Helene Bøe

Real-Time Moving Horizon Estimation as a Soft Sensor for Sugar in a Bioprocess

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Nadav Bar
Co-supervisor: Andrea Tuveri
June 2023

Master's thesis

NTNU

Norwegian University of
Science and Technology

Helene Bøe

Real-Time Moving Horizon Estimation as a Soft Sensor for Sugar in a Bioprocess

NTNU
Norwegian University of
Science and Technology

Abstract

A hindrance in the development of process control and digital twins in the bioprocess industry is the scarcity of real-time direct measurements. Bacteria are used to produce a range of products within the pharmaceutical and food industry, however, they are often described by nonlinear time-varying dynamics, making them hard to model accurately. Moving Horizon Estimation (MHE) can address this issue by combining measurements with a process model, incorporating physical constraints to estimate states. The MHE can thereby act as a soft sensor allowing for real-time monitoring, and enabling control to improve product quality and yield.

This thesis investigates the implementation of real-time MHE to overcome the challenge of limited sugar measurements in continuous cultivation of *Corynebacterium glutamicum*. This consisted of experimental work related to running a bioreactor, and programming using optimization techniques. Online measurements of volume, cell density, and CO_2 were available, with periodic at-line sugar measurements every hour and offline cell dry weight (CDW) samples every 2-4 hours. Calibrations of pumps and an optical density probe were performed. *In silico* Model Predictive Control determined a fixed input profile for continuous cultivation. Parameter Estimation was performed to determine the parameters for the model, followed by tuning of the MHE to the process. Two real-time runs with MHE were performed and evaluated using root mean squared error (RMSE) between the estimates of biomass and sugar compared to offline CDW and sugar measurements.

The study demonstrates the feasibility of real-time sugar state estimation in *C. glutamicum* cultivation using MHE, but questions its current reliability as a soft sensor. The estimator follows the sugar dynamics with a low RMSE for the three experimental runs (two in real-time). The study highlights the importance of reliable measurements from stable and correctly calibrated equipment for optimal soft sensor performance. Suggestions for improvement include revisiting calibration steps, exploring alternative arrival cost updates, and conducting longer cultivation experiments to assess long-term efficiency. Future work should involve closed-loop control experiments to determine the robustness of the soft sensor for control purposes.

Sammendrag

En hinder i utviklingen av prosesskontroll og digitale tvillinger i bioprosessindustrien er mangel på direkte målinger i sanntid. Innen farmasøytisk- og matindustri brukes bakterier til å produsere mange ulike produkter, men de beskrives ofte av ikke-lineære tidvarierende dynamikker som gjør dem vanskelige å modellere nøyaktig. Moving Horizon Estimation (MHE) kan adressere dette problemet ved å kombinere sanntidsmålinger med en prosess-modell, samt inkorporere fysiske begrensninger for å estimere tilstander. MHE kan dermed fungere som en "soft sensor" som tillater sanntids overvåkning og muliggjør kontroll for å forbedre produktkvalitet og utbytte.

Denne oppgaven undersøker implementeringen av sanntids MHE som en løsning på begrensede målinger av sukker i kontinuerlig dyrking av *Corynebacterium glutamicum*. Arbeidet besto av eksperimentelt arbeid relatert til kjøring av bioreaktor og programmering ved bruk av optimalisering. Sanntidsmålinger besto av volum, celletetthet og CO_2, med periodiske målinger av sukker hver time, og offline prøver av tørrvekt av biomasse (CDW) hver 2-4 timer. Pumper og en optisk tetthetsprobe ble kalibrert. *In silico* Model Predictive Control ble brukt til å forhåndsbestemme inn- og utstrømmer til dyrkingen. Parameter-stimering ble utført for å bestemme parametere for modellen, fulgt av tilpasning av MHE til prosessen. To sanntidskjøringer med MHE ble utført og evaluert ved bruk av kvadratisk gjennomsnittsavvik (RMSE) mellom estimater av biomasse og sukker i forhold til offline-målinger av sukker og CDW.

Studien viser at sanntids estimering av sukker er mulig i en *C. glutamicum*-dyrking ved bruk av MHE, men stiller spørsmål til den nåværende robustheten av estimatoren. Esti-matoren følger sukkerdynamikken med lav RMSE for de tre eksperimentelle kjøringene (to i sanntid). Studien understreker viktigheten av pålitelige målinger fra stabilt og rik-tig kalibrert utstyr for optimal ytelse av "soft-sensor" teknologi. Forslag til forbedring av estimator inkluderer ny gjennomgang av kalibreringsprosessen, utforsking av alterna-tive arrival cost oppdateringer og utførelse av lengre dyrkingseksperimenter for å vurdere langtidseffektiviteten av MHE. Fremtidig arbeid bør innebære eksperimenter med lukket sløyfekontroll for å fastslå MHE sin robusthet for kontrollformål.

Preface

This thesis was written in the spring of 2022, at the Department of Chemical Engineering, as the final part of my Master of Science in Chemical Engineering and Biotechnology at the Norwegian University of Science and Technology (NTNU). The project idea was developed in collaboration with my supervisors to incorporate my interest in systems biology and process system engineering.

First I would like to thank my supervisor Professor Nadav Bar for the opportunity to work in a high-tech bioreactor lab, and the freedom to shape my own thesis. I would also like to thank Head Engineer Christopher Sørmo for his practical help in the lab, fixing the HPLC, and saving me from several failed bioreactor runs. I would like to thank Ph.D. student Che Fai Alex Wong for giving me, a process engineer, a crash course in being a microbiologist. Thanks to my fellow master's student Laura Sophie Zintel for her support and great collaboration in the laboratory. I would like to extend my biggest gratitude to Ph.D. graduate and co-supervisor, Andrea Tuveri. Andrea has been an invaluable resource, giving me support and feedback both through my project thesis, and helping me remote from Årdal this spring. Finally, thanks to friends and family for their support, and my fellow students and friends at Chemical Engineering and Biotechnology for making these five years an unforgettable time!

Declaration of Compliance
I declare that this is an independent work according to the exam regulations of the Norwegian University of Science and Technology (NTNU).

Helene Bøe
Trondheim, June 2, 2023

Previous work

In the fall of 2022, the author wrote a project thesis on "Moving horizon optimization strategies for feeding and estimation in bioprocesses". In the project, a closed loop controller with a nonlinear model predictive controller (NMPC) was implemented to optimize feeding for cell density setpoint tracking of *Corynebacterium glutamicum*. The Moving Horizon Estimator (MHE) was used as a state estimator, using volume, cell density, and CO_2 measurements to provide full-state feedback for control. The strategies were applied to a Fed-Batch and Continuously Stirred Tank Reactor (CSTR) in-silico model. The implemented controller successfully reached the Fed-Batch setpoint with minimal input usage, while the CSTR had higher input usage and chattering issues. State estimation for closed-loop systems had low root mean square errors. The code developed in the project thesis serves as a basis for this thesis, where the focus is on real-time state estimation using a fixed input profile for continuous cultivation.

Contents

List of Figures

List of Tables

Symbols and Abbreviations

Symbols used in the text

Symbol	Description	Unit
x	State	
$\hat{x}$	Estimated state	
$\dot{x}$	Differential equation	
u	Input	
y	Measurement/output	
t	Time	h
n_m	Control horizon	min
n_p	Prediction horizon	min
ℓ	Lagrangian interpolation polynomial	
N	Optimization horizon	min
P_L, V, W_k	Weighting matrices for MHE	
Q_w	Covariance matrix for MHE	
ω_i	Noise weight	
Q	Weighting coefficient for set-point importance	
R_1	Weighting coefficient for input usage penalization	
V	Volume	L
X	Biomass density	g/L
S	Sugar density	g/L
CO_2	Carbon dioxide in offgas	% gas flow out
F_i	Flow rate	L/h
CDW	Cell dry weight	g/L
pO_2	Partial pressure of oxygen	
μ_{max}	Maximum growth rate	h^{-1}
K_S	Monod growth constant	$g \cdot L^{-1}$
k_d	Death rate constant	h^{-1}
Y_{XS}	S from X yield	$g \cdot g^{-1}$
Y_{XCO_2}	CO_2 from X yield	$g \cdot g^{-1}$
ρ	Density	g/m^3 or g/L
σ	Standard deviation	
σ^2	Variance	

Abbreviations	
Abbreviation	Definition
ANN	Artificial Neural Networks
CDW	Cell Dry Weight
CU	Concentration Units
DAE	Differential Algebraic Equation
EKF	Extended Kalman Filter
HPLC	High-Performance Liquid Chromatography
MHE	Moving Horizon Estimator
MPC	Model Predictive Control
NIR	Near Infrared
NLP	Non Linear Programming
NMPC	Nonlinear Model Predictive Control
OD	Optical Density
ODE	Ordinary Differential Equation
PID	Proportional-Integral-Derivative
QP	Quadratic Programming
RF	Reinforcement Learning
UKF	Unscented Kalman Filter

Chapter 1

Introduction

1.1 Motivation

Bacteria are versatile and useful organisms that play a vital role in various industries by producing a wide range of products. Their use as fermentors in the dairy industry is well-known where they produce commodities like cheese and yoghurt (Kumar, 2017; Yang et al., 2012). Bacteria are also utilized in the pharmaceutical and biotechnology industry, where they have been used to produce antibiotics and vaccines, enzymes, vitamins and other chemical compounds (Kadner and Rogers, 2023). According to a recent report by Mikulic (2022) the biotechnology and pharmaceutical industry is expected to generate revenue of 505 billion US dollars in 2026. Furthermore, bacteria have been identified as having great potential in the production of biofuels from renewable sources, as they can convert organic matter into usable fuel (Chintagunta et al., 2021; Koppolu and Vasigala, 2016; Liao et al., 2016).

In this master thesis, the aerobic bacteria *Corynebacterium glutamicum* ATCC 13032 (*C. glutamicum*) will be cultivated. This bacteria is known as an industrial workhorse for its ability to produce a range of amino acids, such as L-glutamate and L-lysine (Kalinowski et al., 2003; Wendisch et al., 2016), organic acids (Wieschalka et al., 2012) and to grow on a range of substrates (Becker et al., 2016). It can also grow through fluctuating conditions in O_2, CO_2 and pH (Bäumchen et al., 2007; Follmann et al., 2009; Nishimura et al., 2007), though its main drawback is a relatively small maximum growth rate compared to other bacteria (Graf et al., 2019).

Bacteria, functioning as tiny industrial cell factories, exhibit complex dynamics that are

not easily modeled.Although they are widely used in industrial production, the integration of process monitoring and control remains a challenge in the biopharmaceutical industry (Pittman and Wetterhall, 2022). Developing accurate models of bioprocesses can not only improve the efficiency, performance, and quality of current production methods but also enable new applications and sustainable production. Luttmann et al. (2012) highlights how process monitoring and control are key when transitioning bioprocesses into commercial production, where products need to meet complex regulations. In addressing these challenges, model-based approaches that comprise of a controller and an estimator have shown promise in automating bioprocesses (Lucia et al., 2017).

In control engineering, a comprehensive understanding of process state variables and possibly parameters is crucial for controlling system output. However, creating accurate models for biological systems is difficult and often requires a deep understanding of the process (i.e. Flux Balance Analysis[1] (Jabarivelisdeh et al., 2020)). In practice, bioprocesses are often modelled using unstructured Monod kinetics[2], which uses lumped parameters to describe intracellular phenomena in a simplistic way (Jabarivelisdeh et al., 2020). When relevant state and parameter measurements are not available, state and parameter estimation that exploits online measurements is necessary.

Bioprocess control faces a significant challenge in the limited availability of real-time direct measurements. While online sensors like temperature, optical density, pH, and gas flow rates are commonly used, the options for online glucose measurements are nonexistent. Glucose serves as the primary carbon source for bacterial metabolism, and monitoring its levels provides valuable insights into the current growth state of the bioprocess (Mann et al., 2017). Existing measurement methods mainly rely on offline approaches (Galant et al., 2015), such as enzymatic assays (Blackwell, 2018) for glucose detection and analysis. Although there are some alternatives like single-use in-line optical glucose biosensors (Lederle et al., 2021) and at-line measurements using high-performance liquid chromatography (HPLC) providing periodic values, these methods still have limitations. At-line measurements often involve manual operations and provide infrequent data, making closed-loop feedback control challenging (Alford, 2006).

The lack of online measurements can be addressed by developing soft sensors (state esti-

---

[1]Flux balance analysis (FBA) is a modeling approach based on the genome-scale metabolic reconstruction of an organism. It describes intracellular metabolism by considering a network of metabolites and metabolic fluxes. These models, defined by the network's stoichiometry, determine an optimal metabolic flux distribution by maximizing a biological objective (Orth et al., 2010; Varma and Palsson, 1994).

[2]The Monod model is the commonly used relationship describing the specific growth rate (μ) and substrate concentration (S) for pure cultures on one growth-limiting substrate, described by $\mu = \mu_{max}\frac{S}{S+K_s}$ (Yoon et al., 1977).

mators (Dochain, 2003)), which are capable of accurate real-time estimation of the values of process and product quality variables. A soft-sensor is an indirect measurement that combines a mathematical model and frequently measured process data to predict an output, bypassing the challenge of infrequent or non-existent measurements (Desai et al., 2006). With a soft sensor compensating for lacking measurements, we construct a digital twin of the bioprocess, a digital representation of a real product instance (physical twin) (Schleich et al., 2017).

1.2   Thesis objective

The aim of this master thesis work is to investigate the thesis statement:

> *The implementation of moving horizon estimation can overcome the challenge of limited sugar measurement in bioprocesses, enabling reliable real-time state estimation in a continuous cultivation of C. glutamicum.*

In this work reliability is defined as the MHE's ability to continuously perform its core function, estimating sugar with reasonable values (here 0-20g/L), without disruptions, or significant reductions in performance. To determine the validity of the statement, we will use the Root Mean Square Error (RMSE) as a quantifier to compare the measured (y) cell dry weight (CDW) and sugar, with the estimated ($\hat{x}$) biomass and sugar. We will also compare the real-time state estimation results to those obtained from the nominal state model and online measurements. A soft sensor that provides a low RMSE for sugar, will not only provide insights into the current growth state of the bioprocess but also full state feedback enabling closed-loop control in bioprocesses. The novelty of the master lies in investigating the *real-time* performance of MHE on a bioprocess, not with offline in-silico simulations as has been done before (see Section 3.1).

Table 1.1: Available measurements, where online measurements are volume (V), cell density (X) and CO_2. CDW refers to cell dry-weight samples. * Sugar measurements (S) have a 40-minute delay, before being registered.

Measurement	Frequency	Type
V	60 s	Online
X	60 s	Online
S	1 h*	At-line
CO_2	60 s	Online
CDW	3-4 h	Offline

To check the performance of the real-time MHE, a fixed feeding curve with an initial batch phase of 6 hours is used, followed by continuous feeding and flow out. The input profile is to be determined with model predictive control on a nominal model. Available measurements are shown in Table 1.1, and the estimated states will be volume, cell density, sugar and CO_2. To reduce uncertainty, multiple pieces of equipment are calibrated, including the OD probe, and inlet and outlet pumps. To our knowledge, a continuous cultivation with *C.glutamicum* utilizing the nominal model in Equation (2.1) has not been carried out. Therefore, it is essential to perform a parameter estimation to determine the appropriate parameters to be used in the MHE. The steps mentioned above are presented in Figure 1.1
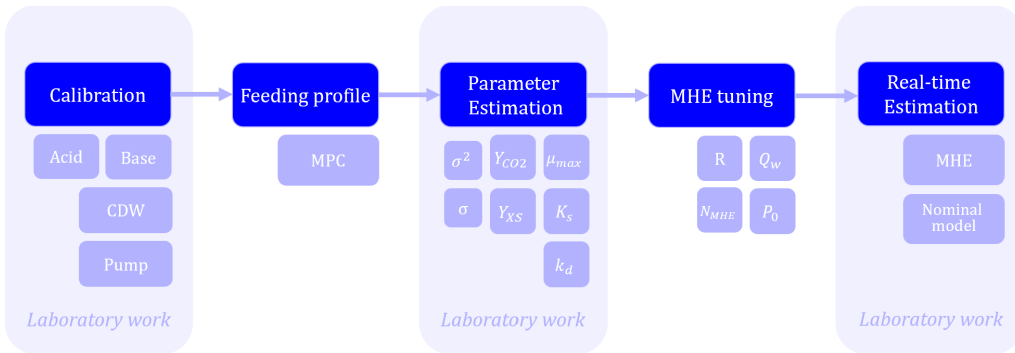
Figure 1.1: Summary of the work performed in this master thesis, presented through a flowsheet. The light purple boxes signify that laboratory work has been involved, in addition to programming. The smaller boxes beneath the main titles indicate either methods used (e.g. MHE & MPC) or instruments calibrated (e.g. acid, base and in/outlet pump, NIR probe), or parameters decided (e.g. μ_{max}, R, N_{MHE}). Note here that to perform parameter estimation a continuous cultivation had to be performed, with a feeding profile equal to the one intended for the real-time MHE. For this reason, a feeding and input profile was found using parameters from (Tuveri et al., 2021), a Fed-batch cultivation, before carrying out a parameter estimation for our continuous process.

1.3   Thesis outline

This section presents the structure of the thesis. The thesis consists of seven chapters including this chapter, and Appendices A to E.

Chapter 2 presents the bioprocess investigated described by a system model of four equations, and the RMSE, the quantitative measure of how well the estimator is performing. Section 2.2 presents the parameter estimation performed to determine the parameters in the system model.

Chapter 3 gives a general introduction to optimization, a state of the art on controllers and

state estimators (summarized in Figure 3.2), before shifting focus to MPC and MHE, the controller and state estimator used in this thesis. The following sections, Section 3.2 and Section 3.3, commence with the underlying theory on the respective topics, including the methodologies used to solve the Nonlinear Programming Problems (NLPs). Lastly, the sections present the methodology that serves as the foundation for the results.

Chapter 4 gives a general introduction to bioprocesses, and the bacteria used, before presenting the experimental setup. Chapter 4 goes into detail on how the experimental work was performed and how measurements were collected through signal processing. Section 4.3 presents how the peristaltic pump and OD-probe were calibrated. The majority of the thesis work consisted of experimental work, and an evaluation of the experimental methods related to their uncertainty is therefore also included. Four experimental bioreactor runs are presented in this thesis. The first was used for calibration of the OD-probe. The second experimental run was used for parameter estimation and MHE tuning. The third and the fourth runs were real-time MHE, referred to as the first and second real-time MHE run in the thesis.

Chapter 5 presents the results of the calibrations, parameter estimations, MHE tuning, two real-time MHE runs and offline tuning on the experimental runs.

Chapter 6 discusses the results presented in the previous chapter with regards to if the implementation of moving horizon estimation can overcome the challenge of limited sugar measurement in bioprocesses, enabling reliable real-time state estimation in a continuous cultivation of *C. glutamicum*. Section 6.1 presents recommendations on further work.

Chapter 7 summarizes the overall conclusions and recommendations for further work.

Appendix A shows statistics on publications and citations on Moving Horizon Estimation over time.

Appendix B presents calculations and details regarding experimental work, like offgas calculations, and linear calibration curves for the OD-probe.

Appendix C lists model parameters used for the MPC.

Appendix D contains the protocol for running the bioreactor.

Appendix E showcases the code developed for the thesis in the form of a file directory accompanied by explanations. The code has been provided as a .zip file, which is attached to the Inspera delivery.

Chapter 2

Process Description

2.1 System model
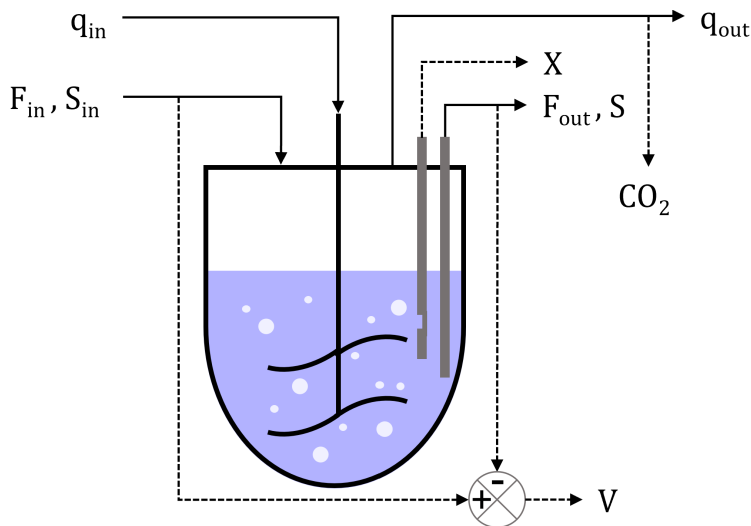
Figure 2.1: The bioprocess system for *Corynebacterium glutamicum* cultivation, described as ODEs in Equation (2.1), includes inputs, outputs, and measurements. Solid lines represent flows, with gas flows (q_{in} and q_{out}) and liquid flow in (F_{in}) having constant composition (S_{in}). Liquid flow out (F_{out}) exhibits a varying sugar concentration (S). Dashed lines represent measurements, including biomass concentrations (X), offgas production (CO_2), and volume (V) determined by the inflow and outflow rates. Air supply (q_{air}) is denoted by bubbles in the tank supplied by q_{in}.

The system model is built upon the work in Tuveri et al. (2021) which describes aerobic bacterial cultivation, using a simple Monod model. For the system model we assume the

tank is mixed homogeneously, temperature and pH are constant, all vital nutrients are available and the microbes in the reactor are a pure culture. The dynamics of the nominal system are described by the following ordinary differential equations (ODEs),

$$
\begin{aligned}
\frac{dV}{dt} &= F_{in} - F_{out} \\
\frac{dX}{dt} &= -\frac{F_{in}}{V}X + \mu_{max}\frac{S}{S + K_s}X - k_d X \\
\frac{dS}{dt} &= \frac{F_{in}}{V}(S_{in} - S) - \mu_{max}\frac{S}{S + K_s}\frac{X}{Y_{XS}} \\
\frac{dCO_2}{dt} &= \mu_{max}\frac{S}{S + K_s}\frac{X}{Y_{XCO_2}} - q_{air}CO_2
\end{aligned}
\tag{2.1}
$$

where F_{in} and F_{out} is the feed into and liquid flow out of the system, and q_{air} represents a constant gas flow of ambient air into the system. The states vector is given by,

$$
x(t) = \begin{bmatrix} V(t) & X(t) & S(t) & CO_2(t) \end{bmatrix}^T
\tag{2.2}
$$

where V is the volume, X is the biomass concentration, S is the substrate concentration, which here is glucose, referred to as sugar, and CO_2 the offgas. The volume is given in liter, the cell mass and substrate as a concentration in g/L, and CO_2 is given as a percentage of the gas flow out. The input variable, u is given by,

$$
u(t) = \begin{bmatrix} F_{in}(t) & F_{out}(t) \end{bmatrix}^T
\tag{2.3}
$$

where F_{in} and F_{out} given in L/h. For continous cultivation F_{in} is set equal to F_{out}. The variables in the system model are shown in the bioprocess in Figure 2.1.

The parameters are given by the vector

$$
\theta = \begin{bmatrix} \mu_{max} & K_S & k_d & Y_{XS} & Y_{XCO_2} \end{bmatrix}^T
\tag{2.4}
$$

Where the parameter values in the equations above are given in Table 5.3, determined by the method in Section 2.2.

The differential equations can be concisely written as,

$$
\dot{x} = f(x, u) = \begin{bmatrix} \frac{dV}{dt} & \frac{dX}{dt} & \frac{dS}{dt} & \frac{dCO_2}{dt} \end{bmatrix}^T
\tag{2.5}
$$

The change in cell density, X, is equal to the formation of new cells (dependent on available

sugar), a decrease in cell concentration (due to dilution caused by flow in), and death. The change in substrate glucose, S, is equal to the substrate added, and substrate decrease due to dilution and bacterial consumption. The change in carbon dioxide is equal to the CO_2 generation (due to cell respiration) and loss through the gas flow out. The term $\frac{S}{S+K_s}$ takes into account the saturation in the bacterial system, also known as the Monod equation, where μ_{max} represents the maximum growth.

As a quantitative measure of how well the estimator is performing, we will use the root mean square error (RMSE) defined as,

$$\text{RMSE}_{y^*} = \sqrt{\frac{1}{n_{y^*}} \sum_{i=1}^{n_{y^*}} (y^* - \hat{x})^2} \tag{2.6}$$

where i is the current timestep for the offline measurements and n_{y^*} the amount of offline measurements . The RMSE is calculated using the deviation between the *offline* measurements (y^*) of biomass (y_X^*) and sugar (y_S^*) and the state estimates of biomass ($\hat{x}_X$) and sugar ($\hat{x}_S$). For comparison to the MHE, the $RMSE_{y^0}$ between the nominal model (y^0) and the *offline* measurements is also calculated.

2.2   Parameter estimation

The parameters in Equation (2.4) must be determined for the process investigated, a continuous cultivation. Parameter estimation is the process of determining the values of unknown quantities in a model (called parameters) based on available data. In biological systems, these parameters may include reaction rates, Hill coefficients[1], or binding affinities (Jeong and Qiu, 2018). Estimating these parameters is important because they help describe the relationships between the variables in the model and lead to a deeper understanding of a given process. However, measuring these parameters directly is often difficult or impossible. The goal of parameter estimation is therefore to find the parameter values that best fit the available data. To achieve this, the estimation process uses available measurements and compares them to predicted values from the model. As an objective function, a norm of the measurement error is typically used. The type of norm used depends on the statistical distribution of the measurement errors. For example, if the errors are independent, normally distributed with a zero mean and known variances, a weighted least squares function

---

[1]The Hill Coefficient (n_H) is a measure of cooperativity in biomolecular binding or enzymatic reactions, indicating how the binding or reaction rate changes with ligand or substrate concentration. A coefficient greater than 1 signifies positive cooperativity, while a coefficient less than 1 suggests negative cooperativity (Nelson and Coc, 2017).

can be interpreted as a maximum likelihood estimate (Bock et al., 2012; Kostina, 2004).

