, T., Yoshida, H., Tamada, T., & Yamamoto, S. (2005). Rational methods for predicting human monoclonal antibodies retention in protein A affinity chromatography and cation exchange chromatography: Structure-based chromatography design for monoclonal antibodies. *Journal of Chromatography A*, *1093*(1), 126–138.

Jakobsson, N., Degerman, M., Stenborg, E., & Nilsson, B. (2007). Model based robustness analysis of an ion-exchange chromatography step. *Journal of Chromatography A*, *1138*(1), 109–119.

Jungbauer, A. (2013). Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology*, *31*(8), 479–492.

Karlsson, D., Jakobsson, N., Brink, K.-J., Axelsson, A., & Nilsson, B. (2004). Methodologies for model calibration to assist the design of a preparative ion-exchange step for antibody purification. *Journal of Chromatography A*, *1033*(1), 71–82.

Konstantinov, K. B., & Cooney, C. L. (2015). White Paper on Continuous Bioprocessing May 20–21 2014 Continuous Manufacturing Symposium. *Journal of Pharmaceutical Sciences*, *104*(3), 813–820.

Krippl, M., Bofarull-Manzano, I., Duerkop, M., & Dürauer, A. (2020). Hybrid Modeling for Simultaneous Prediction of Flux, Rejection Factor and Concentration in Two-Component Crossflow Ultrafiltration. *Processes*, *8*(12), Article 12.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, *51*(11), 1016–1022.

Malmström, A. (2022). *Orbit-simulator for downstream processes*.

Max-Hansen, M., Knutson, H.-K., Jönsson, C., Degerman, M., & Nilsson, B. (2015). Modeling Preparative Chromatographic Separation of Heavy Rare Earth Elements and Optimization of Thulium Purification. *Advances in Materials Physics and Chemistry*, *5*(5), Article 5.

Moorkens, E., Meuwissen, N., Huys, I., Declerck, P., Vulto, A. G., & Simoens, S. (2017). The Market of Biopharmaceutical Medicines: A Snapshot of a Diverse Industrial Landscape. *Frontiers in Pharmacology*, *8*.

Müller-Späth, T., Ströhlein, G., Aumann, L., Kornmann, H., Valax, P., Delegrange, L., Charbaut, E., Baer, G., Lamproye, A., Jöhnck, M., Schulte, M., & Morbidelli, M. (2011). Model simulation and experimental verification of a cation-exchange IgG capture step in batch and continuous chromatography. *Journal of Chromatography A*, *1218*(31), 5195–5204.

Nagrath, D., Messac, A., Bequette, B. W., & Cramer, S. M. (2004). A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes. *Biotechnology Progress*, *20*(1), 162–178.

Narayanan, H., Seidler, T., Luna, M. F., Sokolov, M., Morbidelli, M., & Butté, A. (2021). Hybrid Models for the simulation and prediction of chromatographic processes for protein capture. *Journal of Chromatography A*, *1650*, 462248.

Narayanan, H., Sponchioni, M., & Morbidelli, M. (2022). Integration and digitalization in the manufacturing of therapeutic proteins. *Chemical Engineering Science*, *248*, 117159.

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313.

Nilsson, B., & Andersson, N. (2017). Simulation of Process Chromatography. In *Preparative Chromatography for Separation of Proteins* (pp. 81–110). John Wiley & Sons, Ltd.

Ojala, F., Max-Hansen, M., Kifle, D., Borg, N., & Nilsson, B. (2012). Modelling and optimisation of preparative chromatographic purification of europium. *Journal of Chromatography A*, *1220*, 21–25.

OPC Foundation. (n.d.). *What is OPC?* OPC Foundation. https://opcfoundation.org/about/what-is-opc/ (last visited Aug 20, 2023)

Osberghaus, A., Hepbildikler, S., Nath, S., Haindl, M., von Lieres, E., & Hubbuch, J. (2012). Determination of parameters for the steric mass action model—A comparison between two approaches. *Journal of Chromatography A*, *1233*, 54–65.

Patnaik, P. R. (1999). Applications of neural networks to recovery of biological products. *Biotechnology Advances*, *17*(6), 477–488.

Python Software Foundation. (2023, August 6). *About Python™*. Python.Org. https://www.python.org/about/ (last visited Aug 20, 2023)

Raje, P., & Pinto, N. G. (1997). Combination of the steric mass action and non-ideal surface solution models for overload protein ion-exchange chromatography. *Journal of Chromatography A*, *760*(1), 89–103.