To have a model that gives satisfactory predictions, the model parameters must be both identifiable and estimable. A model is identifiable if a unique input-output behaviour for each set of candidate parameter values exists. Identifiability analysis can be used to uncover problems with model structure, while estimability, on the other hand, is related to whether parameters can be uniquely estimated using existing experimental data (McLean and McAuley, 2011).

To evaluate the accuracy (standard deviation) of estimated parameters the Fisher Information Matrix (FIM) can be applied. FIM contains information on the measurement uncertainty and sensitivities of predicted responses to model parameters at all times. The FIM can be approximated by the number of samples (N), the weighted residual sum of squares (RSS) and the Jacobian matrix (J) for the set of optimal parameters ($p = \hat{p}$) in the following manner (Karakida and Osawa, 2021; Natal A W Van Riel, 2011),

$$FIM_p = N(RSS)^{-1}JJ^T \mid_{p=\hat{p}}$$
(2.7)

The parameter covariance matrix can be calculated as,

$$cov(\hat{p}) = FIM_p^{-1}$$
(2.8)

2.2.1 Method

According to Tuveri et al. (2021) the investigated system in equation (2.1) is structurally identifiable and estimable, making it susceptible to parameter estimation. The experimental data were obtained by running a bioreactor following the procedure given in Section 4.2.1. The parameters were obtained through parameter estimation on the experimental data set of volume, biomass, glucose and CO_2 values from the off-gas analyzer. The estimation was performed using a nonlinear least-squares data fitting algorithm (lsqnonlin © 1994-2023 The MathWorks (2006a)). We assume independent and normally distributed

errors and therefore use the least squares estimator defined as,

$$
\min_{\theta} \quad \sum_{i=1}^{N} \| \frac{y_{i,k} - x_{i,k}(\theta)}{max(y_k)} \|^2
$$

$$
\text{s.t.} \quad x_{i+1} = F(x_i(\theta), u_i) \qquad i = 0 \cdots, N-1 \tag{2.9}
$$

$$
x_0(\theta) = y_0 \qquad\qquad\quad i = 0
$$

$$
\theta_{min} \leq \theta \leq \theta_{max} \qquad\quad i = 0, \cdots, N
$$

where y is the measurements of V, X, S and CO_2, where S is the at-line sugar measurements. $x(\theta)$ are the states calculated from the estimated parameters. Volume is not seen as an important state and is therefore not included in the square estimator, seen by the use of subscript k, which refers to measurements of X, S and CO_2. CDW samples were taken to verify if calibration curves (Equation (5.3)) were sufficient. $max(y_k)$ is used as a scaling factor to make the solver run more reliably (© 1994-2023 The MathWorks, 2006b). θ_{min}, θ_{min} and the initial guess, θ_0 are given in Table 5.2, found by trial and error. The results of the parameter estimation are presented in Chapter 5.

Chapter 3

Optimization

Optimization involves making the best use of resources to maximize or minimize an objective, a quantitative measure of the performance of the system (Nocedal and Wright, 2006). Optimization algorithms are crucial for efficient decision-making in chemical process systems, where small design and operation changes can yield significant improvements in efficiency, product quality, environmental impact, and profitability (Biegler and T., 2010). An optimization problem includes an objective function ($\phi(x)$), a system model (x), variables, and constraints (c(x), g(x)). A general constrained optimization problem can be expressed as,

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & \Phi(x) \\
\text{s.t.} \quad & c(x) = 0 \\
& g(x) \leq 0
\end{aligned}
\tag{3.1}
$$

In optimization, the objective function serves as a metric for evaluating the performance of minimizing (or maximizing) a quantity of interest. This function can take various forms, such as the cost, profit, or yield of a system. States and variables in an optimization problem are typically subject to constraints that restrict their feasible values, either as equality (c(x)) or inequality (g(x)) constraints. Decision variables (x in Equation (3.1)) represent the variables that can be modified to achieve optimal performance. In engineering problems, decision variables can be thought of as the degrees of freedom that govern the system's behaviour.

Dynamic optimization involves solving differential and algebraic equation mathematical models that are time-dependent (Floudas et al., 1999). The Direct method is a common approach for solving dynamic optimization problems. This method discretizes continuous

time variables, transforming the problem into a finite-dimensional nonlinear programming problem (NLP) (Biegler and T., 2010). Multiple shooting and Direct collocation are popular Direct methods, with other approaches also available, as seen in Figure 3.1. Simultaneous methods, such as the previous mentioned, discretize both control inputs and the states, allowing large-scale NLP solvers to find solutions at specified intervals in a time horizon problem (Biegler and T., 2010).
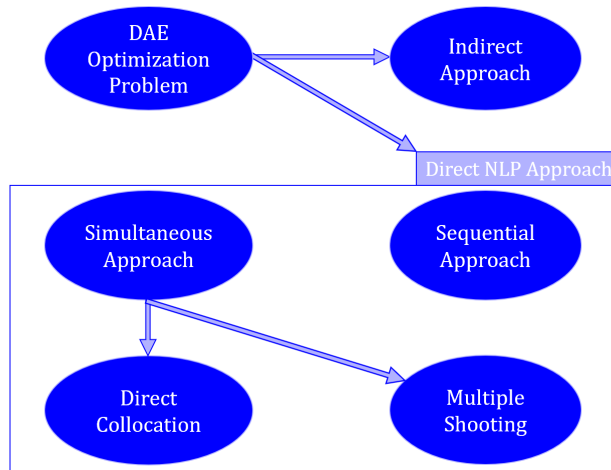
Figure 3.1: The figure shows different approaches to solving dynamic optimization problems. The indirect approach is also known as Optimize then Discretize. The sequential approach is more commonly known as Single Shooting (Biegler and T., 2010).

3.1   Review on Controllers and State Estimators

3.1.1   Controllers

Controller technology is a crucial aspect of commercializing bioprocesses to meet complex regulations (Luttmann et al., 2012) and improve the efficiency, performance and quality of current production. One widely used control strategy in the chemical industry is the Proportional-Integral-Derivative (PID) controller, known for its simplicity and ease of implementation. However, in fermentation processes, simple PI(D) controllers only work well for a restricted phase as their parameters are static and do not change in accordance with the system dynamics (Gnoth et al., 2008). There are alternatives to PI(D)s that can handle non-linearity, like fuzzy control, but these still lack some dynamic adaptability (Rathore et al., 2021).

To overcome the limitations of PID controllers, alternative control strategies have been

investigated. One such strategy is Model Predictive Control (MPC), a control technique that takes into account the dynamic nature of the system and can handle multiple inputs and outputs (Rawlings et al., 2017). MPC has been applied for optimal setpoint tracking in bioprocesses modelled by basic Monod equations (Ramaswamy et al., 2005; Tebbani et al., 2008), and utilized to incorporate economic objectives, such as maximizing product yields (Ashoori et al., 2009; Raftery et al., 2017). One advantage of MPC is its ability to consider future predictions and operational constraints, enabling early detection of potential issues and providing system decoupling by capturing the interactions between input and output variables (Seborg et al., 2016). This does however come at the cost of high computational costs compared to other control strategies (Rathore et al., 2021).

Another emerging approach in control is the use of reinforcement learning, a machine learning approach that focuses on optimizing an agent's behaviour within an environment. The agent learns to select actions based on observed states to maximize an external reward. Training is typically done through episodes, which consist of sequences of states, rewards, and actions until a terminal state is reached, with the total reward obtained during an episode referred to as the return (Kaelbling et al., 1996). Reinforcement learning has shown promise in process control applications and has the potential to challenge traditional MPC and PID controllers (Hedrick et al., 2022; Oh et al., 2022; Pan et al., 2021; Petsagkourakis et al., 2020; Treloar et al., 2020b; Xie et al., 2020). Its advantage lies in its adaptability to varying conditions, as demonstrated in the control of microbial co-cultures in bioreactors (Treloar et al., 2020a). However, a barrier to real-world application is the requirement for large amounts of training data (Dulac-Arnold et al., 2021).

3.1.2 State estimators

A state estimator is a technique developed to address the lack of system measurements by providing estimates of unmeasured variables in a process (Gadkar et al., 2005). It has applications in process monitoring, mathematical model fitting and update, transient data reconciliation, and feedback control (Salau et al., 2012). By estimating essential variables that are not directly measured, such as substrate, biomass, and product concentration, state estimators can be successfully applied for process monitoring and feedback control ensuring the proper functioning of industrial plants (Salau et al., 2012). Several state estimators for nonlinear systems exist, where the extended Kalman filter (EKF), particle filter (PF) and Moving Horizon Estimator (MHE) are well-known Bayesian estimators. Bayesian estimators are a category of estimators that utilize probability distribution estimation to estimate the state variables of a system based on available data. These estimators assume
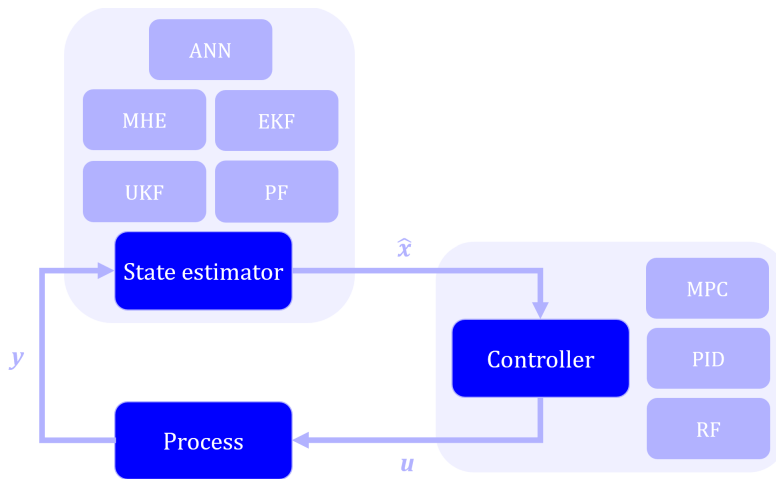
Figure 3.2: Summary of the topics presented in the state of the art for state estimators and controllers of nonlinear processes, and how they are connected. The state estimator receives measurements, y, from the process, and uses one of the methods listed to provide a state estimate ($\hat{x}$); Artificial Neural Networks (ANN), Moving Horizon Estimator (MHE), Extended Kalman filter (EKF), Unscented Kalman Filter (UKF) or Particle Filter (PF). The controller recieves the state estimate ($\hat{x}$) and provides an input action (u) to the process by using one of the strategies listed; PID, Model Predictive Control (MPC) or Reinforcement learning (RF).

that all variables are stochastic, allowing for the determination of the distribution of state variables using measured variables (Ali et al., 2015).

The EKF is an extension of the traditional Kalman filter that can handle nonlinear system models by linearizing them around the current state estimate. It has gained significant attention as a nonlinear state estimator due to its relative simplicity and low computational requirements, demonstrating effectiveness in handling some nonlinear problems (Salau et al., 2012). However, the implementation of EKF suffers from numerical challenges due to linearization, particularly when dealing with highly non-linear processes, it cannot accurately incorporate physical state constraints (Kandepu et al., 2008), and may fail to converge when given a poor initial guess of the state (Haseltine and Rawlings, 2004). The Constrained Extended Kalman Filter (CEKF) is able to incorporate constraints but suffers from similar convergence issues when given poor initial guesses (Salau et al., 2012). Yousefi-Darani et al. (2020) presents several applications of EKF for cultivation processes in the period of 1991–2020, while there are other recent applications as in Tuveri et al. (2021).

Particle Filters (PF) and Unscented Kalman Filters (UKF) are nonlinear estimation tech-

niques used to address the numerical challenges associated with nonlinear systems. PFs use a set of particles to represent the probability density function of the system state. By propagating these particles through the system model and weighting them based on measurement likelihood, PFs provide estimates of the state distribution (Sileshi et al., 2015). PFs have been applied to several bioprocesses such as; estimating states during penicillin production (Golabgir and Herwig, 2016; Kager et al., 2018), and online-monitoring of a cultivation using complex substrate mixtures (Sinner et al., 2021). PFs are flexible in their application as they do not pose any requirements on the functions or distribution, however, they suffer from requiring more computational effort than the EKF and cannot deal with constraints (Stelzer et al., 2017).

On the other hand, UKF is an extension of the unscented transformation to the Kalman filter. It addresses the limitations of the EKF by using a fixed number of deterministically chosen sampling points (sigma points) to represent the state distribution. These sigma points are propagated through the nonlinear function, and the mean and covariance of the resulting propagation are approximated (Julier et al., 1995). The method yields more accurate state estimates, capturing state uncertainty more effectively compared to estimates obtained from the EKF. While this approach incurs a higher computational cost, it remains less demanding than PFs (Kandepu et al., 2008). The UKF has been applied to several bioprocesses such as; bioprocess monitoring of a Fed-Batch cultivation of *C.glutamicum* (Tuveri et al., 2021) , joint estimation for state and parameter in the bio-dissimulation process of glycerol to 1,3-PD in batch culture (Zhu and Feng, 2012) and to estimate the glucose and biomass in lactic acid fermentation (Gonzalez et al., 2015).

Moving Horizon Estimation (MHE) is a state estimator that has gained increasing attention in recent years (of Science, 2021). MHE formulates the state estimation problem as an optimization task, where the objective is to minimize the discrepancy between the predicted system outputs and the measured outputs over a finite time horizon. MHE is particularly useful in cases where the system is subject to noise and uncertainty, and when the system model is nonlinear or time-varying (Rawlings et al., 2017). The MHE can easily incorporate constraints in states and parameters, recover from bad initialization and if expanded, handle delayed multi-rate measurements (Elsheikh et al., 2021b; Kühl et al., 2011b). MHE has been implemented in-silico on several bioprocess (Elsheikh et al., 2021b; Kim et al., 2023; Taylor et al., 2022; Tebbani et al., 2013), however, the implementation of MHE combined with experimental data is slim with only two applications in Goffaux and Wouwer (2008); Tuveri et al. (2022) to the best of our knowledge.

Artificial Neural Networks (ANN) have also been investigated as state estimators. ANN

refers to a network of interconnected nodes, known as neurons, where the connections between neurons (edges) have associated weights. The output of each neuron is calculated by considering the weighted inputs. Typically, an ANN consists of an input layer, an output layer, and one or more hidden layers in between. The hidden layers help in capturing complex relationships and patterns in the input data, while the output layer provides the final predictions or estimates. ANNs are therefore capable of learning complex mappings between inputs and outputs and can be trained to estimate system states based on available measurements (Wang, 2003). Helleckes et al. (2022) discusses several examples of ANNs used as soft sensors in bioprocesses. ANN are black box models, meaning no a priori knowledge about the system is required (Wilson and Zorzetto, 1997). However, high dependency on correct initial conditions for the process states and a significant body of experimental data limits its performance (Kräling et al., 2008).

3.2   Model Predictive Control

3.2.1   Theory

Model predictive control (MPC) is a widely used optimization-based strategy that can handle multivariable, constrained systems. When faced with nonlinear constraints and dynamics, nonlinear MPC (NMPC) strategies can be used. We will continue using the term MPC, not NMPC although we are referring to a nonlinear MPC. It is crucial that the system model is a good representation of the actual process, as an inaccurate model will result in inaccurate predictions of the output variables (Seborg et al., 2016). In MPC, a quadratic objective function is used to control variables, driving them towards their desired values while stabilizing input variable profiles (Biegler and T., 2010). This quadratic programming problem (QP) is solved online, and can be defined by the MPC QP equation,

$$\min_{u,y} \quad \frac{1}{2}(\sum_i^{n_p}(y_i - y_{SP,i})^T Q(y_i - y_{SP,i}) + \quad \sum_i^{n_m}(u_i - u_{i-1})^T R_1(u_i - u_{i-1})$$

$$+ \sum_i^{n_m}(u_i - u_{SP,i})^T R_2(u_i - u_{SP,i}))$$

$$\begin{aligned}
\text{s.t.} \quad & x_{i+1} = F(x_i, u_i) && i = 0, \cdots, n_p - 1 \\
& x_0 = x(0) && i = 0, \cdots, n_p \\
& y_i = C(x_i, u_i) && i = 1, \cdots, n_p \\
& G(x_i, u_i) \leq 0 && i = 1, \cdots, n_p \\
& u_{min} \leq u_i \leq u_{max} && i = 0, \cdots, n_m \\
& -\Delta u_{max} \leq \Delta u_i \leq \Delta u_{max} && i = 1, \cdots, n_m \\
& \Delta u_{max} = 0 && i = n_m + 1, \cdots, n_p
\end{aligned}$$

$$(3.2)$$

where x is the state variables, y is the system output, u is the input variables, n_m is the control horizon, and n_p is the prediction horizon. δu_{max} is the maximum difference the manipulated variable may vary at each iteration, while u_{min} and u_{max} are boundary values for the MV.

The objective function in Equation (3.2) consists of three terms. The first term serves to track the state and penalize deviations from the desired state. The remaining two terms are regularization terms that promote MPC stability and minimize the utilization of inputs. The tuning parameters Q, R_1, and R_2 determine the relative significance of these terms, allowing for prioritization based on specific requirements (Biegler and T., 2010). It is

worth noting that the inclusion of all terms in the objective function is contingent upon the user's preferences and intentions. The constraints in Equation (3.2) encompass both hard constraints (set as equal to or less than zero) and soft constraints. Hard constraints must be strictly satisfied throughout, but in certain scenarios, it may be inevitable to encounter constraint violations, leading to an infeasible Quadratic Programming (QP) solution. An example of such violations can occur when output variables are affected by significant disturbances. To tackle this challenge, the concept of soft constraints is introduced, allowing for violations within a penalty framework incorporated into the cost function. This flexibility empowers MPC to effectively handle constraint violations.

At each time step, an optimization problem is solved over a prediction horizon (the time span where states are included), minimizing an objective (cost function) while obeying the system constraints. When an optimal solution is found, the first component of the vector of control inputs is extracted and applied to the system. As this procedure is repeated at each time instant, the strategy is called a receding-horizon (Morari and Lee, 1999).

Multiple shooting

Direct multiple shooting is a simultaneous approach (see Figure 3.1) used to solve a dynamic optimization problem, for example, MPC problems, as it simultaneously solves the optimization problem and the system ODEs. By applying multiple shooting, the MPC can be solved as a nonlinear program with continuity constraints over the subintervals. This approach, using segmentation, provides a computationally efficient way to solve MPC problems, even for systems with unstable or nonlinear dynamics (Biegler and T., 2010). Multiple shooting is a method used to solve dynamic optimization problems by splitting the time horizon into multiple subintervals. This allows the ordinary differential equation (ODE) to be transformed into a set of nonlinear algebraic equality constraints, referred to as *gap constraint*, seen as $c(\omega)$ in equation (3.5). It is this equality constraint that provides a continuous solution, and forces the states (x) to become decision variables in addition to the inputs (u). The subintervals, $[t_k, t_{k+1}]$ are in the range

$$t_0 < t_1 < \cdots < t_{n-1} < t_n = t_f$$

where t_0 and t_f mark the start- and end-time of the discrete-time horizon respectively, and n is the number of subintervals. At each subinterval, the input function, u, is parameterized,

$$u(t) = u_k t \in [t_k, t_{k+1}], \quad k = 0, \ldots, n-1 \tag{3.3}$$

The parameterization of the initial condition of the state vector and the state trajectories,

$$
\begin{aligned}
y_k &= h(x_k), & k &= 0, \ldots, n-1 \\
\dot{x}_k &= f(x_k(t), u_k, t), & t &\in [t_k, t_{k+1}]
\end{aligned}
\tag{3.4}
$$

The multiple shooting NLP can then be defined in the following manner including the continuity constraints,

$$
\begin{aligned}
\min_{w} \quad & J(w) \\
\text{s.t.} \quad c(w) &= \begin{bmatrix} h_0 - x(0) \\ h_1 - x_0 \\ h_2 - x_1 \\ \ldots \\ h_{N-1} - x_{N-2} \end{bmatrix} = 0 \\
g(w) &= \begin{bmatrix} g(h_0, u_0) \\ \ldots \\ g(h_{N-1}, u_{N-1}) \end{bmatrix} \leq 0
\end{aligned}
\tag{3.5}
$$

where J, is the objective function, and the parameter vector for all subintervals becomes $w = [x_0, u_0, x_1, u_1, \ldots, x_{N-1}, u_{N-1}, x_N]^T$. (Tamimi and Li, 2009)

3.2.2 Method

To enable continuous cultivation to gather experimental data for parameter estimation, and perform real-time Moving Horizon Estimation (MHE), it was necessary to establish an input profile for the inflow and outflow. This section presents the method employed to determine the optimal input profile.

A nonlinear MPC was implemented with multiple shooting to find an optimal feeding profile for a continuous cultivation process. The NLP was solved using IPOPT, an interior point optimizer embedded in CasADI. The integration was solved using the CVODE integrator also embedded in CasADI. The goal of the optimization was to find an optimal feed profile that drives the system to a cell density of $X_{ref} = 15$ g/L throughout the process runtime of 24 hours. The optimization problem is formulated so that the first term of the objective function is setpoint tracking and the second term penalizes for large input usage. The quadratic non-linear problem is formulated in the following manner,

$$\min_{X,u} \quad \frac{1}{2}(\sum_{i}^{n_p}(X_i - X_{ref,i})^T Q(X_i - X_{ref,i}) + \sum_{i}^{n_m} \Delta u_i^T R_1 \Delta u_i)$$

$$
\begin{aligned}
\text{s.t.} \quad & x_{i+1} = F(x_i, u_i) && i = 0, \cdots, n_p - 1 \\
& x(0) = x_0 && i = 0 \\
& 0 \le V_i \le 2.5 && i = 1, \cdots, n_p \\
& 0 \le X_i && i = 1, \cdots, n_p \\
& 0 \le S_i && i = 1, \cdots, n_p \\
& 0 \le CO_{2,i} && i = 1, \cdots, n_p \\
& 0 \le u_i \le 0.210 && i = 1, \cdots, n_m \\
& 0.0178 \le u_i \le 0.210 && t \ge 13h \\
& -0.8u_{max} \le \Delta u_i \le 0.5u_{max} && i = 1, \cdots, n_m \\
& \Delta u_i = 0 && i = n_m + 1, \cdots, n_p
\end{aligned}
$$

$$(3.6)$$

The feed (F_{in}) is considered to be the manipulated variable (MV), and the biomass density (X) is the controlled variable (CV). To ensure continous cultivation F_{in} is set equal to F_{out}. Δu_i is the difference between the current input and the previous input. Increases in input are limited to small changes, while decreases in input can have larger changes. This is done to imitate the pump found in the laboratory. The prediction and control horizon is presented in Table 5.1 and the input profile and dynamics in Chapter 5.

3.3 Moving Horizon Estimation

3.3.1 Theory

Moving Horizon Estimation (MHE) is a constrained optimization-based method utilized for estimating the current state of a dynamic system. The MHE estimates the state by minimizing the sum of squares errors between past measurements and state or output predictions over an N-length sliding window while incorporating the dynamic system model and constraints. The finite sliding-estimation window in MHE is a way to overcome the computational burden associated with solving the full optimization problem for the entire trajectory of states online (Alexander et al., 2020).

MHE can be understood as the dual of Model Predictive Control (MPC), a more well-known constrained optimal control problem on a finite horizon (Kühl et al., 2011a). The
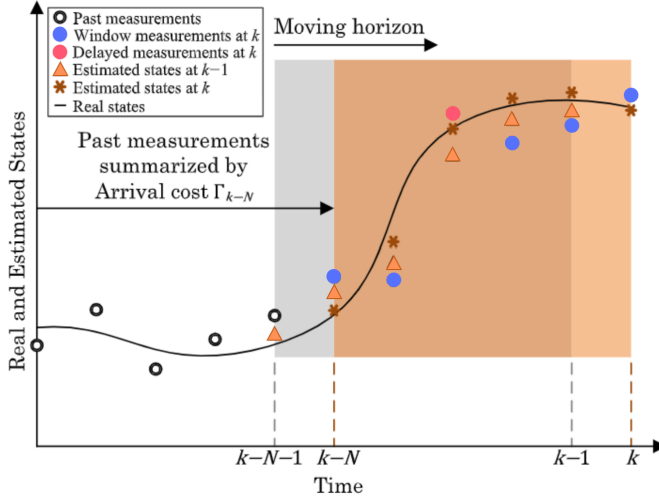
Figure 3.3: The figure shows the concept of MHE. The figure is reprinted and edited with permission from "A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications" (Elsheikh et al., 2021a).

duality between MHE and MPC becomes apparent as both are iteratively solved at each sampling instance. While MPC predicts the future states, the MHE estimates previous states.

The MHE problem, described by Kühl et al. (2011a), can be expressed as a minimization problem,

$$
\begin{aligned}
\min_{x_i, w_i} \quad & \left(\|\hat{x}_L - x_L\|_{P_L}^2 + \sum_{i=L}^{N} \|y_i - h(x_i)\|_V^2 + \sum_{i=L}^{N-1} \|w_i\|_{W_k}^2 \right) \\
\text{s.t.} \quad & x_{i+1} = F(x_i, u_i, w_i) \qquad i = L, \cdots, N-1 \\
& y_i = h(x_i) + v_i \qquad i = L, \cdots, N \\
& x_i \geq x_{min} \qquad i = L, \cdots, N
\end{aligned}
\tag{3.7}
$$

The goal of the optimization is to find the values of x_i and w_i that minimize the sum of squared error subject to the system model and constraints. The objective function consists of three terms. The first term, $\|\hat{x}_L - x_L\|_{P_L}^2$, of the objective function takes into account the arrival cost, which summarizes the effect of measurements prior to the estimation window up until point L. $\hat{x}_L$ denotes the optimal state estimate, while x_L represents the actual state at that point. The second term, $\sum_{i=L}^{N} \|y_i - h(x_i)\|_V^2$, is the measurement error cost, it

21

penalizes the distances of the expected output, $h(x)$, (based on the state estimates) from the actual measurements, y. The third term, $\sum_{i=L}^{N-1} \|w_i\|_{W_k}^2$, takes into account the process noise cost, so the evolution of the state in terms of the state equation. Incorporating process noise in the model is essential to account for model integration errors and modelling approximations.

The weighting matrices, P_L, V, and W_k, are defined as,

$$P_L = P^{-1/2}, \qquad V = R^{-1/2}, \qquad W_k = Q_k^{-1/2} \tag{3.8}$$

Here P, R, and Q_k are covariance matrices for error, measurement noise (v_i) and process noise (w_i). N is the number of data points in the estimation horizon, x_i the vector containing the states, u_i the vector containing inputs, and $F(x_i, u_i, w_i)$ describes the system model, which can be non-linear. P is the uncertainty of the estimated value, reflecting the confidence in the initial condition. So i.e. a high value, or high arrival cost, means high uncertainty in estimates. This means the algorithm trusts the initial state values more, which reduces how much the state value is adjusted by the estimator. R is the measurement uncertainty, so the values in R determine how the estimator weights prediction errors for the different measurements. Q_k is the model uncertainty that reflects the confidence in the process model predictions, so the values in Q_k affect how much the estimator punishes variations in states (Elsheikh et al., 2021b). The process noise matrix, Q_k, can be obtained by multiplying the Jacobian matrix G_k by the covariance matrix Q_w,

$$Q_k = G_k \cdot Q_\omega \cdot G_k^T \tag{3.9}$$

where G_k is the Jacobian

$$G_k = \left[\; \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, \omega)}{\partial \omega} \; \right] \tag{3.10}$$

G_k is the matrix of partial derivatives of the system model with respect to the noise vector $\omega \in \mathbb{R}^{n_x + n\theta}$. The tuning parameter ω is different from w which is the process noise random variable. Q_ω is a constant related to the statistics of the parameter uncertainty

The covariance matrix Q_w has variances of the noise parameters (θ) and states (x) along the diagonal, and covariances of zero for all parameters off the diagonal. The MHE is tuned by adjusting values in the diagonals of the covariance matrices R and Q_w, and the weights for each state in the initial arrival cost (P_0). R and Q_w have constant values, while G_k and P are updated on every iteration by finding an analytical solution to the arrival cost, described in detail in Kühl et al. (2011b).

Orthogonal collocation on finite elements

To solve the MHE NLP, orthogonal collocation can be applied. Direct Orthogonal Collocation on Finite Elements is a method used to solve dynamic optimization problems. The approach is based on a full discretization scheme, meaning that the optimization problem is discretized using symbolic variables instead of using an ODE/DAE solver. This is in contrast to shooting methods that use solvers to approximate the solution.
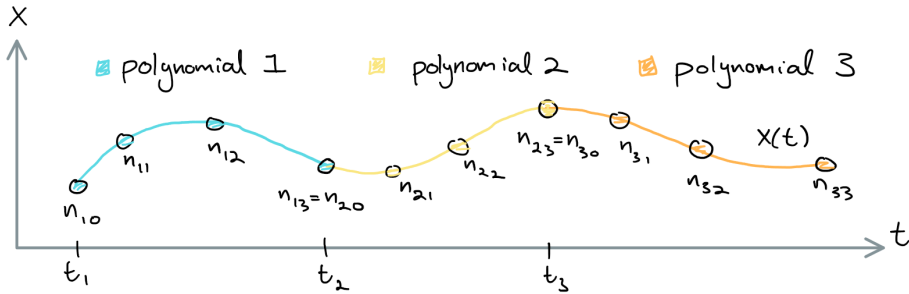
Figure 3.4: Discretization of a continuous time representation using Orthogonal Collocation. This is done to allow large-scale nonlinear programming (NLP) solvers to find solutions to differential equations at specified intervals in a time horizon.

To implement this method, the solution of the ODE is approximated using N polynomials of order K for a given horizon. The polynomial is approximated between each time step using d collocation points. The position of the collocation points is determined by the roots of either Gauss-Legendre, Gauss-Radau, or Gauss-Lobatto in an [0,1] interval and must be scaled to the system's interval (Biegler and T., 2010). The gap constraint ensures that the last state in the current interval is equal to the first state in the new interval,

$$x_k - x_{k,end} = 0 \tag{3.11}$$

To approximate the derivative of the function x(t), we require a function q(x) $\approx$ dx/dt. This function is approximated using a range of orthogonal polynomials by partitioning the horizon into finite elements and performing collocation on each element. For the ODE,

$$\frac{dx}{dt} = f(z(t), t), z(0) = z_0 \tag{3.12}$$

, the solution of the differential equations at discrete time points can be approximated by the Lagrange interpolating polynomial of order K,

$$z^K(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_K t^K \tag{3.13}$$

23

By using Lagrangian interpolation polynomials and K+1 interpolation points, the state of each finite element (*i*) is given by the sum of the product of Lagrangian polynomials and their respective states,

$$z(t) = \sum_{j=0}^{K} \ell_j(\tau) z_{ij} \tag{3.14}$$

where

$$\ell_j(\tau) = \prod_{k=1, \neq j}^{K} \frac{\tau - \tau_k}{\tau_j - \tau_k} \tag{3.15}$$

The collocation equations for the DAEs can be written as a system of equations consisting of a set of equations that enforce the collocation conditions, and another set that enforces the algebraic equations that must hold at each time point. The collocation equations for the DAEs can be written as,

$$\sum_{j=1}^{K} \dot{\ell}_j(\tau_k) z_{ij} - h_i f(z_{ik}, y_{ik}, u_{ik}, p) = 0, \quad i \in \{1, \dots, N\}, k \in \{1, \dots, K\} \tag{3.16a}$$

$$g(z_{ik}, y_{ik}, u_{ik}, p) = 0, \quad i \in \{1, \dots, N\}, k \in \{1, \dots, K\} \tag{3.16b}$$

as given in (Biegler and T., 2010).

These equations can be solved using nonlinear programming solvers to find solutions to differential equations at specified intervals in a time horizon. Where the NLP for orthogonal collocation on finite elements can then be written as,

$$\begin{aligned} \min_{w} \quad & \Phi(w) \\ \text{s.t.} \quad & f(\dot{x}, x, y, u) = 0 \\ & c(x) = 0 \\ & g(x) \leq 0 \end{aligned} \tag{3.17}$$

where $w = [x, x_{quad}, u]^T$, c and g are equality and inequality constraints respectively.

3.3.2   Method

As mentioned in the introduction, the code for the MHE was developed during the project thesis. However, the MHE must be adapted with relevant parameters, tuned, and made compatible to run in real-time in the laboratory.

MHE was implemented in real-time in the laboratory for continuous cultivation using a predefined feeding profile (see Section 3.2.2). The MHE NLP was solved using IPOPT, an interior point optimizer embedded in CasADI. The MHE optimization horizon, or window was divided into N segments and each segment's states were approximated with an interpolation polynomial of degree 3. A sampling time, k, of 60 seconds was used. The setup of the MHE can be seen in Figure 3.5 and the final tuning and horizon presented in Table 5.5 in Chapter 5.
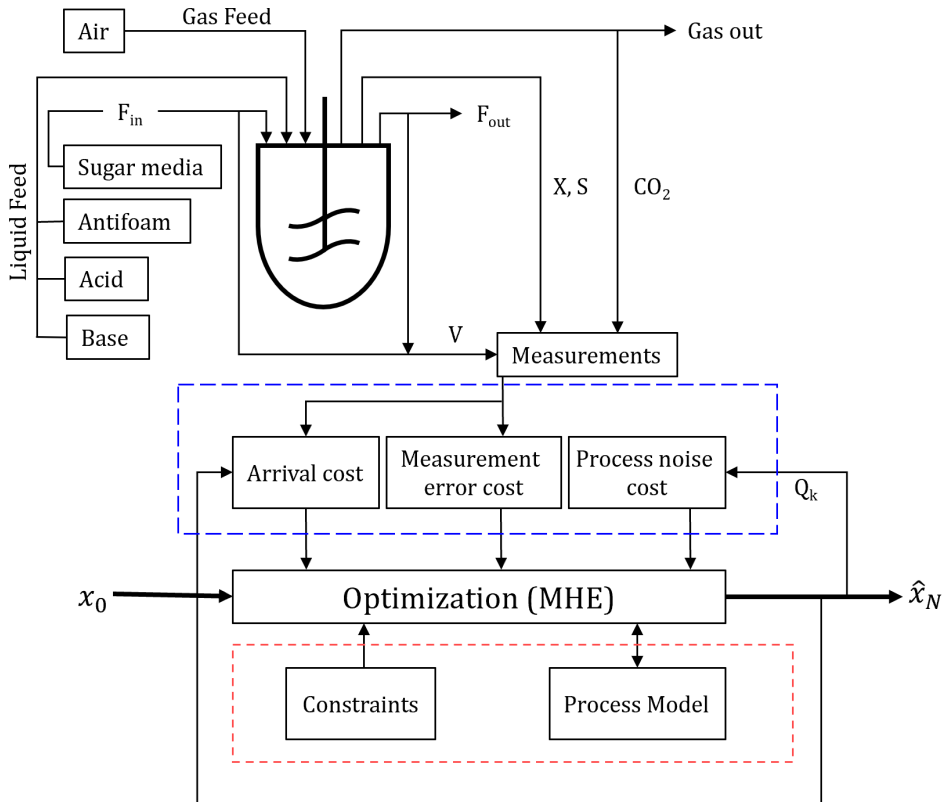
Figure 3.5: The figure shows the experimental setup (top section) and the MHE implementation (below measurements box). The blue dashed box represents the objective (cost) function presented in Equation (3.23). The orange tightly-dashed box is the optimization constraint, including the process model and constraints in states and input. The MHE is provided with an initial state for all states (V, X, S, CO_2). Measurements (V, X, CO_2) are given to both arrival and measurement error cost, while the measurement error covariance matrix is calculated at every iteration, to weight for the model noise. The MHE optimizes the objective at every time step resulting in a state estimate, and the measurement error covariance matrix, Q_k to weight for the model noise. F_{in}, provides the broth with sugar media and F_{out} removes broth from the system. Sugar is measured at-line as a refererence, with a 40-minute delay.

For the MHE the input vector is given by,

$$u(t) = \left[\begin{array}{cc} F_{in}(t) & F_{out}(t) \end{array}\right] \tag{3.18}$$

The outputs from the system are,

$$y = h(x) \tag{3.19}$$

Where the measured outputs, for the MHE input vector, is given by,

$$y = \left[\begin{array}{ccc} V & X & CO_2 \end{array}\right] \tag{3.20}$$

meaning we are blind to sugar measurements. Volume is indirectly measured as it is calculated based on the system flows (see Appendix B.3). The dynamic model is given in a continuous time form and discretized by orthogonal collocation on finite elements to allow for an explicit relationship for the current state x_{k+1} based on the past states x_k and inputs u_k. The discretized system becomes,

$$x_{k+1} = F(x_k, u_k) + w_k \tag{3.21}$$

$$y_k = h(x_k) \tag{3.22}$$

where k denotes the sampling time and w_k is the process noise random variable related to Q_k.

The objective of the MHE problem is to find the states, V, X, S, CO_2, and their noise by minimizing the least squares problem. From equation (3.23) the decision variables are x_i and w_i. The MHE problem is formulated as follows,

$$\begin{aligned} \min_{x_i, w_i} & \quad \left(\|\hat{x}_L - x_L\|_{P_L}^2 + \sum_{i=L}^{N} \|y_i - h(x_i)\|_V^2 + \sum_{i=L}^{N-1} \|w_i\|_{W_k}^2 \right) \\ \text{s.t.} & \quad x_{i+1} = F(x_i, u_i) + w_i \qquad i = L, \cdots, N-1 \\ & \quad x_{min} \leq x_i \leq x_{max} \qquad i = L, \cdots, N \end{aligned} \tag{3.23}$$

where L is the start of the horizon and N is the last point in the horizon. The arrival cost, as mentioned in the theory (section 3.3.1) compensates for not including the previous measurements, and the process noise cost includes the process uncertainty and noise from L to N-1. The states are constrained to have positive values with $x_{min} = [0, 0, 0, 0]$ as a lower bound, to avoid negative and therefore unfeasible concentrations. The estimated

states x_{i+1} are required to follow the system dynamics, given as an equality constraint by $F(x, u) + w$.

Q_w is the covariance matrix given as,

$$Q_w = diag \left[\begin{array}{cccccccc} \omega_{\mu_{max}} & \omega_{K_s} & \omega_{k_d} & \omega_{Y_{XS}} & \omega_{Y_{XCO_2}} & \omega_V & \omega_X & \omega_S & \omega_{CO_2} \end{array} \right] \quad (3.24)$$

where ω_i for the parameters is set to the variance (σ_i^2) found through parameter estimation.

The measurement noise covariance matrix is given as,

$$R = diag \left[\begin{array}{ccc} R_V & R_X & R_{CO_2} \end{array} \right] \quad (3.25)$$

The initial error covariance is given as,

$$P_0 = diag \left[\begin{array}{cccc} P_{0,V} & P_{0,X} & P_{0,S} & P_{0,CO_2} \end{array} \right] \quad (3.26)$$

which according to Schneider and Georgakis (2013) be found by,

$$P_0 = diag((\hat{x} - x_0)^T (\hat{x} - x_0)) \quad (3.27)$$

Chapter 4

Bioprocess

The first section of this chapter provides an introduction to bioprocesses, aiming to familiarize the concepts even for readers without a background in biotechnology. Further, a detailed description of the experimental setup will be presented, providing readers with a comprehensive overview of the components and steps necessary for running a bioreactor, along with their respective significance in the bioprocess. For more in-depth information, readers can refer to Appendix B for the detailed procedures for running the bioreactor and medium protocols.

4.1 Introduction to bioprocesses

Bioprocesses are used to produce a wide range of products, including biopharmaceuticals, food ingredients, and biofuels (Kadner and Rogers, 2023). Bioprocesses are biochemical processes in which microorganisms consume carbon sources, such as sugar, to produce bulk chemicals like amino acids, lipids, and proteins. The major physical and chemical factors that affect microbial growth are temperature, moisture, pH levels, oxygen levels and nutrient availability. Microorganisms require an available source of chemical nutrients, including an energy source, a nitrogen source, as well as minerals like phosphate and trace elements, and nutrient growth factors (KGaA, 2023b). Trace elements are chemical elements required by living organisms in minute amounts, while growth factors are organic compounds that the organism cannot synthesize itself (Britannica, 1998).

Microorganisms grow in a medium, which is a liquid or solid nutrient mixture that contains all of the nutrients required for a microorganism to grow. Rich media and minimal media are two types of culture media used for growing microorganisms in the laboratory.

Rich media refer to nutrient-rich culture media that comprise a diverse blend of organic and inorganic compounds, including peptone, yeast extract, and glucose. These media, which include amino acids, vitamins, and minerals, collectively offer an abundant supply of nutrients, fostering optimal growth conditions. Examples of rich media include Luria-Bertani (LB) medium, Tryptic Soy Broth (TSB) and 2xYT. In contrast, minimal media contain only a minimal set of nutrients, typically a single carbon source, nitrogen source, and inorganic salts. Minimal media are used for specific purposes such as to grow wild-type microorganisms or to select for or against the growth of specific microbes.

The microbial growth cycle for a batch culture consists of four stages: lag, exponential, stationary, and death phases, as shown in Figure 4.1 (Madigan et al., 2018a). The biomass growth rate, measured by cell dry weight (CDW) or optical density (OD), indicates which growth phase a culture is in.
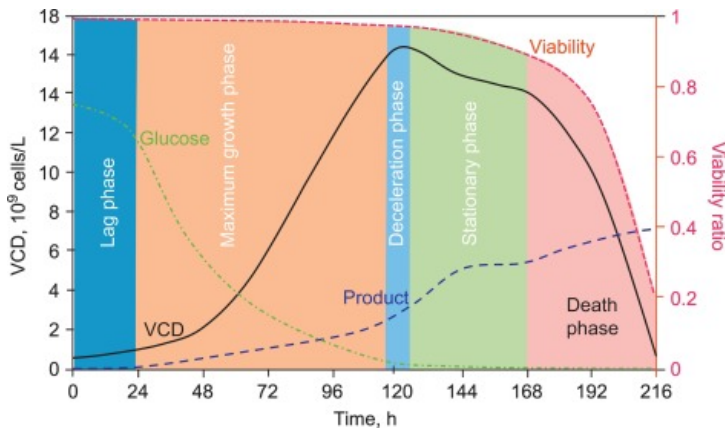
Figure 4.1: A typical growth curve for a bacterial population. The figure shows that the optical density (turbidity) increases with the increase in cell number. A viable count measures the cells in the culture capable of reproducing. The figure is reprinted with permission from (Liu, 2017).

In general, cell cultivation for bioprocesses can be carried out in three modes of operation: Batch, Fed-Batch, and Continuous, where Batch is the most common. The difference lies in the use of input and outputs, and the occurrence of feeding. Several feedings are given in a Fed-batch, whereas in continuous cultivation, the sterilized medium is introduced at a consistent rate, while the broth is removed at the same pace. Continuous bioreactors, such as chemostats, provide precise and controlled process conditions, and constant nutrient supply, and can achieve steady-state conditions resulting in more stable and predictable bioprocess performance (Graf et al., 2019). While continuous cultivation may entail higher costs (Graf et al., 2019) and a higher risk of contamination (HT, 2023), the

operation mode can be justified by its potential for process control, less manual labour (compared to consecutive batch processes), less waste, and continuous product production (Raftery et al., 2017). Moreover, as discussed by Kyslík and Prokop (2010), continuous cultivation has emerged as a potent research tool, demonstrating its potential for advancing new technologies and serving as a production processing mode for various bioprocessing applications.

Effective mixing in tanks plays a vital role in aerobic growth by preventing oxygen from inhibiting the growth process. Good mixing entails setting the stirrer at an appropriate speed, typically around 200 rpm, to ensure uniform dispersion of nutrients and transfer of oxygen from the gas phase to the liquid phase. The oxygen level in the liquid can be monitored through pO_2. Care should be exercised when using antifoam agents, as they can impede efficient oxygen transfer by altering surface tension. This may lead to the collapse of gas bubbles within the bioreactor, reducing the available surface area for gas exchange (HT, 2023).

4.1.1 *Corynebacterium glutamicum*
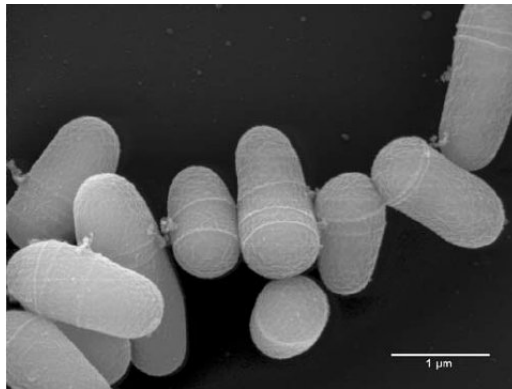
Figure 4.2: Raster electron micrograph of *Corynebacterium glutamicum* ATCC 13032. Figure reprinted with permission from Wittmann and Becker (2007).

Corynebacterium glutamicum (*C. glutamicum*) ATCC 13032, see Figure 4.2, is the wild-type microorganism investigated in this thesis. *C. glutamicum* is a gram-positive, aerobic, rod-shaped bacteria, that serves as an industrial workhorse due to its capabilities in producing various amino acids, such as L-glutamate and L-lysine, which are essential components of many food and feed products (Kalinowski et al., 2003; Lee et al., 2016; Wendisch et al., 2016). The optimal growth conditions for *C. glutamicum* include aerobic conditions, a neutral pH (7), temperatures ranging from 25-37°C, absence of light, and

the presence of preferred carbon sources such as glucose, fructose, or sucrose (Eggeling and Bott, 2005). Remarkably, *C. glutamicum* has exhibited robust growth under fluctuating conditions, including varying oxygen and carbon dioxide levels, as well as pH fluctuations. This adaptability makes it well-suited for industrial-scale fermentations that involve dynamic environmental conditions (Bäumchen et al., 2007; Follmann et al., 2009; Nishimura et al., 2007).

Indeed *C. glutamicum* has a wide range of physiological properties that make it a versatile workhorse. These properties include being generally recognized as safe (GRAS) for human use, fast growth to high cell densities, genetic stability due to the absence of a recombination repair system, and a limited restriction-modification system (defence against foreign DNA) (Gopinath and Nampoothiri, 2014; Hartbrich et al., 2000; Vertès et al., 1993). Furthermore, *C. glutamicum* exhibits no autolysis and maintains metabolic activity even under growth-arrested conditions, which is advantageous for industrial processes (Inui et al., 2004). *C. glutamicum* exhibits a broad spectrum of carbon utilization, including pentoses, hexoses, and alternative carbon sources. It also demonstrates stress tolerance to different carbon sources, further enhancing its adaptability in diverse environments (Becker et al., 2016; Kawaguchi et al., 2006; Sasaki et al., 2008). While *C. glutamicum* possesses numerous advantageous traits, its growth rate is relatively lower compared to some other bacteria, which can impact the overall productivity of fermentation processes (Graf et al., 2019).

4.2   Experimental setup

4.2.1   Method

Figure 4.3 shows the experimental setup of the continous cultivation experiment. To prepare the inoculum, the C.*glutamicum* wildtype ATCC13032 strain was streaked on an agar plate and incubated at 30°C overnight. Subsequently, a preculture was established by selecting a single colony from the plate and inoculating it into 5 mL of 2xYT broth. The preculture was then incubated overnight at 30°C and 200 rpm. From the preculture, 1 mL was transferred to a shake flask containing 200 mL of 2xYT broth, and this process was repeated twice. The shake flask cultures were incubated overnight under the same conditions. After incubation, the optical density (OD600) of the shake flask culture was measured, and the necessary amount of inoculum for the reactor, which had an initial
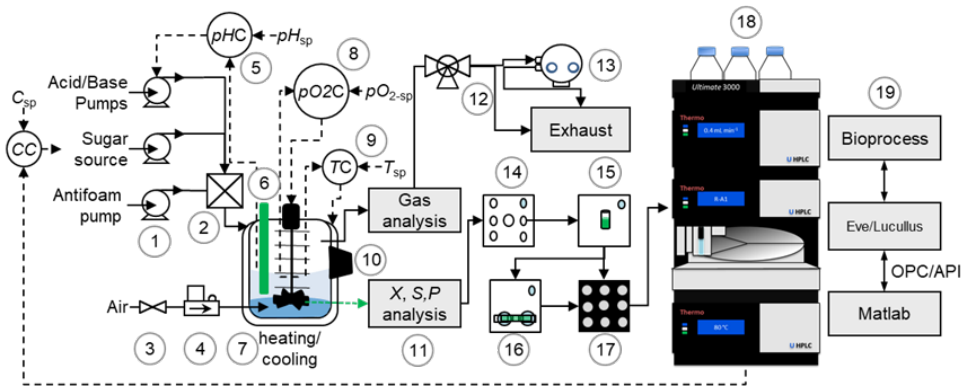
Figure 4.3: Scheme of the laboratory set-up for the microbial cell culture experiments. The system has 1: a set of acid/base, sugar source and antifoam pumps, 2: an inlet liquid flow at the top of the bioreactor, 3: air valve, 4: air mass flow controller, 5: pH Controller (pHC) with a defined set-point (pHsp) that trigger the acid/base pumps, 6: in-situ probes including near-infrared (NIR) measurement with an Optek OD probe or the Back Scaterring Buglab probe 7: heating and cooling jacket, 8: pO2 controller that has a defined set-point (pO2-sp) that controls the stirring of the Rushton type impeller, 9: temperature controller (TC) with a defined Temperature set-point (Tsp), 10: in-situ non-invasive measurement devices, gas multiplexer, 13: blue in one sensor that monitors the O2 and CO2 concentration in the gas phase. The down process consists of a Numera system that takes samples from the liquid phase and it consists of a 14: liquid multiplexer that samples from the different bioreactors, 15: a dilution module, 16: filtration module, 17: an auto-sampler that preserves the analytes into a vial. The filtered sample can also be at-line monitored in the 18: high-performance liquid chromatography (HPLC). The bioreactor setup can be monitored with 19: Eve/Lucullus and the information can be transferred to Matlab via OPC and API. Figure reprinted with permission from Dr.Pedro Antonio Lira.

working volume of 1.5 L, was calculated using the following formula:

$$1500\text{mL} \cdot \text{OD1} = X\,\text{mL} \cdot \text{measured OD} \tag{4.1}$$

The cells were then harvested by centrifugation, the supernatant was discarded, and the cell pellet was washed twice with CGXII medium before being resuspended in 75 mL of CGXII. The inoculum along with 300 mL of a glucose solution (15 g/L in broth), 1 mL of a trace element solution, and 1 mL of a biotin solution were then added to the reactor, already containing 1125mL CGXII, to constitute the reactor broth.