Rathore, A. S., Mishra, S., Nikita, S., & Priyanka, P. (2021). Bioprocess Control: Current Progress and Future Perspectives. *Life*, *11*(6), Article 6.

Rathore, A. S., Nikita, S., Thakur, G., & Deore, N. (2021). Challenges in process control for continuous processing for production of monoclonal antibody products. *Current Opinion in Chemical Engineering*, *31*, 100671.

Rischawy, F., Briskot, T., Hopf, N., Saleh, D., Wang, G., Kluters, S., Studts, J., & Hubbuch, J. (2023). Connected mechanistic process modeling to predict a commercial biopharmaceutical downstream process. *Computers & Chemical Engineering*, *176*, 108292.

Rüdt, M., Gillet, F., Heege, S., Hitzler, J., Kalbfuss, B., & Guélat, B. (2015). Combined Yamamoto approach for simultaneous estimation of adsorption isotherm and kinetic parameters in ion-exchange chromatography. *Journal of Chromatography A*, *1413*, 68–76.

Saleh, D., Wang, G., Müller, B., Rischawy, F., Kluters, S., Studts, J., & Hubbuch, J. (2020). Straightforward method for calibration of mechanistic cation exchange chromatography models for industrial applications. *Biotechnology Progress*, *36*(4), e2984.

Scheffel, J., Isaksson, M., Gomis-Fons, J., Schwarz, H., Andersson, N., Noren, B., Solbrand, A., Chotteau, V., Hober, S., & Nilsson, B. (2022). Design of an integrated continuous downstream process for acid-sensitive monoclonal antibodies based on a calcium-dependent Protein A ligand. In *J Chromatogr A* (Vol. 1664, p. 462806).

Schwarz, H., Gomis-Fons, J., Isaksson, M., Scheffel, J., Andersson, N., Andersson, A., Castan, A., Solbrand, A., Hober, S., Nilsson, B., & Chotteau, V. (2022). Integrated continuous biomanufacturing on pilot scale for acid-sensitive monoclonal antibodies. *Biotechnology and Bioengineering*, *119*(8), 2152–2166.

scipy.integrate.solve_ivp — SciPy v1.11.1 Manual. (n.d.). https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html#r179348322575-5 (last visited Aug 6, 2023)

scipy.optimize.minimize — SciPy v1.11.2 Manual. (n.d.). https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html (last visited Sep 18, 2023)

Sellberg, A., Holmqvist, A., Magnusson, F., Andersson, C., & Nilsson, B. (2017). Discretized multi-level elution trajectory: A proof-of-concept demonstration. *Journal of Chromatography A*, *1481*, 73–81.

Tallvod, S., Andersson, N., & Nilsson, B. (2022). Automation of Modeling and Calibration of Integrated Preparative Protein Chromatography Systems. *Processes*, *10*(5), Article 5.

Tallvod, S., Espinoza, D., Gomis-Fons, J., Andersson, N., & Nilsson, B. (2023). Automated quality analysis in continuous downstream processes for small-scale applications. *Journal of Chromatography A*, *1702*, 464085.

Tallvod, S., Yamanee-Nolin, M., Gomis-Fons, J., Andersson, N., & Nilsson, B. (2022). A novel process design for automated quality analysis in an integrated biopharmaceutical platform. In L. Montastruc & S. Negny (Eds.), *Computer Aided Chemical Engineering* (Vol. 51, pp. 619–624). Elsevier.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual: (python documentation manual part 2)*. CreateSpace Independent Publishing Platform.

von Lieres, E., & Andersson, J. (2010). A fast and accurate solver for the general rate model of column liquid chromatography. *Computers & Chemical Engineering*, *34*(8), 1180–1191.

Walch, N., Scharl, T., Felföldi, E., Sauer, D. G., Melcher, M., Leisch, F., Dürauer, A., & Jungbauer, A. (2019). Prediction of the Quantity and Purity of an Antibody Capture Process in Real Time. *Biotechnology Journal*, *14*(7), 1800521.

Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, *40*(12), Article 12.

Williams, A., Read, E. K., Agarabi, C. D., Lute, S., & Brorson, K. A. (2017). Automated 2D-HPLC method for characterization of protein aggregation with in-line fraction collection device. *Journal of Chromatography B*, *1046*, 122–130.

Wu, Y.-S., & Forsyth, P. A. (2008). Efficient Schemes for Reducing Numerical Dispersion in Modeling Multiphase Transport through Heterogeneous Geological Media. *Vadose Zone Journal*, *7*(1), 340–349.