The 2xYT used consists per liter of : 16 g Tryptone, 10 g Yeast Extract and 5 g NaCl. The composition of the feeding solution for the reactor, or enriched CGXII per litre is as follows: 200 g glucose, 50 g $(NH_4)_2SO_4$, 5 gurea, 1 g KH_2PO_4, 1 g K_2HPO_4 0.01325 g $CaCl_2\,x\,2\,H_2O$, 0.25g $MgSO_4\,x\,7\,H_2O$, 1 mL of biotin solution (0.2 g/L), 1 mL of a trace element solution (consisting of 16.4 g/L $FeSO_4\,x\,7\,H_2O$, 10 g/L $MnSO_4\,x\,H_2O$), 1 g/L

ZnSO$_4$ x 7 H$_2$O, 0.31 g/L CuSO$_4$, and 0.02 g/L NiCl$_2$ x 6 H$_2$O. The CGXII medium used for washing has no glucose, biotin, or trace elements and a lower amount of (NH$_4$)$_2$SO$_4$ at 20 g/L. The reactor was prepared and autoclaved according to the protocol in Appendix B with the latter CGXII medium of 1125 mL.
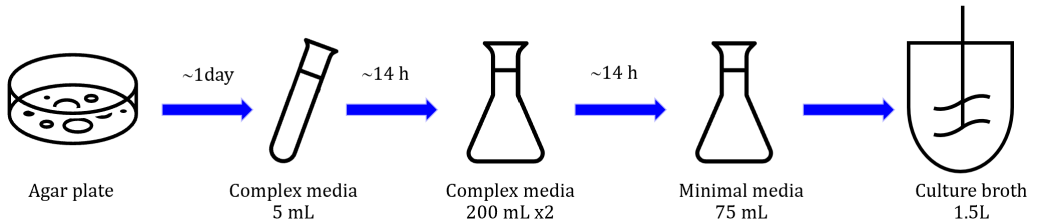
Figure 4.4: Growth media preparation steps. Cells are streaked on an agar plate and incubated to develop into colonies. A single colony is picked and inoculated in increasing volumes of complex meadia to grow a pre-culture. The necessary amount preculture is collected, centrifuged, washed and resuspended in minimal media to be used as inoculum for the culture broth.

The cultivation was performed in a 2.7L baffled stirred tank reactor, called Labfors5 (Infors AG, Switzerland). The reactors were equipped with two six-bladed Rushton impellers, with a distance from the bottom of the reactor of 6 cm and 12 cm. The feed consisted of enriched CGXII with 200 g/L of glucose. An absorbance probe for biomass monitoring was used to measure biomass in terms of CDW in the culture broth and an infrared off-gas analyzer for offgas composition, details are in Section 4.2.2. Dissolved oxygen was controlled above 30% by stirrer speed (200–1100 rpm), while the reactor was kept at 1 bar and aerated with 2 NL/L (normal air liter per minute) pressurized air. The temperature was kept at 30 °C and the pH was maintained at 7 by the addition of KOH (4M) and H$_3$PO$_4$ (10%). Offline CDW samples were taken using a supersafe sampler from Infors HT. Antifoam was added manually as necessary, as no ports were available for automatic antifoam dispensing.

4.2.2   Signal processing

Signals were collected every 60s from measurement devices and processed through the Process Information Management System Lucullus (Securecell, Switzerland) taking signals from both the Labfors 5 reactor (Infors HT), and HPLC as seen in Figure 4.5. The volume is an indirect measurement (see Appendix B.3) based on signals from the inlet and outlet weighing scale. The off-gas analyzer is a non-invasive infrared (IR) measurement device (BlueInOne Ferm, Blue-Sens GmbH) that measures the concentration of CO$_2$ in the outflow in a range between 0%–25%. The offgas calculations can be found in Ap-

pendix B.8. Signals from the absorbance probe (wavelength range 840–910 nm) were converted to CDW by using the calibration curves seen in Figure 5.2. The absorbance probe is an invasive near-infrared (NIR) probe (ASD12-N Absorption Probe, Optek GmbH) that measures absorbance in the culture broth in a range of concentration units (CU). The range of the absorbance probe is listed as 0-4 CU, however, during experimental runs, a saturation was repeatedly experienced around 1.4 CU. The calibration curve from CU to cell dry weight (CDW) is given in Chapter 5. If the NIR measurement is lower than 0.0136, CDW is set to zero to avoid negative values. This is also done for very high (> 3) CU values.
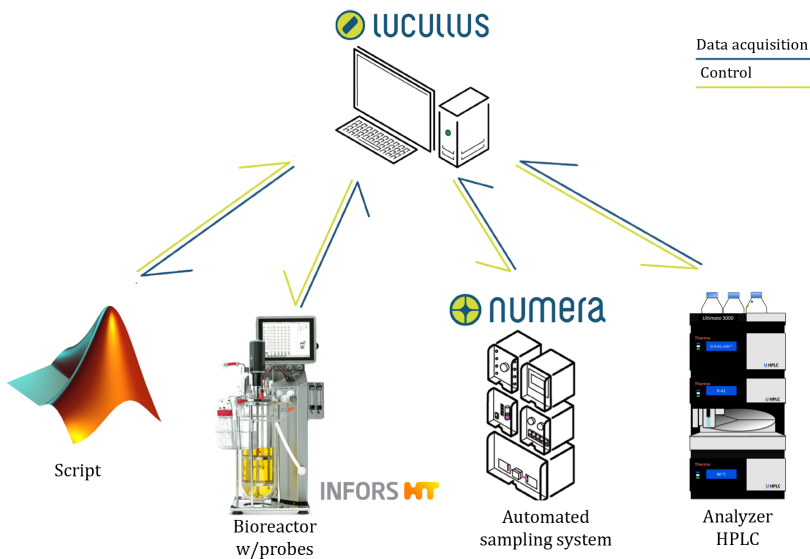
Figure 4.5: The figure shows the signal processing during an experimental run. Lucullus is a bioprocess software that integrates all elements of a bioprocess workflow into an intelligent suite on a fully digital platform. Lucullus can provide real-time data from the INFORS reactor and pass it to the Matlab script. The script can be used to control and monitor various aspects of the bioprocess. Numera is an advanced bioprocess sampling solution, automated samplings can be scheduled from Lucullus. In the same way, analyses through HPLC can be scheduled from Lucullus in combination with the Numera samplings.

4.3   Calibration

Figure 4.3 shows the many measurement devices and instruments used to monitor and control the bioreactor and process. To ensure accuracy all of these instruments must be calibrated. For this thesis, a calibration of the OD probe and pump was performed. A calibration for the acid and base pumps was also performed, in addition to finding a relation

between the laboratory spectrophotometer (OD600) and cell-dry weight (CDW), seen in Appendices B.5 and B.7. Acid and base addition was observed insignificant for volume changes and therefore not included in the model, and this section.

4.3.1 Peristaltic pumps

The pumps used in the laboratory are peristaltic pumps. They work by compressing a flexible tube to move fluid through the pump, creating flow. The setpoint of the feed is given in a flow rate (L/h), so a conversion to pump unit is necessary. The flow rate can be calculated by dividing the volume of fluid being pumped per unit time. The unit of pump rate in the lab is expressed in terms of percentage and can operate somewhat above 100 %. The most important advantage of using peristaltic pumps in biological processes is that the pump ensures the liquid is sterile. Other advantages are bidirectional flow and easy maintenance (Pump, 2022).

The setup for the calibration was one large container filled with tap water (assuming density equal to 1 g/cm^3) placed on a scale and connected to a second container through a silicon tube that passed through the peristaltic pump. The pump percentage was varied between 2 and 100 and the flow was measured indirectly by logging the weight change after 60s. This was repeated four times. The result of the calibration is seen in Chapter 5, with the corresponding regression curves.

4.3.2 OD probe

The signals obtained from the absorbance probe, were as mentioned operating within a wavelength range of 840–910 nm, and reported in concentration units (CU). In order to interpret these values in a meaningful unit, such as cell dry weight (CDW) measured in g/L, a calibration curve was required. In addition, a calibration curve was also made between the OD600 measurements of the Thermo Scientific Genesys 10S UV-Vis Spectrophotometer and CDW. The results of the latter are shown in Appendix B.7.

To obtain the experimental data a bioreactor experiment was performed, in the same method as described above in Section 4.2.1 for two reactors, but with sugar feeding of 200 g/L and an initial sugar concentration of 15 g/L. The feeding profile was chosen arbitrarily (see Appendix B.6) to ensure sufficient sugar for growth. Samples of the reactor broth were taken in triplicates every 1.5 h through a super safe sampler, for a duration of 14 hours. At every sampling the OD600 was measured (at high OD 1:10 or even 1:100 dilutions with CGXII were made), a 9 mL sample was centrifuged and washed with CGXII,

then resuspended to 9 mL, and 3 mL x 3 filtered under vacuum onto a preweighed 0.22µm MCE membrane filter. The filter cake was then dried for 72 hours at 70 °C and weighed again.

The regression was performed on the means of the CDW values at each sample time, as the samples were dependent. The result, the sinusoidal regression curve, is shown in Equation (5.3) and Figure 5.3. A linear regression was also performed, which can be seen in Appendix B.7 and figure 5.3 where each growth region has it's own curve. The corresponding linear equations are given in Equations (B.3) and (B.4). Although the R^2 and RMSE are quite similar for both the linear and sinusoidal fit, the advantage of a sinusoidal curve is that we do not have to "switch" between linear curves at specific NIR values.

4.4   Evaluation of experimental methods

In this section, we will perform a qualitative evaluation of the experimental methods. The purpose is to show transparency and what contributed to uncertainty in the results.

One of the main sources of uncertainty when working with living organisms is the potential for mutations and variability within a population. Despite our efforts to select a single colony, mutations (transcription error 1 per every 100,000 nucleotides (Pray, 2008)) can occur during cultivation, potentially leading to unexpected dynamics. However, since the cultivation time is relatively short and we are working with a wild-type (not GMO) under neutral circumstances this is unlikely to have an effect. Another source of uncertainty is contamination, which may occur after autoclaving and filtering, but the effect of this on the cultivation is seen as low.

Regarding the preparation of media and inoculum, we used two different weighing instruments with different accuracies (Sartorius ENTRIS64- 1S Analytical Balance   0.0000 g and VWR LP-1002 precicion balance 0.00 g) to ensure precise measurements. However, slight inaccuracies in weighing the compounds should not have affected growth behaviour since we added the media compounds (see Section 4.2.1) in such an amount that only the substrate should be the limiting growth factor. The accuracy of adjustable micropipettes is affected by factors such as the viscosity of the fluid, the angle at which the micropipette is held, and the user's skill in operating the micropipette. These factors became apparent when taking CDW samples, as more liquid remained in the test tube of 9 mL as the CDW increased, despite filtering 3x3 mL each time.

The offgas analysis in the BlueInOne can measure CO_2 in the range of 0 - 25 Vol.%. To measure accurately it must be heated and calibrated with the air used during the experiments. The volume is indirectly measured by measuring the weight difference in the inlet and outlet bottle over a timestep (see details in Appendix B.3). These weights were affected by vibrations in adjacent equipment.

The OD measurements are dependent on correct NIR-CDW calibration. The NIR measurements in themselves have no physical meaning, other than showing us the changes in density, so the calibration is necessary to have a quantitative measure in the form of CDW. The OD probe must be calibrated to 0 (run on the Infors reactor) before adding the inoculum for the correct reading.

Sugar measurements were done at-line in the HPLC, which was calibrated and heated before experimental work by authorized lab personnel. There may arise leaks in the HPLC tubing and column, resulting in faulty sugar measurements. This is not a problem for the MHE, as the samples may be rerun after the experiment (this was the case for the parameter estimation and run 1), but this is an issue if proceeding with multirate MHE. In addition, the amount of sugar measured is highly dependent on how the program is calibrated.

During the run pH and the pO2 level is adjusted automatically through setpoint control by the INFORS system. The pH adjustment contributes to some volume changes, but these were registered as insignificant compared to the total volume. The pO2 level is adjusted by the stirring rate. If the setpoint is set too high the stirring speed may disturb other measurements in the reactor. If the setpoint is set too low the oxygen becomes a limiting factor for bacteria growth. In Figures B.7, B.8a and B.8b the pO2 level for the runs can be seen, where manual adjustments to the setpoints are made to reduce disturbance and to provide a sufficient oxygen level. Antifoam was added manually during the run. From Figure 5.11 we see that a high level of antifoam may disturb the probes, here seen as an unnatural spike in the OD measurement around 3 hours. Too much antifoam will also affect the surface tension, and at its worst decrease the rate of oxygen transfer from the gas to the liquid phase.

Finally, the reproducibility of the results is dependent on instruments and calibrations at other laboratories, but most of all the methods used in the preparation of the pre-culture. Therefore, it is important to consider the potential sources of uncertainty when interpreting the results obtained in this study.

Chapter 5

Results

The aim of this thesis, presented in Section 1.2, was to investigate whether Moving Horizon Estimation (MHE) can address the challenge of limited sugar measurement in bioprocesses, allowing for reliable real-time state estimation in a continuous cultivation of *Corynebacterium glutamicum* (*C. glutamicum*). To accomplish this, the real-time performance of MHE as a sugar estimator was tested in a *C. glutamicum* continuous cultivation process with only three available measurements.

The results of the two real-time experiments can be seen in Figures 5.9 and 5.11. To achieve the intended input profile for continuous cultivation, shown in Figure 5.4, Model Predictive Control (MPC) was used on the nominal model given in Equation (2.1) using the tuning parameters provided in Table 5.1, and model parameters from a Fed-Batch cultivation of *C. glutamicum* from Tuveri et al. (2021).

In order to ensure the proper functioning of the MHE, it was necessary to tune it to the specific process investigated, the continuous cultivation of *C. glutamicum*. This involved performing parameter estimation, described in Section 2.2.1, on the experimental data obtained by following the method outlined in Section 4.2.1, using the input profile obtained from MPC (Figure 5.4). The resulting parameters and variances for the MHE and covariance matrix (Q_w) are presented in Figure 5.6 and Table 5.2. Further, the MHE was tuned using R, Q_w, and P_0, presented in Table 5.5, and the resulting performance of the MHE tuning is shown in Figure 5.8.

The overall performance of the MHE, quantified by Root Mean Square Error (RMSE), is presented in Table 5.4. The acquisition of experimental data involved running a bioreactor, which involved extensive laboratory work, including calibration. The peristaltic pumps

responsible for inlet and outlet flows were calibrated, and the calibration results are shown in Figure 5.1. The Optical Density (OD) probe was also calibrated in order to measure values from the bioreactor in terms of Cell Dry Weight (CDW) [g/L]. The calibration results, represented by a sinusoidal curve, are presented in Figures 5.2 and 5.3.

The two real-time MHE runs encountered some issues, such as incorrect initial conditions and disturbances in the OD measurements caused by antifoam. To assess the impact of tuning on the estimations, the MHE was retuned offline by varying R and P_0 as shown in Figures 5.13 and 5.15. Additionally, to observe the calibration curve's effect on the biomass and sugar estimates, the MHE was rerun offline using a linear calibration curve for the OD measurements.

The results for each section are presented below and further discussed in Chapter 6 in relation to the initial objective outlined in Section 1.2. It should be noted that while all red crosses in the figures are labeled as "offline", the CDW samples are obtained offline, while the sugar measurements are obtained at-line.

5.1   Calibration

Figure 5.1 shows the experimental data and calibration curve for the two peristaltic pumps in the laboratory used for inlet (Figure 5.1a) and outlet (Figure 5.1b) flow. The coefficient of determination (R^2) is 0.99988 for pump 1, and 0.99922 for pump 2. The R^2 value indicates that the linear fit can explain nearly 100% of the variation in pump flow. The average deviation of the observations from the regression, the RMSE, is 0.24362 for pump 1, and 0.87418 for pump 2. The largest deviations are observed at high flow rates, however, this does not affect the subsequent experiments since only flow rates below 1 L/h are utilized. The graphic results are shown in figure 5.1, while the calibration equations are given below,

$$\text{Pump \%}_1 = 45.2176 \cdot \text{Flowrate} \tag{5.1}$$

$$\text{Pump \%}_2 = 44.7667 \cdot \text{Flowrate} \tag{5.2}$$

(a) Calibration pump 1   **(b)** Calibration pump 2
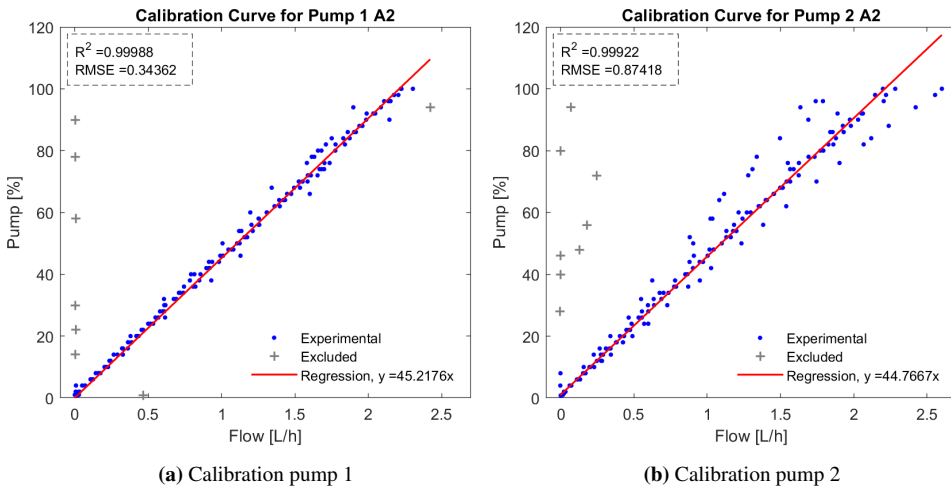
Figure 5.1: Calibration of the peristaltic pump in the lab. Experimental data are seen as blue dots showing the measured flow rate at a given pump percentage. Grey '+' markers show outliers that have been excluded from the regression. The red curve is the linear regression curve. The R^2 tells us how many of the variations in data points (pump flow) the linear fit, y, is able to explain. The root-mean-square error (RMSE) is the average deviation of the observations from the regression.

Calibration of the OD probe was performed by running a bioreactor according to Section 4.3.2 for 20 hours and taking CDW samples every 2-3 hours. The calibration curve from NIR ([CU]) to CDW ([g/L]) was obtained by regression to a sum of sine functions from experimental data. This method of obtaining the curve is further elaborated on in Section 4.3.2. The sinusoidal regression curve found was,

$$\text{CDW} = -15212.545 \cdot \sin(2.0532 \cdot \text{NIR} + 2.5893) + 15197.4245 \cdot \sin(2.055 \cdot \text{NIR} + 2.5887)$$

$$(5.3)$$

Figure 5.2 shows the obtained sinusoidal calibration curve for NIR to CDW, along with the experimental CDW and the corresponding OD probe (NIR) values. A linear curve was also considered (Appendix B.7), but the sinusoidal curve had a slightly better fit, so this was chosen. The coefficient of determination (R^2) is 0.98596 indicating that the sinusoidal fit is able to explain nearly 100% of the variation in CDW within the tested intervals of 0-1.2 CU for the NIR-signal and 0-25 g/L for CDW. The average deviation of the observations from the regression, the RMSE, is 1.3903. Figure 5.3 shows the conversion of NIR measurements to CDW using the calibration curve, along with manual CDW samples and NIR values from the OD probe for the 20-hour cultivation. Apart from the second to last sample, the converted NIR is in close proximity, $\pm\ 1$ g/L, to the actual CDW samples.
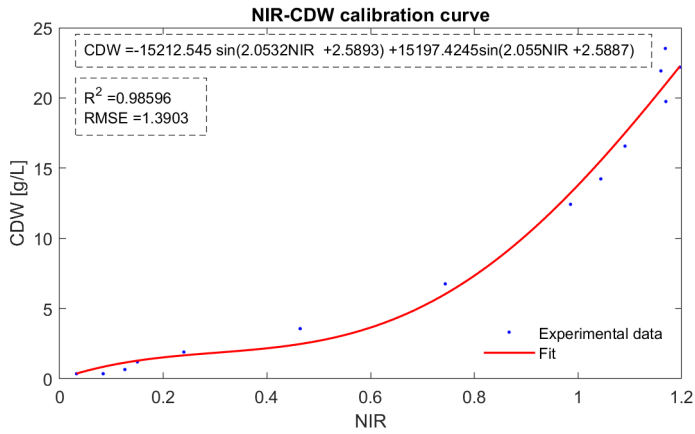
Figure 5.2: NIR-CDW calibration curve for the OD probe. Experimental measurements of NIR vs. the mean value of the triplicate CDW measurements are seen as blue dots. The red curve is the sinusoidal fit, where the CDW regression equation is given at the top of the figure. The coefficient of determination (R^2) and RMSE are in the box below the equation.

Figure 5.3: Convertion of NIR measurements to CDW, for visual evaluation of fit during the cultivation and comparison to experimental CDW. The blue curve with purple points 'o' shows the NIR converted to CDW with the calibration curve given in Equation (5.3) and at the top of the figure. The red 'x' show the experimental CDW and the orange band, the variance in the triplicates. The blue '+' are the NIR values from the OD probe at the CDW sampling time.

5.2   Optimal feeding profile

An optimal feeding profile for the 24-hour cultivation was found using in-silico MPC with setpoint-tracking of biomass at 15 g/L, given in Equation (3.6). The initial conditions used were $x_0 = [V_0, X_0, S_0, CO_{2,0}]^T 4 = [1.5, 0.8, 15, 0]^T$. To enable continuous production the flow in and out was set to be above 0.0178 L/h[1] after the batch phase. The control horizon was set to two minutes, and the prediction horizon 20 minutes, found by trial and error. The MPC tuning and system parameters are shown in Table 5.1. For the in-silico MPC simulations, model parameters from Tuveri et al. (2021) were applied (listed in Table C.1), because parameter estimation for the continuous cultivation had not been performed yet, as this depended on determining an input profile first.

The MPC suggests initiating feeding at 10 hours (purple profile) in Figure 5.4, but based on experience from previous runs (Figure B.1), the feeding was adjusted to start after 6 hours (red profile) to account for the approximate duration of the batch phase before sugar depletion. The dynamics associated with the feeding profile are shown in Figure 5.5. The increase in cost, seen in Figure 5.4 is due to divergence from the biomass setpoint (see purple X profile in Figure 5.5) due to keeping a sustained flow.

Table 5.1: Parameters for the MPC given in equation (3.6) finding an optimal feeding profile. The time step (dt) and the simulation time (T) are in hours, while the horizons, n_m and n_p, are in minutes. Q and R_1 are tuning parameters given in Equation (3.6).

Parameter	Value
n_m	4
n_p	20
T	24
dt	1/60
Q	$500 \cdot diag(0, 1, 0, 0)$
R_1	100

---

[1]0.0178 L/h is the lowest registered flow for the peristaltic pump
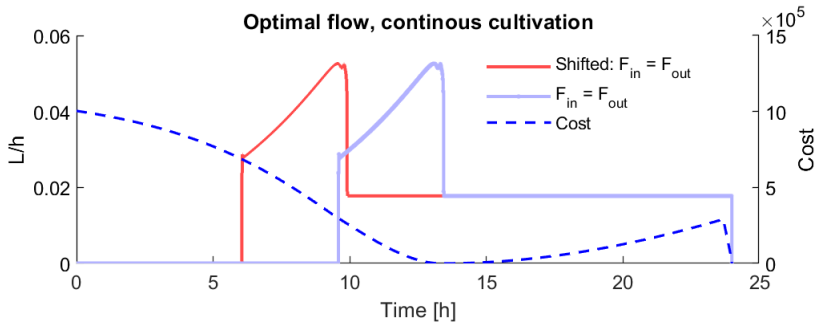
Figure 5.4: Optimal input profile (inlet, F_{in}, & outlet, F_{out}) found by MPC. Fixed (shifted) feeding and flow out are seen in red, the intended input profile for all cultivations. The MPC cost function is seen in dashed blue and is related to the original optimal flow. For the shifted input, the batch phase is 6 hours, following an increase in feed, until around 11 hours when the feed is kept constant at 0.0178 L/h. The volume is kept constant at 1.5 L by the outlet flow.
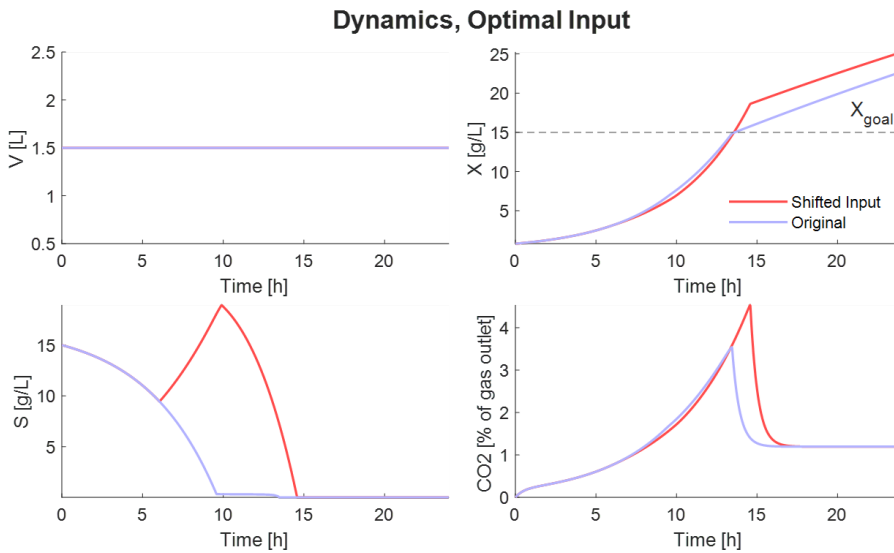
Figure 5.5: State dynamics for the optimal flow (purple) found using MPC and the shifted flow (red) related to the input profiles in Figure 5.4. The dashed black line shows the setpoint for the biomass (X_{goal}) in Equation (3.6).

5.3   Parameter estimation

Expermental data was obtained by performing a cultivation using the method in Section 4.2.1 using the feeding profile found in the previous section (Figure 5.4). Following the method in Section 2.2.1, the parameters with standard deviation in Table 5.3 were ob-

tained. 23 data points were used, as 23 sugar measurements were available, with the initial guess and bounds for the parameters given in Table 5.2. Figure 5.6 shows the experimental data and integration of the nominal model (Equation (2.1)) using the obtained parameters in Table 5.2. The feed and flow out during the cultivation are seen in Figure 5.7 where a large initial addition of feed is observed, before the intended feed is added, shifted by 3 hours. The cause of this deviation was a pause in the experiment due to code glitch. Due to pump malfunction, the outlet flow does not follow the intended profile as seen inFigure 5.7. Details on the flow anomalies are given in Appendix B.9. The initial state for parameter estimation was $x_0 = [1.5000, 0.3832, 11.8238, 0]^T$.

Table 5.2: The initial guess (θ_0) and constraints (θ_{min}, θ_{max}) for parameter estimation.

Parameter	θ_{min}	θ_0	θ_{max}
μ_{max}	0.09	0.22846	0.3
K_s	0.001	0.0077687	0.01
k_d	0.001	0.0025287	0.009
Y_{XS}	0.1	0.43004	0.5
Y_{XCO_2}	0.1	0.39328	0.9

Table 5.3: Values of the estimated model parameters in Equation (2.1) with the unit and standard deviations.

Parameter	Description	Value	Unit	Std. Dev.
μ_{max}	Maximum growth rate	0.2296	$[\text{h}^{-1}]$	$1.000 \cdot 10^{-3}$
K_S	Monod growth constant	0.0077556	$[\text{g} \cdot \text{L}^{-1}]$	$9.4749 \cdot 10^{-5}$
k_d	Death rate constant	0.0025183	$[\text{h}^{-1}]$	$4.4482 \cdot 10^{-5}$
Y_{XS}	S from X yield	0.42931	$[\text{g} \cdot \text{g}^{-1}]$	$5.0084 \cdot 10^{-3}$
Y_{XCO_2}	CO$_2$ from X yield	0.39387	$[\text{g} \cdot \text{g}^{-1}]$	$1.0611 \cdot 10^{-3}$

The dynamics in Figure 5.6 show that the volume is not held constant at 1.5 L as intended. This is caused by difference in flow in and out, seen in the input profile in Figure 5.4. The fitted data follows the measurement dynamics without being overfitted[2]. The RMSE for the parameter estimation shown in Figure 5.6 is 9.0377 for the CDW measurements and 1.4125 for sugar. The RMSE between the online biomass value and the value from the simulation with parameters from the estimation is 1.8953.

---

[2]Overfitting is the production of an analysis which corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably (Press, 2023).
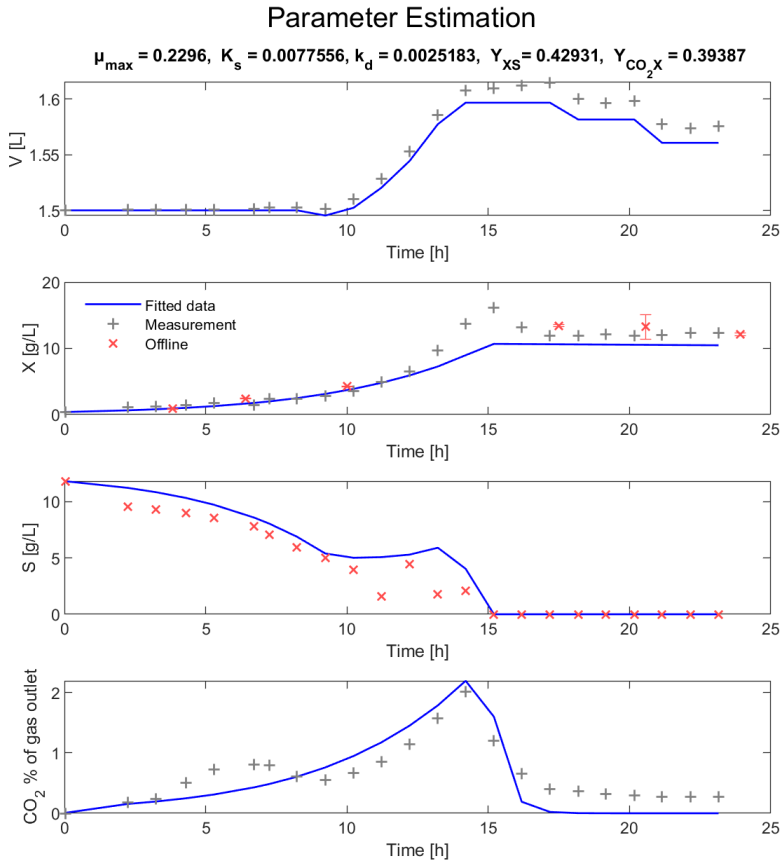
Figure 5.6: Parameter estimation on 23 data points. The header shows the parameter values. The standard deviations of the parameters are given in Table 5.3. The blue line is the fitted data, found by integration of the model using the feeding profile in Figure B.7 and estimated parameters. The grey '+' is the experimental data used corresponding to the timestamp of the at-line sugar measurements. The red 'x' are at-line measurements by HPLC and offline manual CDW samples.
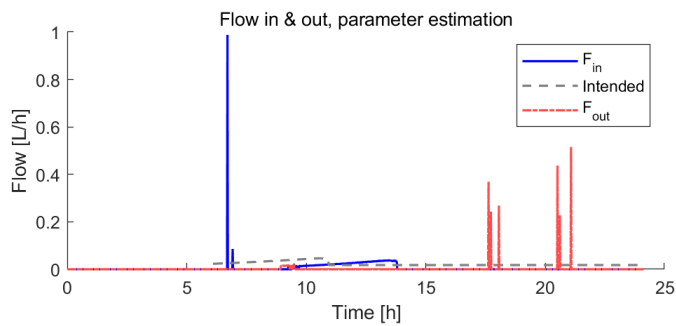
Figure 5.7: Inlet (blue line) and outlet (red line) flow of the reactor during the cultivation. The aim of the experiment was to obtain data for parameter estimation. The grey dashed line shows the intended flow in and flow out during the experimental run. The deviation from the intended feed and flow out, is discussed in Appendix B.9.

5.4   Moving horizon estimation

This section presents the three runs MHE was applied on. First, the MHE tuning parameters in Equations (3.24) to (3.26) were determined by trial and error to the values shown in Table 5.5, resulting in state estimates with the lowest possible RMSE. The RMSE for all MHEs, including parameter estimation is shown in Table 5.4. The horizon length was determined by trial and error. The tuned MHE was subsequently applied in real-time for two experiments, as described in Section 4.2.1.

Table 5.4: RMSE for all runs, including Parameter estimation, MHE tuning and the two real-time MHE runs. The calculation of the RMSE is given in equation (2.6). [1] The first real-time MHE experiment seen in Figure 5.9 and [2] the second real-time MHE experiment seen in Figure 5.11

Variable	Parameter Estimation	Tuning		Real-time[1]		Real-time[2]	
		Model	MHE	Model	MHE	Model	MHE
X	9.0377	1.2145	0.8651	0.4021	0.8626	2.1408	1.3877
S	1.4125	2.4968	0.8440	1.1233	1.6319	1.15986	2.6279

5.4.1   MHE tuning

Figure 5.8 shows the results of the MHE tuning using the obtained parameters in Table 5.2. The initial state given to the MHE was $x_0 = [1.5000, 0.3832, 11.8238, 0]^T$. The values of Q_w, R and P_0 are listed in Table 5.5. As a similar experimental setup was used as in Tuveri et al. (2021), the same P_0 values from this paper were applied instead of using Equation (3.27). The optimization horizon, N, used was 40 minutes. The horizon is equal to the current time it takes to process sugar measurements in the HPLC. For each iteration, the average runtime was 0.3469s, while the maximum was 1.2296s. From the experimental data seen in Figure 5.8 sugar is observed present in the system when feeding commences at 6 hours. After around 6 hours there are two jumps in the MHE sugar estimate which are observed to occur simultaneously with the short feeding in Figure 5.7. Both for the biomass (X) and CO_2 there is a small offset between the MHE and measurements as the system moves towards a steady state.

Table 5.5: Tuning parameters used for MHE. R, Q_ω and P_0^+ are described in Section 3.3.2. R_i and $Q_{\omega,i}$ and $P_{0,i}^+$ describe the values along the diagonal of each matrix corresponding to the measurement or parameter in the column 'Variable'.

Variable	Type	$\mathbf{R}_i$	$\mathbf{Q}_{\omega,i}$	$\mathbf{P}_{0,i}^+$
V	State/measurement	0.1	$1.0 \cdot 10^{-2}$	$2.0946 \cdot 10^{-8}$
X	State/measurement	0.1	$9.0 \cdot 10^{-2}$	$1.0975 \cdot 10^{-5}$
S	State/measurement	-	$1.0 \cdot 10^{-2}$	$1.0852 \cdot 10^{-4}$
CO_2	State/measurement	0.001	$1.0 \cdot 10^{-1}$	$2.1669 \cdot 10^{-5}$
μ_{max}	Parameter	-	$1 \cdot 10^{-8}$	-
K_s	Parameter	-	$1 \cdot 10^{-8}$	-
k_d	Parameter	-	$1.9786 \cdot 10^{-9}$	-
Y_{XS}	Parameter	-	$2.508 \cdot 10^{-5}$	-
Y_{XCO_2}	Parameter	-	$1 \cdot 10^{-8}$	-
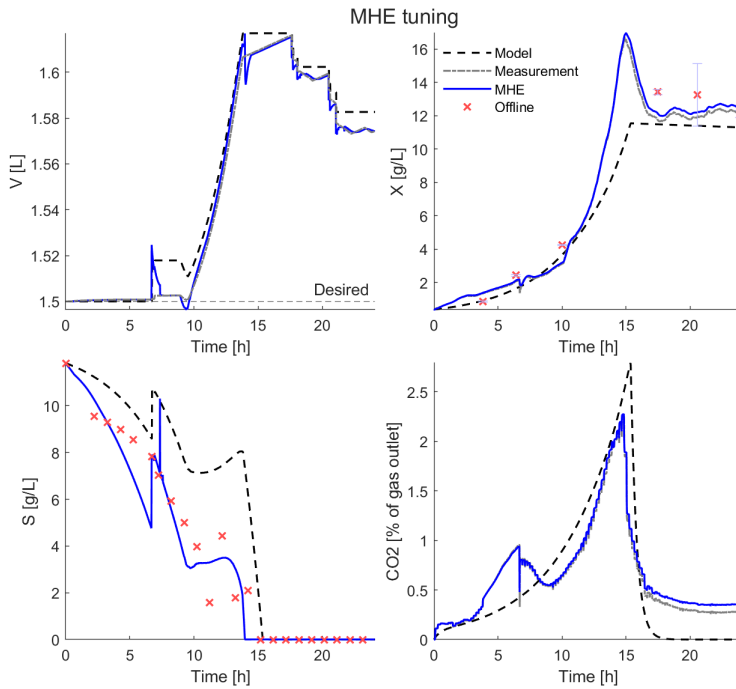
Figure 5.8: MHE tuning with estimated parameters. The black dashed line is the integration of the nominal model, the blue line is the MHE state estimate, the grey dashed line the online measurements, and the red 'x' offline CDW and at-line sugar measurements. The initial state given to the MHE was $x_0 = [1.5000, 0.3832, 11.8238, 0]^T$

.

5.4.2 Real-time MHE

The initial state given to the MHE for the first real-time MHE run was x_0 = [1.5000, 0.3798, 11.0000, 0.0006]^T. In the first real-time MHE run, it can be observed from Figure 5.9 that an incorrect initial condition for sugar was provided to the MHE. This is apparent as the MHE starts at 11 g/L, while the first HPLC measurement indicates a value just below 10 g/L. The reason for this discrepancy was most likely a leak found in the HPLC. All samples were rerun offline after the experiment, resulting in a lower initial sugar value at ∼10g/L. The MHE estimates a slightly higher sugar consumption rate than what the HPLC measurements show, while the nominal model has a slightly lower sugar consumption rate compared to the MHE. Figure 5.10 shows that the cultivation had failed feeding, so no feed was given. There is registered flow out at ∼6h, however, this was a disturbance caused by a fluctuation in the outlet scale, not actual flow.
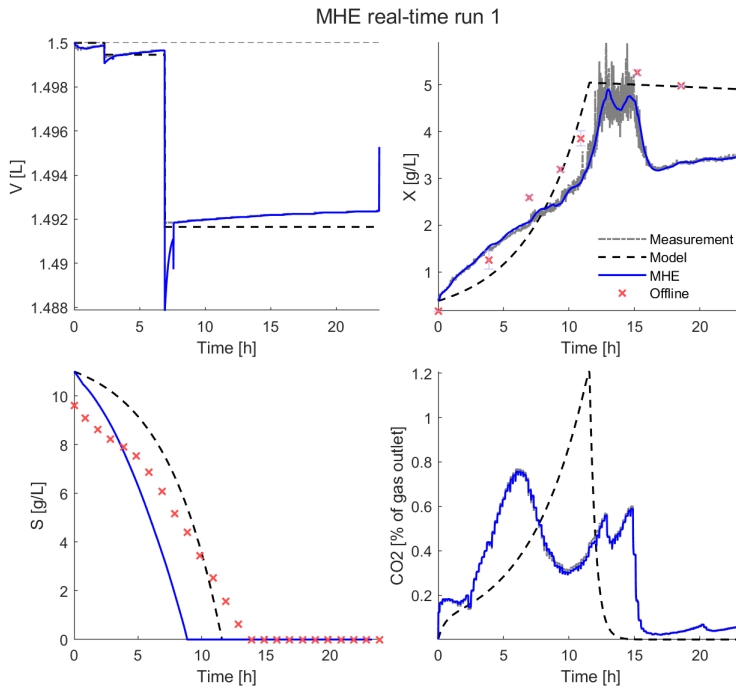
Figure 5.9: First real-time MHE experiment. The black dashed line is the integration of the nominal model, the blue line is the MHE state estimate, the grey dashed line is the online measurements, and the red 'x' offline CDW and at-line sugar measurements. The initial state given to the MHE was $x_0 = [1.5000, 0.3798, 11.0000, 0.0006]^T$. This experiment had a failed feeding, so no feed was given in the 24-hour cultivation (see Figure 5.10). The CO_2 measurements do not follow the model dynamics, as is the case for the other experimental runs.
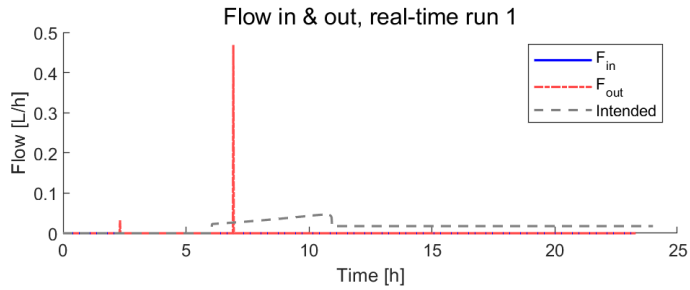
Figure 5.10: Inlet (blue line) and outlet (red line) flow of the reactor during the real-time MHE run 1. The grey dashed line shows the intended flow in and flow out during the experimental run. The deviation from the intended flow out is discussed in Appendix B.9.

The initial state given to the MHE for the second real-time MHE run was $x_0 = [1.5000,$ $0.3601, 10.0000\ 0]^T$. In Figure 5.11 the MHE estimates a significantly higher sugar consumption rate than what the HPLC measurements show. Sugar was present in the system when feeding began at 6 hours. The nominal model has a lower sugar consumption rate than the MHE, closer to the HPLC measurements. In Figure 5.12 we see that the feed profile is slightly below the intended feed. Also here there was an issue with the outlet flow resulting in manual removal of broth to keep the volume close to 1.5 L. The disturbance in biomass at 3-4 hours in the cultivation was caused by large amounts of antifoam in the reactor. The measurement dropped to its real value when antifoam was manually added ($\sim$4.5h).
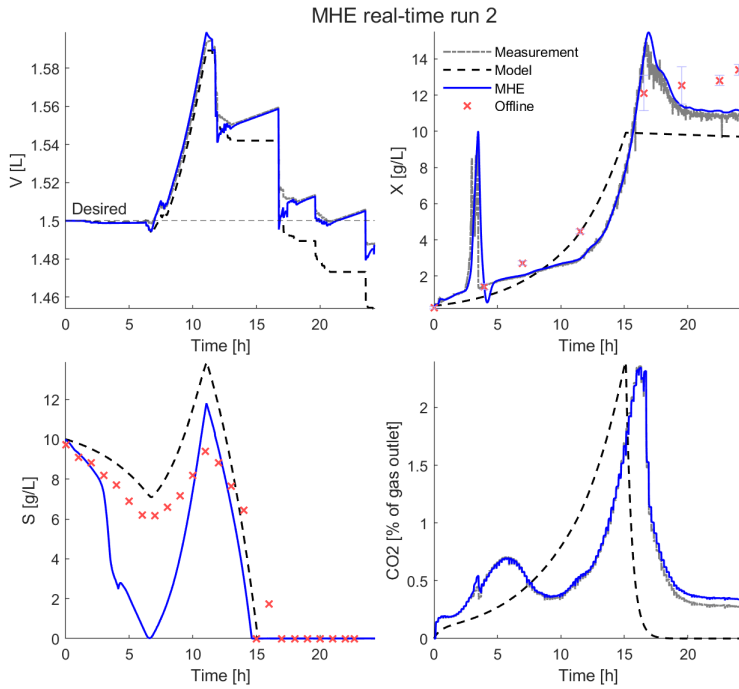
Figure 5.11: Second real-time MHE experiment. This experiment had a failed feeding, so no feed was given in the 24-hour cultivation seen in Figure B.8b. The black dashed line is the integration of the nominal model, the blue line is the MHE state estimate, the grey dashed line is the online measurements, and the red 'x' offline CDW and at-line sugar measurements. The large increase in biomass at ∼3 hours is due to the presence of foam. After adding antifoam the biomass measurement goes down to its actual value. The initial state given to the MHE was $x_0 = [1.5000, 0.3601, 10.00000]^T$.
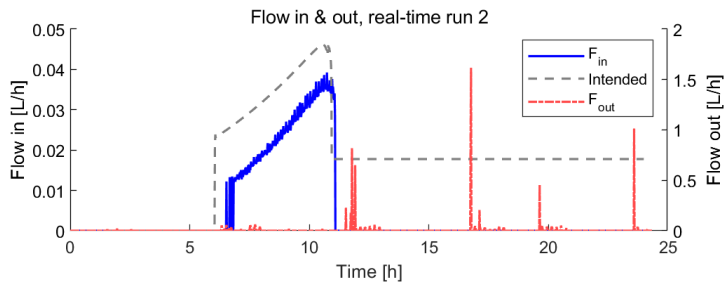
Figure 5.12: Inlet (blue line) and outlet (red line) flow of the reactor during the real-time MHE run 2. The grey dashed line shows the intended flow in and flow out during the experimental run. The deviation from the intended flow out is discussed in Appendix B.9.

5.4.3   Offline MHE with adjusted tuning

Figure 5.13 shows MHE on run 1 with an initial covariance matrix with varying value for sugar, $P_{0,S} \in$ [3, 2, 1, 1e-1, 1e-2, 1e-3, 1e-4, 1e-5, 1e-6, 1e-7, 1e-8] (see Equations (3.8) and (3.26)). A high value of P_0 means low uncertainty in the initial state, while a low value of P_0 means high uncertainty in the initial state. The purpose of rerunning the MHE with varying P_0 values was to see if the MHE could recover from inaccurate initial conditions by adjustment in the initial covariance matrix. The $P_{0,S}$ used during the real-time run was 1.0852e-4 (see Table 5.5), so values below this have increased uncertainty. The figure shows the variance in state values given different $P_{0,S}$ (transparent red), between the max (red line) and min (blue) and the average value (purple dashed). From Figure 5.14 we observe that the maximum value for sugar in Figure 5.13 is for index 8 in $P_{0,S}$ which is 1e-8, where the initial sugar is estimated to 17 g/L. So at high uncertainty, the sugar estimate is at its furthest from the true value. Regardless of the $P_{0,S}$ value, the estimate for sugar never goes below the initial value of 11 g/L. We observe that changes in the value for sugar in the initial covariance matrix have little effect on the other estimates, V, X, CO_2, visible in Figure 5.13 where the states have close to no variance.

Figure 5.15 figure shows MHE run offline on the data from the second MHE real-time run using a measurement noise covariance matrix with varying values for biomass, $R_X \in$ = [5, 2, 1, 1e-1] (see Equations (3.8) and (3.25)), and a fixed value for w_x at $9 \cdot 10^{-4}$ (see Equations (3.8) and (3.24)). The purpose of the adjustments in tuning was to see if the initial disturbance could be tuned out. A high value of R means high uncertainty in the measurements, while a low value of R means low uncertainty in the measurements. Figure 5.15 shows the variance in state values given different R_X (transparent red), between the max (red line) and min (blue) and the average value (purple dashed). From Figures 5.15 and 5.16 we see that the R_X that is able to reduce the noise in X the most at 2-4h, is index 1, corresponding to $R_X = 5$. (see bottom graph in Figure 5.16). We see that $R_X = $ 1e-4, gives the highest X estimate after 15h, which is higher than the value in Figure 5.11 using the same R. This is due to the increase in additive noise for biomass (w_x) from $9 \cdot 10^{-2}$ to $9 \cdot 10^{-4}$.
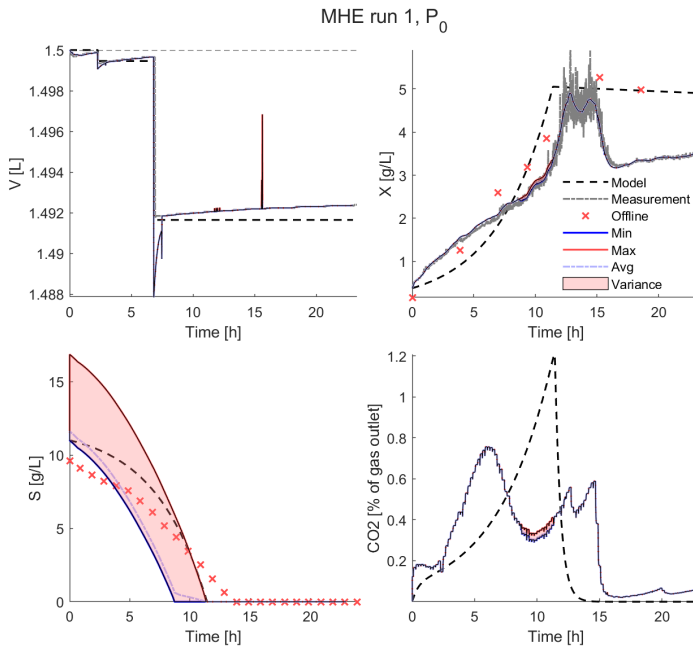
Figure 5.13: Offline MHE run on the experimental data from the first real-time run, using a initial covariance matrix with varying values for sugar, $P_{0,S} \in [3, 2, 1, 1e\text{-}1, 1e\text{-}2, 1e\text{-}3, 1e\text{-}4, 1e\text{-}5, 1e\text{-}6, 1e\text{-}7, 1e\text{-}8]$. The figure shows the variance in state values (translucent red) given different $P_{0,S}$, between the max (red line) and min (blue line) and also the average values (purple dashed). The red 'x' mark offline samples for CDW and at-line values for sugar, while the wide dashed black line is the nominal model value based on the input profile, and the dashed grey line is online measurements.
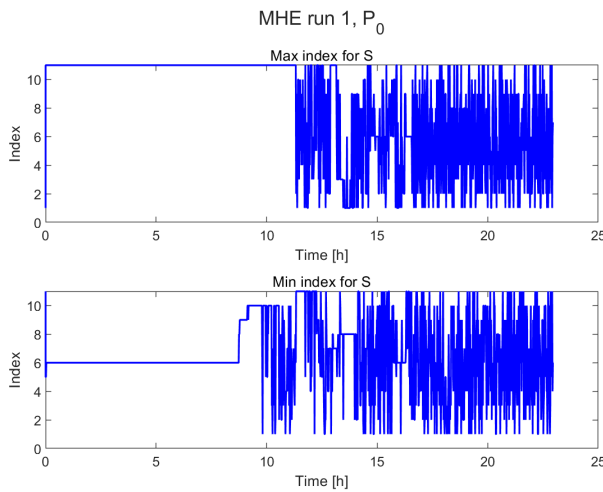
Figure 5.14: The figure shows the indices in the list, $P_{0,S} \in [3, 2, 1, 1e\text{-}1, 1e\text{-}2, 1e\text{-}3, 1e\text{-}4, 1e\text{-}5, 1e\text{-}6, 1e\text{-}7, 1e\text{-}8]$, related to the maximum and minimum values of sugar (S) in Figure 5.13.
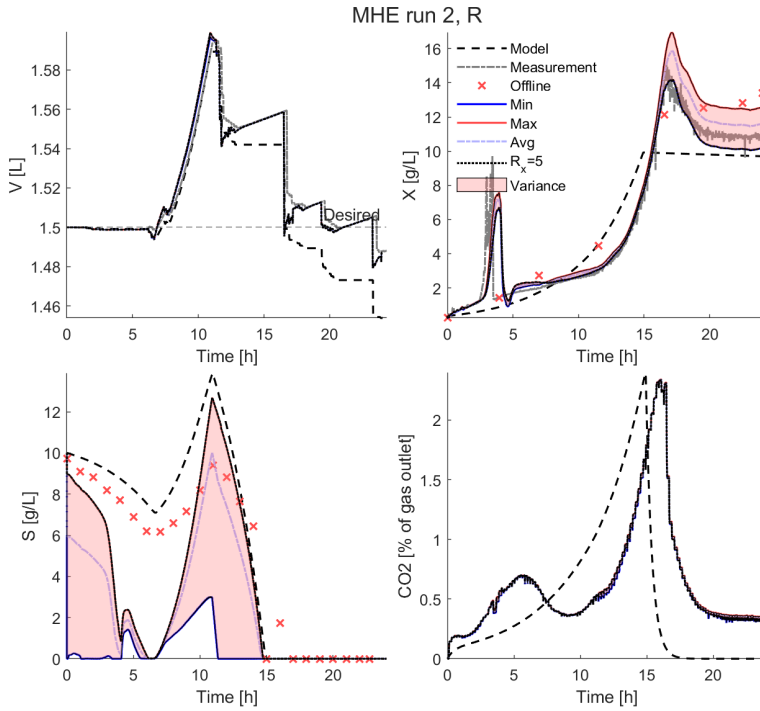
Figure 5.15: Offline MHE run on the experimental data from the second real-time run, using a measurement noise covariance matrix with varying values for biomass, $R_X \in = [5, 2, 1, 1e\text{-}1]$ and $w_x = 9 \cdot 10^{-4}$. The purpose was to see if the initial disturbance could be tuned out. The figure shows the variance in state values (translucent red) given different $P_{0,S}$, between the max (red line) and min (blue line) and also the average values (purple dashed). The red 'x' mark offline samples for CDW and at-line values for sugar, while the wide dashed black line is the nominal model value based on the input profile, and the dashed grey line is online measurements.
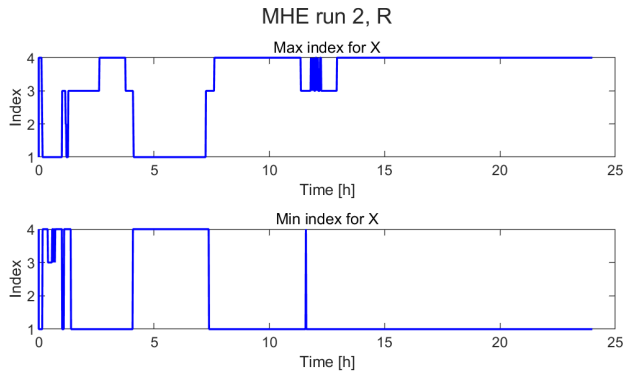
Figure 5.16: The figure shows the indices in the list, $R_X \in = [5, 2, 1, 1e\text{-}1]$, related to the maximum and minimum values of biomass (X) in Figure 5.15.

5.4.4   Offline MHE with linear OD probe calibration

To see the effect the biomass calibration has on the estimator, all experimental bioreactor runs, both MHE tuning and real-time MHE, were rerun with MHE on the experimental data, using a linear calibration (Equations (B.3) and (B.4)) of the OD probe, as opposed to the sinusoidal curve used during cultivation. The results of the adjustment are seen in Figures 5.17, 5.18a and 5.18b, and quantified in terms of RMSE in Table 5.6. The model profile for sugar in Figure 5.17 deviates from Figure 5.8. This is due to different initial conditions in biomass caused by the different calibration curves ($X = 0.2631$ for linear, $X = 0.3839$ for sinusoidal).

Table 5.6: The table shows the RMSE for all offline MHE simulations using alinear calibration curve for the OD probe. MHE was rerun on all experimental data from tuning ((Figure 5.17), and the two real-time experiments. The calculation of the RMSE is given in equation (2.6). [1] The first MHE experiment, run offline, seen in Figure 5.18a and [2] the second MHE experiment, run offline, seen in Figure 5.18b

Variable	Tuning		Real-time[1]		Real-time[2]	
	Model	MHE	Model	MHE	Model	MHE
X	1.4877	0.5945	0.4021	0.3123	2.1408	0.8872
S	5.0070	1.4666	1.1233	1.7136	1.15986	4.2531
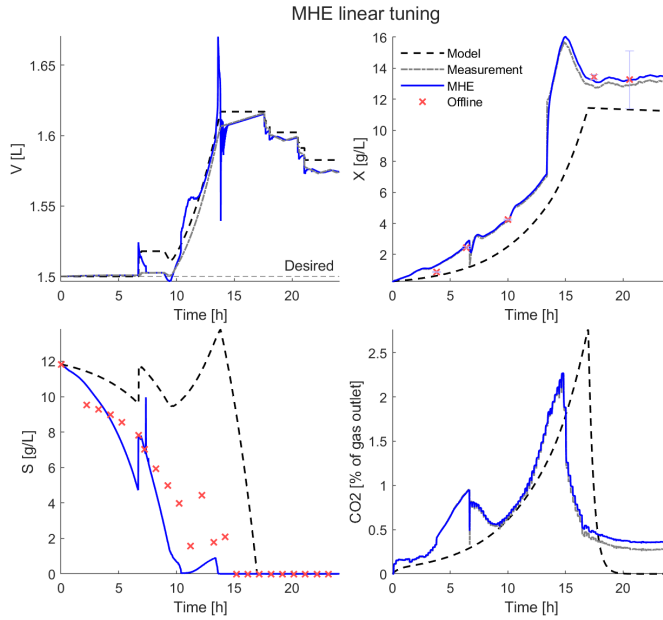
Figure 5.17: Offline MHE run on data from the experimental run for parameter estimation, using a linear calibration curve for the OD probe. The purpose was to see the effect the calibration in biomass had on the estimator. The RMSE is given in Table 5.6
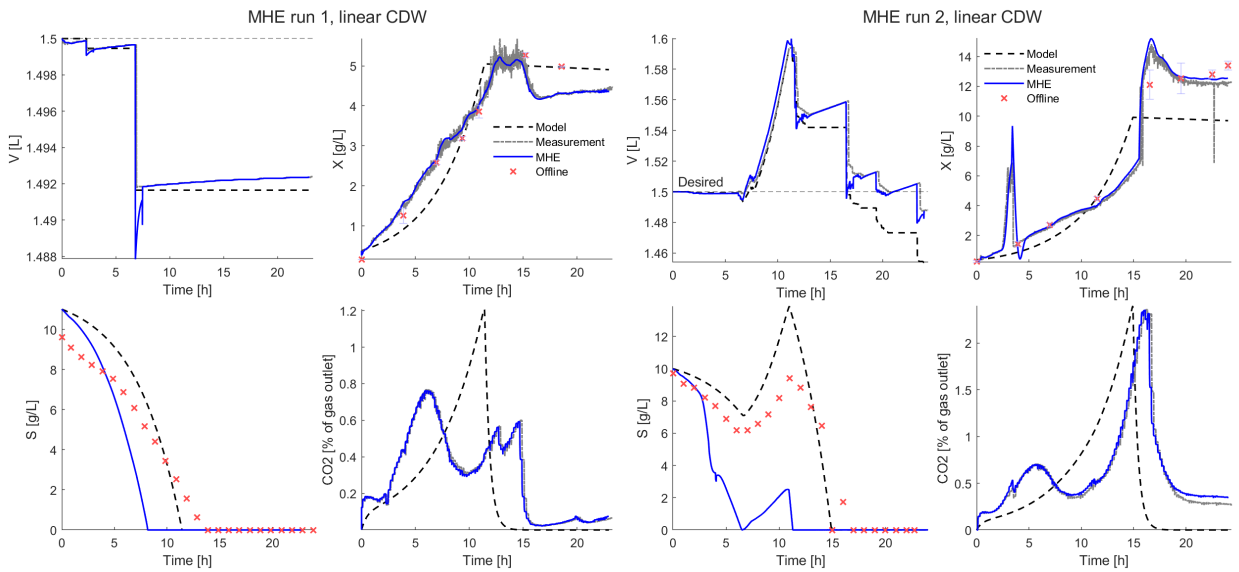
(a) MHE on run 1 with linear calibration curve for biomass.   **(b)** MHE on run 2 with linear calibration curve for biomass.

Figure 5.18: Offline MHE on experimental data from a) run 1, b) run 2, using a linear calibration curve. The purpose was to see the effect of the calibration curve of the OD probe on the MHE performance. The RMSE is given in Table 5.6.

5.4.5   Offline MHE on shortened second run

MHE was rerun offline for run 2, seen in Figure 5.11. The purpose of this was to see the effect the disturbance in biomass (see Figure 5.11) had on the estimator. The MHE was run with the same tuning given in Section 3.3.2, but run on the data after the disturbance, so 5 hours into the run. The result of the estimation is seen in Figure 5.19a with all RMSEs presented in Table 5.7. The RMSE is 1.6017 for X and 1.9404 for S, so a slight increase in X and a significant decrease in S compared to the original run.

Table 5.7: RMSE for offline MHE on the experimental run 2 using data from 5-24 hours (Figure 5.19a), and the same cropped experiment, using a linear calibration curve for the OD probe (Section 5.4.5). The calculation of the RMSE is given in Equation (2.6).

Variable	Cropped		Cropped & Linear	
	Model	MHE	Model	MHE
X	2.6244	1.6017	2.6399	1.0166
S	2.3458	1.9404	2.7281	1.8093

(a) MHE on cropped run 2

(b) MHE on shortened run 2, using the linear calibration curve for biomass.
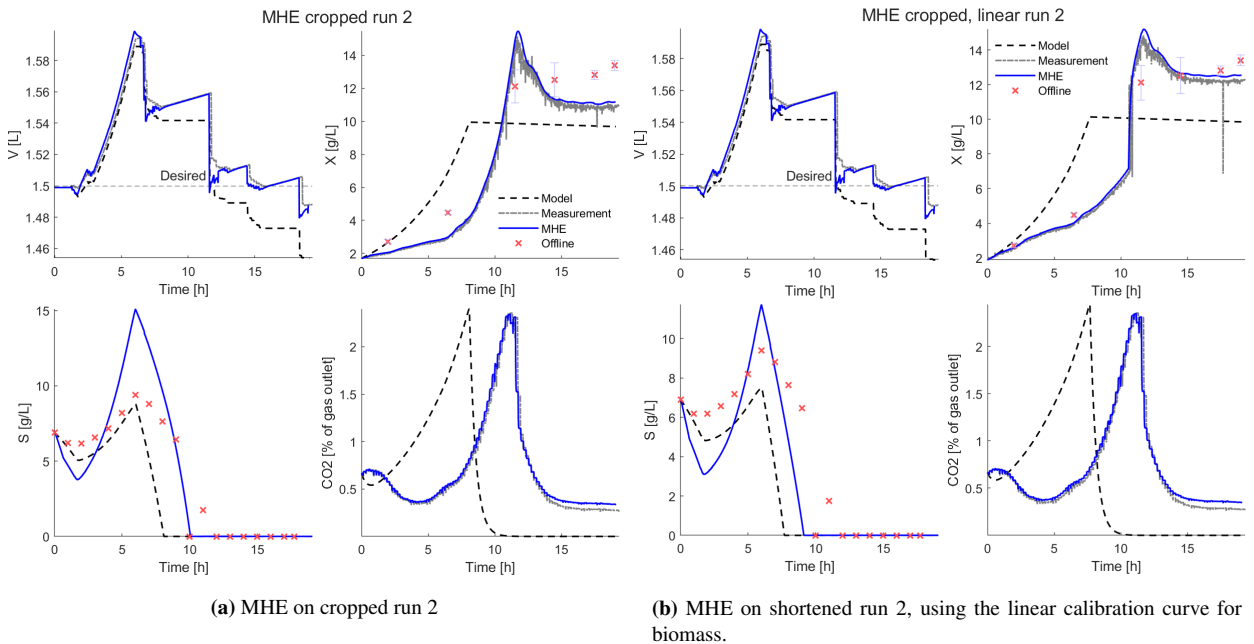
Figure 5.19: Offline MHE on experimental data from run 2, using data from 5-24 hours, here shown as 0-19 hours. a) Shows MHE on solely the cropped data, while b) also uses the linear calibration curve for biomass. The purpose was to see the effect of the disturbance in biomass (see Figure 5.11) The RMSE is given in Table 5.7.

Chapter 6

Discussion

In Section 1.3 we presented the thesis statement:

> The implementation of moving horizon estimation can overcome the challenge of limited sugar measurement in bioprocesses, enabling reliable real-time state estimation in a continuous cultivation of *Corynebacterium glutamicum*.

Figures 5.9 and 5.11 show that real-time state estimation of sugar using MHE is feasible, where the first run gives a low sugar RMSE of 1.6319, while the second run has a higher RMSE of 2.6279. The estimates are obtained solely based on the nominal model and measurements of volume, biomass, and CO_2. However, despite the feasability there are some challenges related to the reliability of the estimates.

Firstly, the MHE performance is highly dependent on the quality of the online measurements. This is seen clearly in Figure 5.11 where the MHE is not able to ignore the large initial disturbance in biomass caused by a high presence of foam in the reactor. The large increase in the X measurement causes the MHE to believe that large amounts of sugar are consumed, which is not the case. We see from the at-line sugar measurement that the sugar level is in fact much higher than the estimate, in addition to the CDW sample showing that the biomass in the reactor should be lower than what the OD probe measures. Had the MHE been able to counter this disturbance the sugar estimate would be more accurate. In fact, rerunning the MHE on the data after the disturbance, as is done in Figure 5.19a, an RMSE for S of 1.9404 is obtained, as opposed to 2.6279 for the real-time run.

The quality of the online measurements themselves relies on the proper calibration of the

equipment. As Brunner et al. (2021) discusses; "If the input to a soft sensor is faulty, there is a high probability that the output is faulty as well.", if the information the estimator receives is wrong or too uncertain, we cannot expect to get an accurate sugar estimate. The nominal model in Equation (2.1) states that the consumption of sugar is dependent on the current amount of bacteria (X), meaning the biomass measurement is important for the MHE's sugar estimate. Section 4.3.2 presents how the OD probe was calibrated, and in Figures 5.2 and 5.3 we see that a sinusoidal calibration curve for NIR-CDW was used for the OD probe. In Section 4.3.2 we mentioned that a linear calibration curve was also made and considered (see Appendix B.7, Equations (B.3) and (B.4)). In Figures 5.17 and 5.18 and table 5.6 we see that using a linear calibration curve (switching linear curves at NIR = 0.9) for the OD probe gives a significant better biomass estimate for all runs. For the MHE tuning for example (Figure 5.17), the biomass estimate is improved from an RMSE of 0.8626 to 0.3123 with the linear curve. If we for the second MHE run, in addition, remove the initial disturbance in X, seen in Section 5.4.5 and table 5.7, we achieve an RMSE for sugar of 1.8093, much better than the original 2.6279. Although the sugar estimates for all runs are not significantly improved with the linear calibration curve, the sugar dynamics are better reflected in the MHE for all data sets, seen especially in Figures 5.17 and 5.18b, where the sugar estimate follows the sugar measurement curve.

A possibility to improve the quality of the measurements would be to use alternative and additional measurement devices. The robustness of the soft sensor will then increase with the feedback information. For example, the use of biomass probes (if there are available ports) to improve biomass measurements by choosing the probe with more stable measurements.

Secondly, the MHE performance is dependent on its tuning and process parameters. Table 5.4 shows that "MHE tuning" (Figure 5.8) had the lowest RMSE, which is the experimental run the MHE was tuned for. If we *retune* the estimator using the tuning parameters R_x and w_x for the second MHE run, we see in Figure 5.15 that the MHE estimate able to counter some of the disturbance in X. However, even with an increase in uncertainty in the measurements from $R_x = 0.1$ to 5, we are not able to tune out the noise significantly. To bypass the biomass disturbance, one solution could be to add a filter to tune out fast changes in X, or to add an automatic antifoam dispenser (if available ports) to avoid large foam increases. When the performance of the state estimator is dependent on tuning and process parameters, it means we can get a very good estimator for a specific process, like in Figure 5.8. On the other hand, this means that the estimator is not versatile in the sense that it must be retuned for a slightly different process. This is the exact reason we performed a

parameter estimation and tuned the MHE for this continuous process as opposed to using the parameters and tuning in Tuveri et al. (2022) for a fed-batch process, even though they both used the same model (although F_{out} was set to zero for the fed-batch). Two of the experimental runs (Figures 5.8 and 5.11) exhibit a steady-state offset in biomass X and CO_2. This behaviour was also observed for a fed-batch system in Tuveri et al. (2022). The offset could be removed by lowering the process noise (w) for biomass and CO_2 in the tuning, but this may cause a general overfit to the measurements.

Thirdly the MHE performance is dependent on the quality of the initial guess. In the first real-time MHE run the estimator is given the wrong initial sugar concentration. We see in Figure 5.9 that the estimate (solid blue) and model (dashed black) are above the sugar measurements (red 'x'). P_0, the initial error weighting matrix, reflects the confidence in the initial condition. By decreasing the weight for sugar we can increase the uncertainty in the initial condition for sugar, possibly improving the estimate. In Figure 5.13 we reran the MHE for a range of different $P_{0,S}$ values, however, the sugar estimates did not come closer to the true value, instead they diverged further away. Here it is also apparent how the arrival cost can affect the sugar estimate, causing a large variance in the sugar estimates based on $P_{0,S}$. The maximum sugar estimate is 17g/L (red line in Figure 5.13) for the $P_{0,S}$ values of 1e-8, the highest uncertainty simulated. We see that when this max curve reaches zero, the actual sugar is 3g/L. The same behaviour is seen for all the $P_{0,S}$s, the consumption rate is higher than reality. This suggests that the growth rate μ might be wrong, or even that μ changes with the density of X. To recover from inaccurate initial conditions it may be of interest to look into alternate methods to calculating the arrival cost, where some methods are discussed in Elsheikh et al. (2021b). The same paper discusses how multirate MHE can improve sugar estimates by including delayed sugar measurements within the moving horizon.

From the above arguments, the MHE does not currently seem reliable as a real-time state estimator. However, let us discuss some of the MHE's strengths observed, and factors that may have affected its performance negatively in our experimental work. First, let us compare the MHE to the nominal model. The nominal model is not affected by disturbances, as it receives no information on the measurements, apart from the initial state and input (F_{in} and F_{out}). Table 5.4 shows that the model's RMSE for sugar is only slightly higher than the MHEs for the two first experimental runs, and lower for the second real-time MHE run. From this it may be tempting to just use the nominal model for estimates, however, the MHE offers several advantages over the nominal model.

First and foremost the MHE handles plant model mismatch. In practice, there can be

discrepancies between the nominal model and the actual system, this mismatch may be structural, caused by unmodeled dynamics, or parametric, whereby the model structure is correct but the parameters are not (Simkoff et al., 2018). We know that there is a structural mismatch present. This is seen in the difference between the MPC's predicted states (Figure 5.5) and the observed experimental states (Figures 5.8, 5.9 and 5.11). The biomass steady state observed during successful feeding in the experiments (Figures 5.8 and 5.11) was around 12g/L, which is lower than the predicted values above 15g/L from the feeding profile generated by MPC (Figure 5.4). In addition, there is a small bump in the CO_2 at 2-10 h seen in all of the experimental runs, that is not captured by the model. Indeed the nominal model cannot explain these unexpected dynamics, the MHE in contrast, considers online measurements thereby capturing the system dynamics. The MHE's ability to take the system state into account is also visible in Figure 5.11 where the MHE follows the system's CO_2, which is shifted to the right compared to the nominal model.

Another example of plant-model mismatch is visible in all runs after around 16 hours, where we see that the model shows no CO_2 in the system, in contrast to the measurements. The model behaviour is obvious in Equation (2.1); when no sugar is present (as is the case) the first term in dCO_2/dt disappears, resulting in a negative rate of CO_2. This is because the CO_2 dynamics of the model assumes growth is the sole reason for CO_2 release, while in reality production is also linked to metabolism. Although there is no growth, the bacteria still consumes oxygen, causing the release of CO_2 through respiration. A model that takes oxygen consumption into account may help compensate for unexpected changes in the dynamics, and in this way reduce plant-model mismatch like the one observed. However, this would come at the cost of the simplicity of the nominal model in Equation (2.1).

A contribution, but not cause, to the mismatch between the expected dynamics from the MPC in Figure 5.5 and the observed dynamics in the experimental runs, is the difference in initial conditions for biomass and sugar. For example, the parameter estimation and run 2 used different values of X_0 and S_0 (parameter estimation: X=0.2631, S=11.8238, run 2: X=0.3601, S=10.0000) than those used by MPC (X=0.8, S=10). Therefore the feed amount provided by the MPC may not have been sufficient to achieve the desired biomass value of 15 g/L for the experimental runs, due to the use of lower initial conditions than reality.

Secondly, the MHE provides state estimates that reflect the current state of the system. This is essential for closed-loop control integration, as the controller makes decisions based on the current states, implementing strategies to actively adapt and optimize the system's behaviour in real-time. With closed-loop control the feeding (input and output

flow) strategy could be adapted in real-time to reach a biomass of 15 g/L. From previous runs (see Figure B.1) the batch phase lasted around 6 hours before the sugar was completely depleted, therefore the input start from the MPC was shifted from 10 to 6 hours (Figure 5.4). This "guessing" of feed start would be avoided with a closed-loop controller, that can make the optimal choice as defined by the objective function in Equation (3.6).

Overall, the performance of the MHE might have been better if the experiments were performed consistently. The major cause of the experimental discrepancies were issues with the outflow. In all three experimental runs, the outlet flow presented an issue, leading to a failure to maintain a constant volume. This is reflected in Figures 5.7, 5.10 and 5.12 which indicates several spikes in the outflow. These spikes were caused by the manual pumping of liquid out of the reactor, and occasionally disturbances caused by fluctuations in the scale. Fluctuations in the outlet scale were a disturbance and did not result in actual volume changes, but they were still recorded as such (as seen in V in Figure 5.9). More information on this issue is provided in Appendix B.9. The unstable outflow means that we were not able to have continuous cultivation as intended.

Not related to the MHE performance, but what may pose an issue for closed-loop control, is a mismatch between the intended flow and the actual feed, as evident from Figures 5.7 and 5.12. Possible causes for this discrepancy include errors in the pump calibration shown in Figure 5.1, the use of pure water instead of the actual feed during calibration (which has a higher density, as explained in Appendix B.2), or errors in the scales.

6.1   Further work

There are several areas for further work to improve the accuracy and reliability of the soft sensor, including improving the experimental setup.

To improve the reliability of MHE estimates, there are several steps that can be taken. One important aspect is improving the quality and consistency of experimental measurements. This can be achieved by taking steps to isolate scales to avoid disturbances from adjacent equipment. Additionally, ensuring correct volume and flow calculations is essential. An alternative to using the scales to calculate flow would be to use the registered pump rotations, however, this may also give inaccuracies if the pump is rotating, but not moving any liquid. Using these rotation measurements in these experiments would have given a completely wrong flow out (larger than reality) as the outlet pump was not able to pump liquid out at a low rate during the run (Figures 5.7, 5.10 and 5.12). However, the most crucial step is to address the irregular outlet flow issue, which is necessary to enable continuous

cultivation.

Another way to improve MHE performance is by recalibrating the NIR-probe and CDW with a linear calibration curve, as we saw this gave better biomass estimates. Additionally, pumps should be recalibrated with higher-density liquids, especially to ensure accurate input (u) if moving forward with closed-loop control. The inlet pump should be calibrated with the intended feed (e.g. 15% sugar feed), and the outlet could be calibrated at different OD densities so that the most relevant calibration curves can be used during cultivation.

To improve the MHE further, the option of using multirate MHE to incorporate delayed measurements can be explored. This could be particularly useful now that a new HPLC processing method has been developed in the laboratory, reducing the sugar analysis time to just 20 minutes. This means that multiple sugar measurements can be taken within the sliding window, making multirate MHE a possibility. Additionally, one could also consider using an alternate arrival cost calculation to improve the estimates as discussed in Elsheikh et al. (2021b).

Plant model mismatch was observed in the experiments, it may therefore be of interest to explore alternate models that can explain more of the dynamics. This may for example be the inclusion of oxygen consumption, or CO_2 generation related to metabolism.

The overarching objective of developing a soft sensor is to enable monitoring, and process control through full-state feedback, facilitating automation of the bioprocess industry. Before implementing closed-loop control strategies like model predictive control (MPC), it is necessary to evaluate the performance of the Moving Horizon Estimation (MHE) technique. The accuracy of state estimation is critical for closed-loop controller performance, so comprehensive testing is needed before moving forward. This will help distinguish issues related to the estimator from the controller. This evaluation should include a longer cultivation period of at least 48 hours, to unveil potential weaknesses or strengths in the MHE, or reveal new dynamics that need to be addressed before proceeding. For example, a steady-state offset was observed in Figures 5.8 and 5.11, which may also be identified during a longer cultivation period, potentially causing problems for control.

Chapter 7

Conclusion

The aim of this master thesis was to investigate the implementation of Moving Horizon Estimation (MHE) as a solution to the challenge of limited sugar measurement in continuous cultivation of *Corynebacterium glutamicum* (*C. glutamicum*). The results obtained in this study provide valuable insights into the capabilities and performance of the MHE approach and demonstrates the potential of MHE for real-time state estimation.

Our findings demonstrate that real-time state estimation of sugar is achievable using MHE in the cultivation of *C. glutamicum* using three available measurements, but questions its current reliability as a soft sensor for closed-loop control. Although the estimator did not consistently estimate the true sugar value, it captured the sugar dynamics, as evidenced by the low root mean square error (RMSE) observed in the experimental runs. Notably, the largest deviation in estimation was attributed to disturbances in the biomass measurements. We observed that using a linear calibration curve, as opposed to the sinusoidal used in real-time, for optical density (OD) measurements in offline data analysis, improved the RMSE for biomass estimation and subsequently enhanced the dynamics of sugar estimation.

Furthermore, the results highlight the significance of reliable measurements from stable and correctly calibrated equipment, such as peristaltic pumps and OD probes, for optimal performance of soft sensors. It is recommended to revisit and improve these steps for future work. Moreover, it was observed that the MHE partially mitigated the effects of plant-model mismatch. However, to further reduce such discrepancies, it may be of interest to explore new models that incorporate factors like CO_2 production related to metabolism or oxygen consumption. Additionally, for optimal performance the processes on which the soft sensor is employed should closely resemble the process for which the MHE was tuned.

Although issues with the outlet flow hindered a definitive assessment of MHE efficiency for continuous cultivation, the obtained results are encouraging. Despite disruptions in the outlet and biomass measurements, the MHE approach exhibited a low RMSE for sugar estimation, suggesting that a more continuous and less disrupted system could yield even more accurate estimates. Exploring alternative arrival cost updates is proposed to further enhance the performance of the estimator, especially to recover from inaccurate initial conditions. Furthermore, conducting longer cultivation experiments would provide insights into the long-term efficiency of the MHE approach.

To ascertain the reliability of the soft sensor for closed-loop control, a real-time experiment involving control with full-state feedback through MHE is recommended. This would allow for a comprehensive evaluation of the system's performance and its applicability in closed-loop control scenarios.

Bibliography

Ronald Alexander, Gilson Campani, San Dinh, and Fernando V. Lima. Challenges and opportunities on nonlinear state estimation of chemical and biochemical processes. *Processes*, 8:1462, 11 2020. ISSN 2227-9717. doi: 10.3390/pr8111462.

Joseph S. Alford. Bioprocess control: Advances and challenges. *Computers & Chemical Engineering*, 30:1464–1475, 9 2006. ISSN 00981354. doi: 10.1016/j.compchemeng. 2006.05.039.

Jarinah Mohd Ali, N. Ha Hoang, M.A. Hussain, and Denis Dochain. Review and classification of recent observers applied in chemical process systems. *Computers & Chemical Engineering*, 76:27–41, 5 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015. 01.019.

Ahmad Ashoori, Behzad Moshiri, Ali Khaki-Sedigh, and Mohammad Reza Bakhtiari. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. *Journal of Process Control*, 19(7):1162–1173, July 2009. doi: 10.1016/j. jprocont.2009.03.006.

Carsten Bäumchen, Arnd Knoll, Bernward Husemann, Juri Seletzky, Bernd Maier, Carsten Dietrich, Ghassem Amoabediny, and Jochen Büchs. Effect of elevated dissolved carbon dioxide concentrations on growth of corynebacterium glutamicum on d-glucose and l-lactate. *Journal of Biotechnology*, 128(4):868–874, March 2007. doi: 10.1016/j.jbiotec. 2007.01.001.

Judith Becker, Gideon Gießelmann, Sarah Lisa Hoffmann, and Christoph Wittmann. Corynebacterium glutamicum for sustainable bioproduction: From metabolic physiol-

ogy to systems metabolic engineering. In *Synthetic Biology – Metabolic Engineering*, pages 217–263. Springer International Publishing, 2016. doi: 10.1007/10_2016_21.

Biegler and Lorenz T. *Nonlinear Programming*, pages 1–5, 244–246, 292. Society for Industrial and Applied Mathematics, 2010. doi: 10.1137/1.9780898719383.

Anne Blackwell. Bioreactor monitoring: More to measure than glucose in, mab out. *BioPharm International*, 31(6):14–19, 2018. URL https://www.biopharminternational. com/view/bioreactor-monitoring-more-measure-glucose-mab-out.

Hans Georg Bock, Stefan Körkel, and Johannes P. Schlöder. Parameter estimation and optimum experimental design for differential equation models. In *Contributions in Mathematical and Computational Sciences*, pages 1–30. Springer Berlin Heidelberg, August 2012. doi: 10.1007/978-3-642-30367-8_1.

Encyclopedia Britannica. Trace element | biology. https://www.britannica.com/science/ trace-element, 1998. [Accessed 20-Feb-2023].

Vincent Brunner, Manuel Siegl, Dominik Geier, and Thomas Becker. Challenges in the development of soft sensors for bioprocesses: A critical review. *Frontiers in Bioengineering and Biotechnology*, 9, 8 2021. ISSN 2296-4185. doi: 10.3389/fbioe.2021.722202.

Anjani Devi Chintagunta, Gaetano Zuccaro, Mahesh Kumar, S. P. Jeevan Kumar, Vijay Kumar Garlapati, Pablo D. Postemsky, N. S. Sampath Kumar, Anuj K. Chandel, and Jesus Simal-Gandara. Biodiesel production from lignocellulosic biomass using oleaginous microbes: Prospects for integrated biofuel production. *Frontiers in Microbiology*, 12, August 2021. doi: 10.3389/fmicb.2021.658284.

Kiran Desai, Yogesh Badhe, Sanjeev S. Tambe, and Bhaskar D. Kulkarni. Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochemical Engineering Journal*, 27(3):225–239, January 2006. doi: 10.1016/j.bej.2005.08.002.

Denis Dochain. State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13(8):801–818, 2003. ISSN 0959-1524. doi: 10.1016/S0959-1524(03)00026-X.

Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9):2419–2468, April 2021. doi: 10.1007/s10994-021-05961-4.

Lothar Eggeling and Michael Bott, editors. *Handbook of Corynebacterium glutamicum*. CRC Press, March 2005. doi: 10.1201/9781420039696.

Mohamed Elsheikh, Rubin Hille, Alexandru Tatulea-Codrean, and Stefan Krämer. A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Computers & Chemical Engineering*, 146:107219, 2021a. ISSN 0098-1354. doi: 10.1016/j.compchemeng.2020.107219.

Mohamed Elsheikh, Rubin Hille, Alexandru Tatulea-Codrean, and Stefan Krämer. A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Computers & Chemical Engineering*, 146:107219, 3 2021b. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.107219.

Paul Evans. Properties of carbon dioxide at atmospheric pressure, Jun 2020. URL https://theengineeringmindset.com/properties-of-carbon-dioxide-at-atmospheric-pressure/.

Christodoulos A. Floudas, Pãnos M. Pardalos, Claire S. Adjiman, William R. Esposito, Zeynep H. Gümüş, Stephen T. Harding, John L. Klepeis, Clifford A. Meyer, and Carl A. Schweiger. Dynamic optimization problems. In *Handbook of Test Problems in Local and Global Optimization*, pages 351–412. Springer, Boston, MA, 1999. doi: 10.1007/978-1-4757-3040-1_15.

Martin Follmann, Ines Ochrombel, Reinhard Krämer, Christian Trötschel, Ansgar Poetsch, Christian Rückert, Andrea Hüser, Marcus Persicke, Dominic Seiferling, Jörn Kalinowski, and Kay Marin. Functional genomics of pH homeostasis in corynebacterium glutamicum revealed novel links between pH response, oxidative stress, iron homeostasis and methionine synthesis. *BMC Genomics*, 10(1), December 2009. doi: 10.1186/1471-2164-10-621.

Kapil G. Gadkar, Sarika Mehra, and James Gomes. On-line adaptation of neural networks for bioprocess control. *Computers & Chemical Engineering*, 29(5):1047–1057, April 2005. doi: 10.1016/j.compchemeng.2004.11.004.

A.L. Galant, R.C. Kaufman, and J.D. Wilson. Glucose: Detection and analysis. *Food Chemistry*, 188:149–160, December 2015. doi: 10.1016/j.foodchem.2015.04.071.

Stefan Gnoth, Marco Jenzsch, Rimvydas Simutis, and Andreas Lübbert. Control of cultivation processes for recombinant protein production: a review. *Bioprocess and Biosystems Engineering*, 31:21–39, 1 2008. ISSN 1615-7591. doi: 10.1007/s00449-007-0163-7.

G. Goffaux and A. Vande Wouwer. Design of a robust nonlinear receding-horizon observer - application to a biological system. *IFAC Proceedings Volumes*, 41(2):15553–15558, 2008. doi: 10.3182/20080706-5-kr-1001.02630.

Aydin Golabgir and Christoph Herwig. Combining mechanistic modeling and raman spectroscopy for real-time monitoring of fed-batch penicillin production. *Chemie Ingenieur Technik*, 88(6):764–776, March 2016. doi: 10.1002/cite.201500101.

Karen Gonzalez, Sihem Tebbani, Didier Dumur, Filipa Lopes, Dominique Pareau, Aurore Thorigne, and Sebastien Givry. Feedback linearizing controller coupled to an unscented kalman filter for lactic acid regulation. In *2015 19th International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, October 2015. doi: 10.1109/icstcc. 2015.7321296.

V. Gopinath and K.M. Nampoothiri. Corynebacterium glutamicum. In *Encyclopedia of Food Microbiology*, pages 504–517. Elsevier, 2014. doi: 10.1016/b978-0-12-384730-0. 00076-8.

Michaela Graf, Thorsten Haas, Felix Müller, Anina Buchmann, Julia Harm-Bekbenbetova, Andreas Freund, Alexander Nieß, Marcus Persicke, Jörn Kalinowski, Bastian Blombach, and Ralf Takors. Continuous adaptive evolution of a fast-growing corynebacterium glutamicum strain independent of protocatechuate. *Frontiers in Microbiology*, 10, August 2019. doi: 10.3389/fmicb.2019.01648.

A. Hartbrich, G. Schmitz, D. Weuster-Botz, A. A. de Graaf, and C. Wandrey. Development and application of a membrane cyclone reactor for in vivo NMR spectroscopy with high microbial cell densities. *Biotechnology and Bioengineering*, 51(6):624–635, March 2000. doi: 10.1002/(sici)1097-0290(19960920)51:6<624::aid-bit2>3.0.co;2-j.

Eric L. Haseltine and James B. Rawlings. Critical evaluation of extended kalman filtering and moving-horizon estimation. *Industrial & Engineering Chemistry Research*, 44(8): 2451–2460, June 2004. doi: 10.1021/ie034308l.

Elijah Hedrick, Katherine Hedrick, Debangsu Bhattacharyya, Stephen E. Zitney, and Benjamin Omell. Reinforcement learning for online adaptation of model predictive controllers: Application to a selective catalytic reduction unit. *Computers & Chemical Engineering*, 160:107727, April 2022. doi: 10.1016/j.compchemeng.2022.107727.

Laura M. Helleckes, Johannes Hemmerich, Wolfgang Wiechert, Eric von Lieres, and

Alexander Grünberger. Machine learning in bioprocess development: from promise to practice. *Trends in Biotechnology*, November 2022. doi: 10.1016/j.tibtech.2022.10.010.

INFORS HT. *The Cookbook BASIC CONCEPTS, RECIPES AND STRATEGIES FOR BIOPROCESSES INVOLVING CELL CULTURES AND MICROORGANISMS*. IN-FORS HT, 2023.

Thermo Fisher Scientific Inc. Potassium hydroxide, 4m, chem lab | fisher scientific. https://www.fishersci.no/shop/products/potassium-hydroxide-4m/11905049, 2023. [Accessed 16-Mar-2023].

Masayuki Inui, Shikiko Murakami, Shohei Okino, Hideo Kawaguchi, Alain A. Vertès, and Hideaki Yukawa. Metabolic analysis of corynebacterium glutamicum during lactate and succinate productions under oxygen deprivation conditions. *Microbial Physiology*, 7(4): 182–196, 2004. doi: 10.1159/000079827.

Banafsheh Jabarivelisdeh, Lisa Carius, Rolf Findeisen, and Steffen Waldherr. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Computers & Chemical Engineering*, 135:106744, 4 2020. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.106744.

Jenny E. Jeong and Peng Qiu. Quantifying the relative importance of experimental data points in parameter estimation. *BMC Systems Biology*, 12(S6), November 2018. doi: 10.1186/s12918-018-0622-6.

Simon Julier, Jeffrey Uhlmann, and Hugh Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of 1995 American Control Conference - ACC'95*, volume 3, pages 1628–1632. American Control Conference Seattle, Washington, June 1995. doi: 10.1109/ACC.1995.529783.

Robert J Kadner and Kara Rogers. Bacteria in industry, Jan 2023. URL https://www.britannica.com/science/bacteria/Bacteria-in-industry.

L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, May 1996. doi: 10.1613/jair.301.

Julian Kager, Christoph Herwig, and Ines Viktoria Stelzer. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. *Chemical Engineering Science*, 177:234–244, February 2018. doi: 10.1016/j.ces.2017.11.049.

Jörn Kalinowski, Brigitte Bathe, Daniela Bartels, Nicole Bischoff, Michael Bott, Andreas Burkovski, Nicole Dusch, Lothar Eggeling, Bernhard J Eikmanns, Lars Gaigalat, Alexander Goesmann, Michael Hartmann, Klaus Huthmacher, Reinhard Krämer, Burkhard Linke, Alice C McHardy, Folker Meyer, Bettina Möckel, Walter Pfefferle, Alfred Pühler, Daniel A Rey, Christian Rückert, Oliver Rupp, Hermann Sahm, Volker F Wendisch, Iris Wiegräbe, and Andreas Tauch. The complete corynebacterium glutamicum ATCC 13032 genome sequence and its impact on the production of l-aspartate-derived amino acids and vitamins. *Journal of Biotechnology*, 104(1-3):5–25, September 2003. doi: 10.1016/s0168-1656(03)00154-8.

Rambabu Kandepu, Bjarne Foss, and Lars Imsland. Applying the unscented kalman filter for nonlinear state estimation. *Journal of Process Control*, 18(7-8):753–768, August 2008. doi: 10.1016/j.jprocont.2007.11.004.

Ryo Karakida and Kazuki Osawa. Understanding approximate fisher information for fast convergence of natural gradient descent in wide neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124010, December 2021. doi: 10.1088/1742-5468/ac3ae3.

Hideo Kawaguchi, Alain A. Vertès, Shohei Okino, Masayuki Inui, and Hideaki Yukawa. Engineering of a xylose metabolic pathway in corynebacterium glutamicum. *Applied and Environmental Microbiology*, 72(5):3418–3428, May 2006. doi: 10.1128/aem.72.5.3418-3428.2006.

Merck KGaA. Phosphoric acid 10%. https://www.merckmillipore.com/NO/en/product/Phosphoric-acid-100-0,MDA_CHEM-480951, 2023a.

Merck KGaA. Microbiology introduction. https://www.sigmaaldrich.com/NO/en/technical-documents/technical-article/microbiological-testing/microbial-culture-media-preparation/microbiology-introduction, 2023b. [Accessed 24-May-2023].

Jong Woo Kim, Niels Krausch, Judit Aizpuru, Tilman Barz, Sergio Lucia, Peter Neubauer, and Mariano Nicolas Cruz Bournazou. Model predictive control and moving horizon estimation for adaptive optimal bolus feeding in high-throughput cultivation of e. coli. *Computers & Chemical Engineering*, 172:108158, April 2023. doi: 10.1016/j.compchemeng.2023.108158.

Veerendra Koppolu and Veneela KR Vasigala. Role of escherichia coli in biofuel production. *Microbiology Insights*, 9:MBI.S10878, January 2016. doi: 10.4137/mbi.s10878.

Ekaterina Kostina. Robust parameter estimation in dynamic systems. *Optimization and Engineering*, 5(4):461–484, December 2004. doi: 10.1023/b:opte.0000042035.67293. 92.

D. Krämer and R. King. A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point kalman filter. *Journal of Process Control*, 82:91–104, October 2019. doi: 10.1016/j.jprocont.2017.11.008.

Michael Kräling, Dorothee Barth, Helmut Röck, and Joachim Hörrmann. State estimation in biotechnological processes using a software-sensor combining full-horizon observer and neural networks. *IFAC Proceedings Volumes*, 41:9655–9660, 2008. ISSN 14746670. doi: 10.3182/20080706-5-KR-1001.01633.

Anil Kumar. Role of microbes in dairy industry. *Nutrition & Food Science International Journal*, 3(3), September 2017. doi: 10.19080/nfsij.2017.03.555612.

Pavel Kyslík and Aleš Prokop. The impact of ivan málek's continuous culture concept on bioprocessing. *Journal of Industrial Microbiology & Biotechnology*, 37(12):1249–1256, November 2010. doi: 10.1007/s10295-010-0881-8.

Peter Kühl, Moritz Diehl, Tom Kraus, Johannes P. Schlöder, and Hans Georg Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers & Chemical Engineering*, 35(1):71–83, 2011a. ISSN 0098-1354. doi: 10.1016/j.compchemeng. 2010.07.012.

Peter Kühl, Moritz Diehl, Tom Kraus, Johannes P. Schlöder, and Hans Georg Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers & Chemical Engineering*, 35:71–83, 1 2011b. ISSN 00981354. doi: 10.1016/j.compchemeng. 2010.07.012.

Mario Lederle, Mircea Tric, Tatjana Roth, Lina Schütte, Anke Rattenholl, Dirk Lütkemeyer, Stefan Wölfl, Tobias Werner, and Philipp Wiedemann. Continuous optical in-line glucose monitoring and control in CHO cultures contributes to enhanced metabolic efficiency while maintaining darbepoetin alfa product quality. *Biotechnology Journal*, 16(8):2100088, June 2021. doi: 10.1002/biot.202100088.

Joo-Young Lee, Yoon-Ah Na, Eungsoo Kim, Heung-Shick Lee, and Pil Kim. The actinobacterium corynebacterium glutamicum, an industrial workhorse. *Journal of Microbiology and Biotechnology*, 26(5):807–822, May 2016. doi: 10.4014/jmb.1601.01053.

James C. Liao, Luo Mi, Sammy Pontrelli, and Shanshan Luo. Fuelling the future: micro-

bial engineering for the production of sustainable biofuels. *Nature Reviews Microbiology*, 14(5):288–304, March 2016. doi: 10.1038/nrmicro.2016.32.

Shijie Liu. How cells grow. *Bioprocess Engineering*, pages 629–697, 2017. doi: 10.1016/B978-0-444-63783-3.00011-3.

Sergio Lucia, Lisa Carius, and Rolf Findeisen. Adaptive nonlinear predictive control and estimation of microaerobic processes. *IFAC-PapersOnLine*, 50(1):12635–12640, July 2017. doi: 10.1016/j.ifacol.2017.08.2230.

Reiner Luttmann, Daniel G. Bracewell, Gesine Cornelissen, Krist V. Gernaey, Jarka Glassey, Volker C. Hass, Christian Kaiser, Christian Preusse, Gerald Striedner, and Carl-Fredrik Mandenius. Soft sensors in bioprocessing: A status report and recommendations. *Biotechnology Journal*, 7(8):1040–1048, April 2012. doi: 10.1002/biot.201100506.

Michael T. Madigan, Kelly S. Bender, Daniel H. Buckley, W. Matthew Sattley, and David A. Stahl. *Brock Biology of Microorganisms*. Pearson, 15 edition, 2018a. ISBN 9781292235103.

Michael T Madigan, Kelly S Bender, Daniel H Buckley, W Matthew Sattley, and David A Stahl. *Brock biology of microorganisms, global edition*. Pearson Education, London, England, 15 edition, January 2018b.

Riti Mann, Leigh G. Monahan, Elizabeth J. Harry, and Amy L. Bottomley. We are what we eat: True for bacteria too. *Frontiers for Young Minds*, 5, September 2017. doi: 10.3389/frym.2017.00054.

Kevin A. P. McLean and Kim B. McAuley. Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering*, 90(2):351–366, October 2011. doi: 10.1002/cjce.20660.

Matej Mikulic. World pharmaceutical revenue by technology 2012-2026, Nov 2022. URL https://www.statista.com/statistics/309457/world-pharmaceutical-revenue-distribution-by-technology/.

Manfred Morari and Jay H. Lee. Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23:667–682, 5 1999. ISSN 00981354. doi: 10.1016/S0098-1354(98)00301-9.

Natal A W Van Riel. A template for parameter estimation with matlab optimization toolbox; including dynamic systems, 2011.

David L. Nelson and Michael M. Coc. *Lehninger Principles of Biochemistry*. W. H. Freeman and Company, 7 edition, 2017. ISBN 978-1-4641-2611-6.

Taku Nishimura, Alain A. Vertès, Yoshifumi Shinoda, Masayuki Inui, and Hideaki Yukawa. Anaerobic growth of corynebacterium glutamicum using nitrate as a terminal electron acceptor. *Applied Microbiology and Biotechnology*, 75(4):889–897, March 2007. doi: 10.1007/s00253-007-0879-y.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer New York, 2 edition, 2006. doi: 10.1007/978-0-387-40065-5.

Clarivate Web of Science. 1270 results from web of science core collection for moving horizon estimation, 2021. URL https://www.webofscience.com/wos/woscc/summary/6728ca3b-b43e-495d-a487-dd051a85d90d-8586cac0/relevance/1.

Tae Hoon Oh, Hyun Min Park, Jong Woo Kim, and Jong Min Lee. Integration of reinforcement learning and model predictive control to optimize semi-batch bioreactor. *AIChE Journal*, 68(6), March 2022. doi: 10.1002/aic.17658.

Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, March 2010. doi: 10.1038/nbt.1614.

Elton Pan, Panagiotis Petsagkourakis, Max Mowbray, Dongda Zhang, and Antonio del Rio-Chanona. Constrained q-learning for batch process optimization. *IFAC-PapersOnLine*, 54(3):492–497, 2021. doi: 10.1016/j.ifacol.2021.08.290.

P. Petsagkourakis, I.O. Sandoval, E. Bradford, D. Zhang, and E.A. del Rio-Chanona. Reinforcement learning for batch bioprocess optimization. *Computers & Chemical Engineering*, 133:106649, February 2020. doi: 10.1016/j.compchemeng.2019.106649.

Nick Pittman and Magnus Wetterhall. Bioprocess monitoring and control: New and continuing needs in the biopharmaceutical industry. *BioProcess International*, 20(8), 2022.

Leslie A. Pray. Errors in dna replication. https://www.nature.com/scitable/topicpage/dna-replication-and-causes-of-mutation-409/, 2008. [Accessed 03-May-2023].

Oxford University Press. overfitting, n. In *Oxford University Dictionary*. Oxford University Press, 2023. URL https://www.oed.com/view/Entry/258314?rskey=8rjJW8&result=2.

Albin Pump. How do peristaltic pumps work?, 2022. URL https://www.albinpump.com/en-us/news/how-peristaltic-pumps-work.

Jonathan P. Raftery, Melanie R. DeSessa, and M. Nazmul Karim. Economic improvement of continuous pharmaceutical production via the optimal control of a multifeed bioreactor. *Biotechnology Progress*, 33(4):902–912, February 2017. doi: 10.1002/btpr.2433.

S. Ramaswamy, T.J. Cutright, and H.K. Qammar. Control of a continuous bioreactor using model predictive control. *Process Biochemistry*, 40(8):2763–2770, July 2005. doi: 10.1016/j.procbio.2004.12.019.

Anurag S. Rathore, Somesh Mishra, Saxena Nikita, and Priyanka Priyanka. Bioprocess control: Current progress and future perspectives. *Life*, 11:557, 6 2021. ISSN 2075-1729. doi: 10.3390/life11060557.

James Blake Rawlings, David Q Mayne, and Moritz Diehl. *Model predictive control: Theory, computation, and design*, volume 2. Nob Hill Publishing Madison, 2017. ISBN 978-0975937754.

Nina P.G. Salau, Jorge O. Trierweiler, and Argimiro R. Secchi. State estimators for better bioprocesses operation. In *Computer Aided Chemical Engineering*, pages 1267–1271. Elsevier, 2012. doi: 10.1016/b978-0-444-59520-1.50112-3.

Miho Sasaki, Toru Jojima, Masayuki Inui, and Hideaki Yukawa. Simultaneous utilization of d-cellobiose, d-glucose, and d-xylose by recombinant corynebacterium glutamicum under oxygen-deprived conditions. *Applied Microbiology and Biotechnology*, 81(4):691–699, December 2008. doi: 10.1007/s00253-008-1703-z.

Benjamin Schleich, Nabil Anwer, Luc Mathieu, and Sandro Wartzack. Shaping the digital twin for design and production engineering. *CIRP Annals*, 66(1):141–144, 2017. doi: 10.1016/j.cirp.2017.04.040.

René Schneider and Christos Georgakis. How to NOT make the extended kalman filter fail. *Industrial & Engineering Chemistry Research*, 52(9):3354–3362, February 2013. doi: 10.1021/ie300415d.

Dale E Seborg, Thomas F Edgar, Duncan A Mellichamp, and Francis J Doyle, III. *Process dynamics and control*. John Wiley & Sons, September 2016.

B.G. Sileshi, C. Ferrer, and J. Oliver. Accelerating techniques for particle filter implementations on FPGA. In *Emerging Trends in Computational Biology, Bioinformatics,*

and Systems Biology, pages 19–37. Elsevier, 2015. doi: 10.1016/b978-0-12-802508-6.00002-8.

Jodie M. Simkoff, Siyun Wang, Michael Baldea, Leo H. Chiang, Ivan Castillo, Rahul Bindlish, and David B. Stanley. Plant–model mismatch estimation from closed-loop data for state-space model predictive control. *Industrial & Engineering Chemistry Research*, 57(10):3732–3741, February 2018. doi: 10.1021/acs.iecr.7b04917.

Peter Sinner, Marlene Stiegler, Christoph Herwig, and Julian Kager. Noninvasive online monitoring of corynebacterium glutamicum fed-batch bioprocesses subject to spent sulfite liquor raw material uncertainty. *Bioresource Technology*, 321:124395, February 2021. doi: 10.1016/j.biortech.2020.124395.

Ines V. Stelzer, Julian Kager, and Christoph Herwig. Comparison of particle filter and extended kalman filter algorithms for monitoring of bioprocesses. In *Computer Aided Chemical Engineering*, pages 1483–1488. Elsevier, 2017. doi: 10.1016/b978-0-444-63965-3.50249-x.

Jasem Tamimi and Pu Li. Nonlinear model predictive control using multiple shooting combined with collocation on finite elements. *IFAC Proceedings Volumes*, 42:703–708, 2009. ISSN 14746670. doi: 10.3182/20090712-4-TR-2008.00114.

Josh A. Taylor, Alain Rapaport, and Denis Dochain. A sequential convex moving horizon estimator for bioprocesses. *Journal of Process Control*, 116:19–24, August 2022. doi: 10.1016/j.jprocont.2022.05.012.

Sihem Tebbani, Didier Dumur, and Ghizlane Hafidi. Open-loop optimization and trajectory tracking of a fed-batch bioreactor. *Chemical Engineering and Processing: Process Intensification*, 47(11):1933–1941, October 2008. doi: 10.1016/j.cep.2007.10.009.

Sihem Tebbani, Laurent Le Brusquet, Emil Petre, and Dan Selisteanu. Robust moving horizon state estimation: Application to bioprocesses. In *2013 17th International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, October 2013. doi: 10.1109/icstcc.2013.6689014.

Neythen J. Treloar, Alex J. H. Fedorec, Brian Ingalls, and Chris P. Barnes. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLOS Computational Biology*, 16(4):e1007783, April 2020a. doi: 10.1371/journal.pcbi.1007783.

Neythen J. Treloar, Alex J. H. Fedorec, Brian Ingalls, and Chris P. Barnes. Deep reinforce-

ment learning for the control of microbial co-cultures in bioreactors. *PLOS Computational Biology*, 16(4):e1007783, April 2020b. doi: 10.1371/journal.pcbi.1007783.

Andrea Tuveri, Fernando Pérez-García, Pedro A. Lira-Parada, Lars Imsland, and Nadav Bar. Sensor fusion based on extended and unscented kalman filter for bioprocess monitoring. *Journal of Process Control*, 106:195–207, oct 2021. doi: 10.1016/j.jprocont. 2021.09.005.

Andrea Tuveri, Haakon Eng Holck, Caroline S.M. Nakama, José Matias, Johannes Jäschke, Lars Imsland, and Nadav Bar. Bioprocess monitoring: A moving horizon estimation experimental application. *IFAC-PapersOnLine*, 55:222–227, 2022. ISSN 24058963. doi: 10.1016/j.ifacol.2022.07.448.

Amit Varma and Bernhard O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology*, 12(10):994–998, October 1994. doi: 10.1038/nbt1094-994.

A.A. Vertès, M. Inui, M. Kobayashi, Y. Kurusu, and H. Yukawa. Presence of mrr- and mcr-like restriction systems in coryneform bacteria. *Research in Microbiology*, 144(3): 181–185, January 1993. doi: 10.1016/0923-2508(93)90043-2.

Sun-Chong Wang. Artificial neural network. In *Interdisciplinary Computing in Java Programming*, pages 81–100. Springer US, 2003. doi: 10.1007/978-1-4615-0377-4_5.

Volker F. Wendisch, João M. P. Jorge, Fernando Pérez-García, and Elvira Sgobba. Updates on industrial production of amino acids using corynebacterium glutamicum. *World Journal of Microbiology and Biotechnology*, 32(6), April 2016. doi: 10.1007/ s11274-016-2060-1.

Stefan Wieschalka, Bastian Blombach, Michael Bott, and Bernhard J. Eikmanns. Bio-based production of organic acids with *Corynebacterium glutamicum*. *Microbial Biotechnology*, 6(2):87–102, December 2012. doi: 10.1111/1751-7915.12013.

J.A. Wilson and L.F.M. Zorzetto. A generalised approach to process state estimation using hybrid artificial neural network/mechanistic models. *Computers & Chemical Engineering*, 21(9):951–963, June 1997. doi: 10.1016/s0098-1354(96)00336-5.

Christoph Wittmann and Judith Becker. The l-lysine story: From metabolic pathways to industrial production. In *Amino Acid Biosynthesis ~ Pathways, Regulation and Metabolic Engineering*, pages 39–70. Springer Berlin Heidelberg, feb 2007. doi: 10.1007/7171_2006_089.

Huimin Xie, Xinghai Xu, Yuling Li, Wenjing Hong, and Jia Shi. Model predictive control guided reinforcement learning control scheme. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2020. doi: 10.1109/ijcnn48605.2020. 9207398.

En Yang, Lihua Fan, Yueming Jiang, Craig Doucette, and Sherry Fillmore. Antimicrobial activity of bacteriocin-producing lactic acid bacteria isolated from cheeses and yogurts. *AMB Express*, 2(1), September 2012. doi: 10.1186/2191-0855-2-48.

H. Yoon, G. Klinzing, and H. W. Blanch. Competition for mixed substrates by microbial populations. *Biotechnology and Bioengineering*, 19(8):1193–1210, August 1977. doi: 10.1002/bit.260190809.

Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, and Bernd Hitzmann. The kalman filter for the supervision of cultivation processes. In *Advances in Biochemical Engineering/Biotechnology*, pages 95–125. Springer International Publishing, 2020. doi: 10.1007/10_2020_145.

Xi Zhu and Enmin Feng. Joint estimation in batch culture by using unscented kalman filter. *Biotechnology and Bioprocess Engineering*, 17(6):1238–1243, December 2012. doi: 10.1007/s12257-012-0290-0.

Inc. © 1994-2023 The MathWorks. lsqnonlin, 2006a. URL https://se.mathworks.com/ help/optim/ug/lsqnonlin.html#d124e124964.

Inc. © 1994-2023 The MathWorks. When the solver fails, 2006b. URL https://se. mathworks.com/help/optim/ug/when-the-solver-fails.html#br53duy-1.

Publications & citations on Moving Horizon Estimation
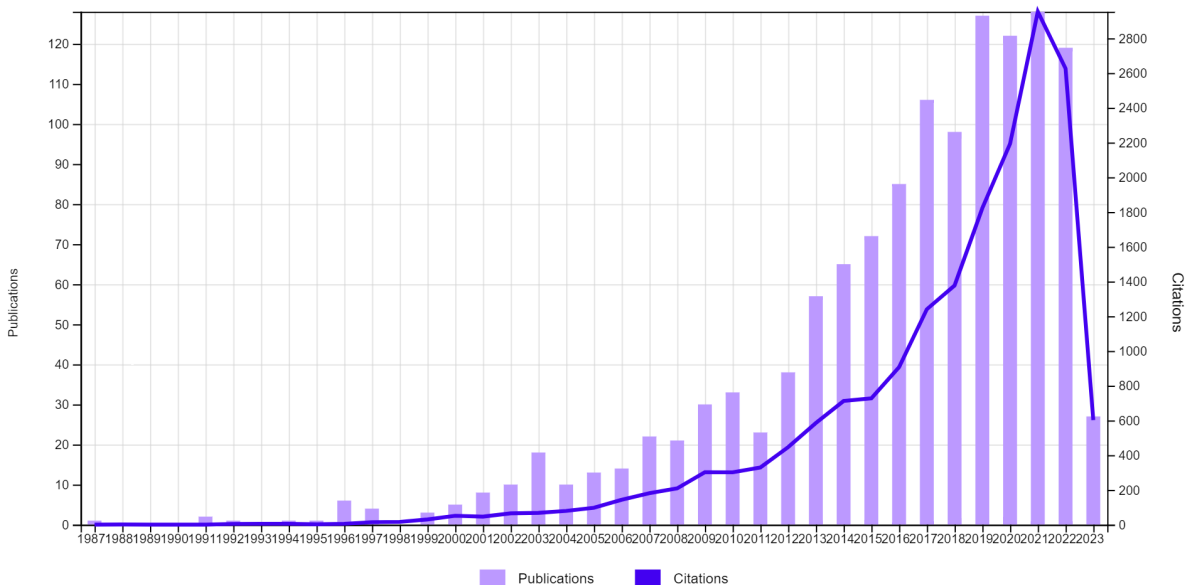
Figure A.1: Publications (purple) and citations (blue) over time, from 1987 to 2023, for publications on Moving Horizon Estimation. Citation Report graphic is derived from Clarivate Web of Science, © Copyright Clarivate 202_. All rights reserved (of Science, 2021). Collected 27.04.2023.

Calculations & details of laboratory work

B.1 Inoculum calculations

The necessary cell mass and OD values are found through back-calculations. A final OD of 1 in the bioreactor is desired for initial conditions in 1.5 L broth media.

The necessary initial volume of seed culture to reach this OD is found after measuring the OD600 of the seed culture.

$$V_{seed} = \frac{OD1 \cdot 1.5\text{L}}{OD600}$$

This volume is then spun down and washed 2 times before resuspending in 75 mL CGXII which is then added to the reactor.

B.2 Feeding density

To calculate the inlet and outlet flows the feed density is needed. The flows are calculated at every iteration (in the laboratory code Appendix E) in the following manner,

$$F_{i,x} = \frac{w_i - w_{i-1}}{\rho_{solution} \cdot dt}$$

where w is the weight of either the inlet or outlet bottle, i is the current iteration and dt is the iteration time.

The feeding mixture is a 20% glucose solution in 1.5 L. Let us find the theoretical value. In a 20% glucose solution there is 200 g glucose (solute) and 800 g water (solvent). The density of glucose ($C_6H_{12}O_6$) is 1560 kg/m^3, and 997 kg/m^3 at 25 °C. The feed consists of more chemicals than glucose (see CGXII solution), however, we are only considering glucose and water here, resulting in the theoretical density,

$$\rho_{solution} = \frac{200}{1000} \cdot 1560 + \frac{800}{1000} \cdot 997 = 1109.6\text{g/m}^3$$

The experimental value is based on the weight and volume of the feed,

$$\rho_{solution} = \frac{w_{\text{empty bottle}} + w_{\text{feed in bottle}}}{V} = \frac{3808.16\text{g} - 2190.46\text{g}}{1.5L} = 1078.467\text{g/L}$$

The experimental value was used for feed calculations.

B.3   Volume measurement

The volume is an indirect measurement based on the initial volume and volume added to, or removed from the system.

$$V_{i,x} = \frac{w_{i,x} - w_{0,x}}{\rho_{solution}}$$

Where V_i is the volume at the current timestep and x denotes if it is volume added or removed. The weight of the scale x at timestep i is $w_{i,x}$, and $w_{0,x}$ is the scale weight at the experiment start. The volume at timestep i is then determined by,

$$V_i = V_0 + V_{i,in} - Vi, out$$

B.4   Bioreactor run 12/02/23

Figure B.1 shows one of several experimental runs. Here we see the batch phase lasts around 5 hours. Several experimental runs were not included in the thesis due to some factor compromising the experiment. For example, we did not realise til later on that we should take CDW samples during all runs to not be biased by the calibration curve
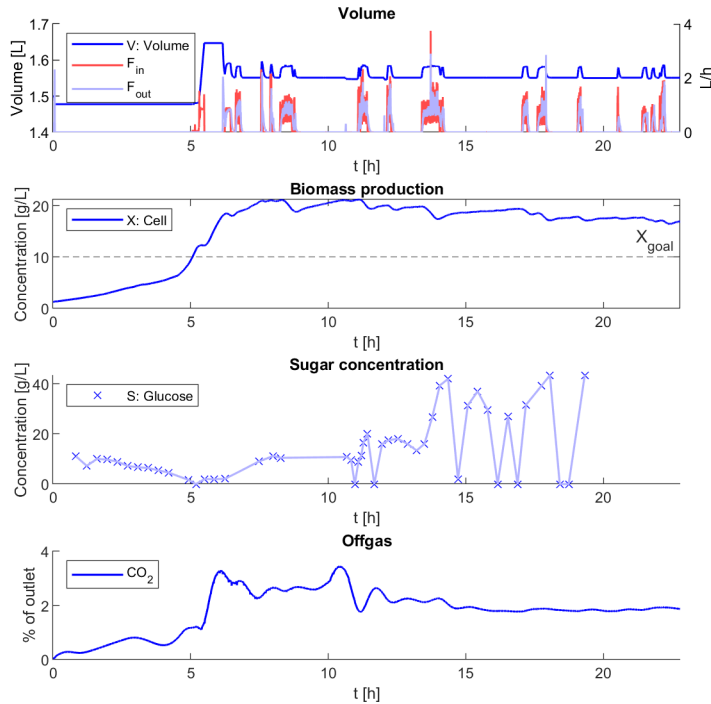
Figure B.1: The figure shows one of several experimental runs performed, but not included in the report. Here we see the batch phase lasts around 5 hours, the time before all sugar is depleted. The purple 'x' are sugar measurements from the HPLC.

in Equation (5.3). The MPC controller available in the lab, used in this experiment and several others, did not function properly, so we later landed on a fixed inlet and outlet profile.

B.5   Acid and base pump calibration

B.5.1   Acid and base density

To calculate the amount of volume added to the bioreactor to acid and base, their density is needed, which can be found in table B.1.

Table B.1: Acid and base density values taken from KGaA (2023a) and Inc. (2023)

Chemical	Density
10% H_3PO_4	1053 g/L
KOH	1170 g/L

The volume of either acid and base can be found by,

$$V_{i,x} = \frac{w_{i,x} - w_{i-1,x}}{\rho_x}$$

where w is the logged weight of x (acid or base) added.

B.5.2 Calibration

The pH is automatically regulated by the Labfoors 5 software keeping the pH at 7. To see if the amount of acid and base added to the system was substantial (and to be included in Equation (2.1)) for the 2.7 L system a calibration was performed. To perform the calibration the pumps tubes were primed with 10% H_3PO_4 (acid), and later 4M KOH (base), leading into a container placed on a scale. Acid and base were used as opposed to water to account for density differences. From the INFORS Labfors 5 screen the pumps were run on, 25%, 50%, 75% and 100% speeds for 4 minutes and the weight changes were logged, this can be seen in Figure B.2. The weight change after 4 minutes at 25% was given to the INFORS system which calculated a pump factor, f to convert the logged pump data, the number of pump rotations, to weight change.

$$f = \frac{\text{Value}}{\text{Duration}} \tag{B.1}$$

where *Value* is the delivered quantity in weight (or mL) and *Duration* is the number of pump revolutions.

The chosen pump factors were,

$$f_{acid} = 0.00549003 \qquad f_{base} = 0.00578613 \tag{B.2}$$

as these matched closest to the observed weight change in Figure B.2.

(a) Acid pump calibration

(b) Base pump calibration
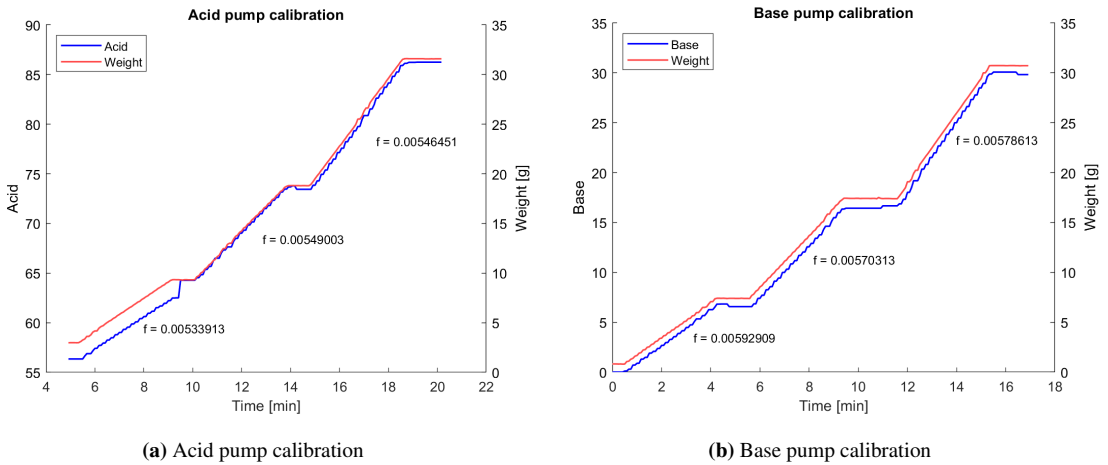
Figure B.2: Acid and base calibration. f is the pump factor used to convert the number of pump rotations to a weight change. The logged data weight does not necessarily start at 0, as seen in *(a)*, as the system logs *all* liquid added through the current Lucullus session.

B.6   Feeding profile NIR - CDW - OD600 calibration

The CDW calibration is explained in detail in Section 4.3.2. The feeding profile used for the calibration experiment is seen in Figure B.3. The profile was found using MPC (Equation (3.6)), and then a constant flow was added from 4-9+h to ensure that sugar was in surplus in the system for efficient growth.
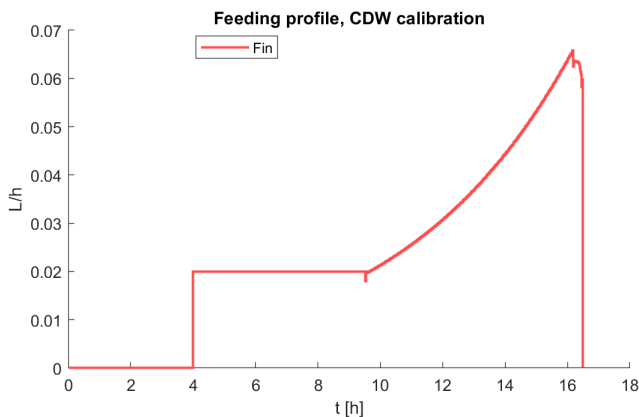
Figure B.3: Input profile (feed) used for calibration of the OD probe, CDW and OD600. The feeding profile was made with simulation using MPC with setpoint tracking of biomass at 15 g/L and parameters in Table 5.1. A constant flow of feed at 0.2L/h was added from hour 4 to 9.5 to account for model mismatch and ensure sufficient sugar level in the broth for growth.

B.7 CDW-OD600-NIR calibration

Simultaneously as the NIR and CDW calibration was performed, the relation between OD600 and CDW was also found. This can be useful if you want to begin the experiment at a specific CDW, and not just aim for OD 1 in the reactor, as we have done in this thesis. The calibration curve was found to be the following,

OD600 <= 15

$$CDW = 1.1241 \cdot \text{OD600} - 0.65028$$

OD600 > 15

$$CDW = 1.0372 \cdot \text{OD600} + 0.022053$$
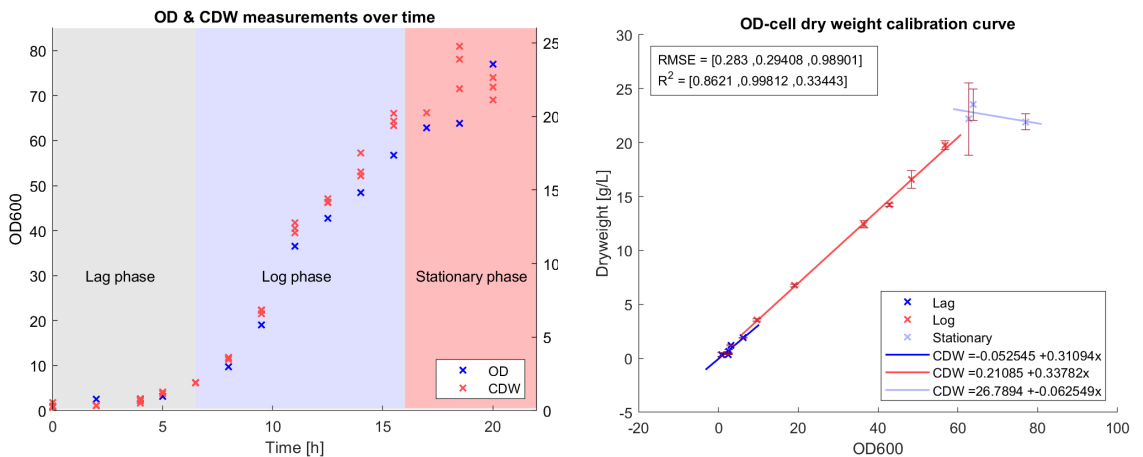
(a) OD600 vs. CDW over time. Blue marks are OD600 measurements, and red marks are CDW measurements.

(b) OD-CDW calibration curve

Figure B.4: Experimental measurements and the linear calibration curve between OD600 and CDW. The dry weight measurements are plotted as means, including their variance. The RMSE and R^2 for the linear fits are also included in the figures. Based on the dynamics seen from the NIR measurements, the samples are divided into 3 phases, lag, log and stationary. Only the two first phases should be used during a run (due to negative slope in the last one.)

A linear regression was also performed for CDW-NIR, which can be seen in appendix B.7, but a sinusoidal curve was chosen as it had a lower RMSE.

NIR <= 0.9

$$CDW = 7.959 \cdot \text{NIR} - 0.11656 \tag{B.3}$$

NIR > 0.9

$$CDW = 22.2695 \cdot \text{NIR} - 8.1916 \tag{B.4}$$
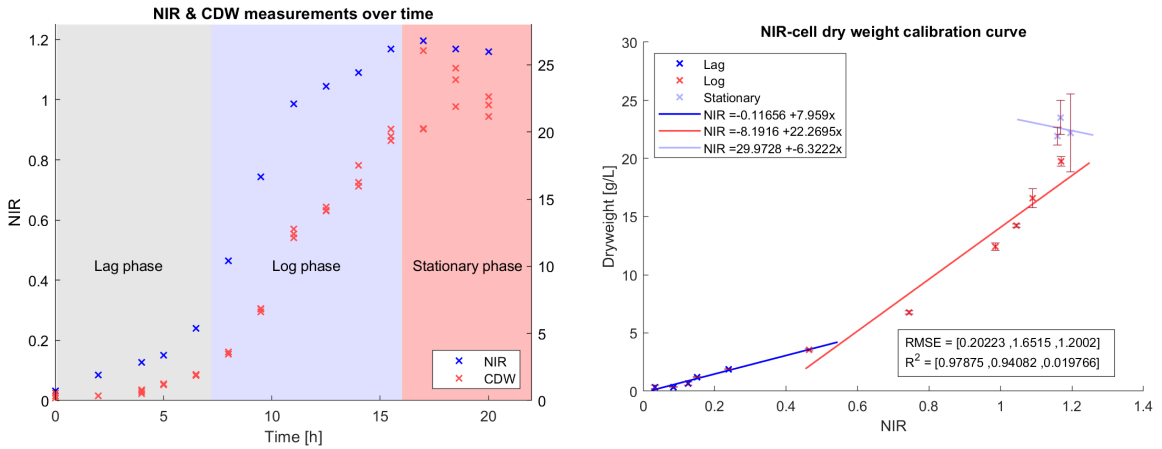
(a) OD600 vs. NIR over time. Blue marks are NIR measurements, and red marks are CDW measurements.

(b) NIR-CDW calibration curve

Figure B.5: Experimental measurements and the linear calibration curve between NIR and CDW. The dry weight measurements are plotted as means, including their variance. The RMSE and R^2 for the linear fits are also included in the figures. Based on the dynamics seen from the NIR measurements, the samples are divided into 3 phases, lag, log and stationary .

The RMSE is highest for the log phase for both calibration curves, the average deviation between the measurement and the linear fit. The R^2 for the lag and log phase tells us that the linear fit is able to explain around 90% of the variation in CDW. If this linear calibration curve is used, only the two first sections should be used, as seen in Equations (B.3) and (B.4). A calibration for the same strain was performed in Tuveri et al. (2021). For the linear fit, the slopes are similar, but the intercept is slightly different.

B.8 Offgas calculation

The offgas measurements show the CO_2 produced by the cells in the bioreactor. We assume that there is no transfer of CO_2 from the gas to the liquid phase, only from the liquid to the gas phase. The total volume of the reactor, V_{total}, is 2.7 L and the airflow into the reactor, q_{air} is constant at $2L/min$.
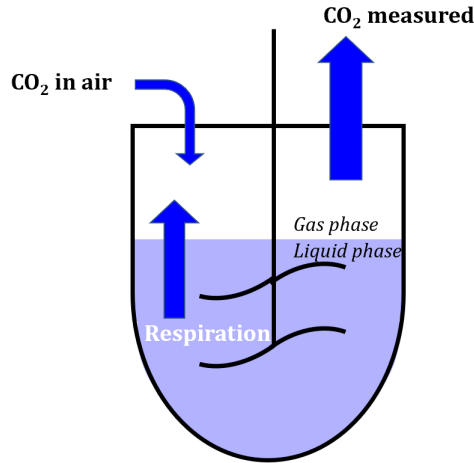
Figure B.6: The figure shows the flows and sources of CO_2 in the system. There is a small amount of CO_2 in the air inlet stream (0.04%) which also will be measured by the BlueInOne BlueSens gas sensor. CO_2 is also generated by the *C. glutamicum* cells in their aerobic respiration, an indication of growth.

To find the CO_2 exerted from the biosystem due to the metabolism of the bacteria, we need to take into account the CO_2 already present in the air. In general CO_2 makes up 0.04% of the air. Adding a bit of margin, if the CO_2 measurement is below 0.06% we assume CO_2 in the air to just be the measurement itself. The density of CO_2, ρ_{CO_2}, is $1.7\overline{7}$ g/L at atmospheric pressure and 30 °C (Evans, 2020). The liquid volume is constant at 1.5 L. In the following equations we consider CO_2 with the following unit: L_{CO2}/L_{air}. The measurements of CO_2 are made by the BlueInOne BlueSens gas sensor, described in Section 4.2.2.

The CO_2 in the system can be described by,

$$q_{air} \cdot CO_2 = q_{air} \cdot 60 \cdot CO_2 \cdot \rho_{CO_2} \frac{CO_{2\text{BlueInOne}}}{100 \cdot (V_{tot} - V_{liquid})}$$

Where we multiply by 60 to convert from per minute to per hour. By withdrawing the amount of CO_2 in the inlet air the final equation is,

$$y_{\text{CO}_2} = \rho_{CO2} \cdot 60 \frac{CO_{2\text{BlueInOne}} - CO2\text{air}}{100 \cdot (V_{tot} - V_{liquid})} \tag{B.5}$$

The calculations are based on Krämer and King (2019).

B.9   Flow during experimental runs

The intended feeding profile for the experimental runs for parameter estimation and MHE is shown in Figure 5.4. The experimental input profile is shown in Figure B.7. The deviation in the two mentioned input profiles is due to unforeseen events during the experimental run. Firstly, after 4 hours we noticed the sugar measurements from the HPLC were not logged in Matlab. After debugging a breakpoint was not removed, leading to a 2.9-hour stop in the planned input profile and data logging. To account for the sugar feed being delayed by 3 hours, some feed was manually added. The missing data was after the experiment manually taken from Lucullus into an Excel file and merged with existing data. The parameter estimation experiment was not run again as it was time critical to begin parameter estimation for the MHE. At this point several attempts at a "good" experiment had failed, so we chose to move on.

The deviation from the planned flow out in Figures B.7, B.8a and B.8b is due to the feed not properly being pumped out of the system. There was an issue with the tube out from the reactor, and even at max rpm, it was difficult to remove liquid to get to V = 1.5 L. One cause of the problem may be the density of the liquid (although it worked fine with water). Another may be equipment malfunction - in the outlet pipe or tubing, or changes after autoclavation.
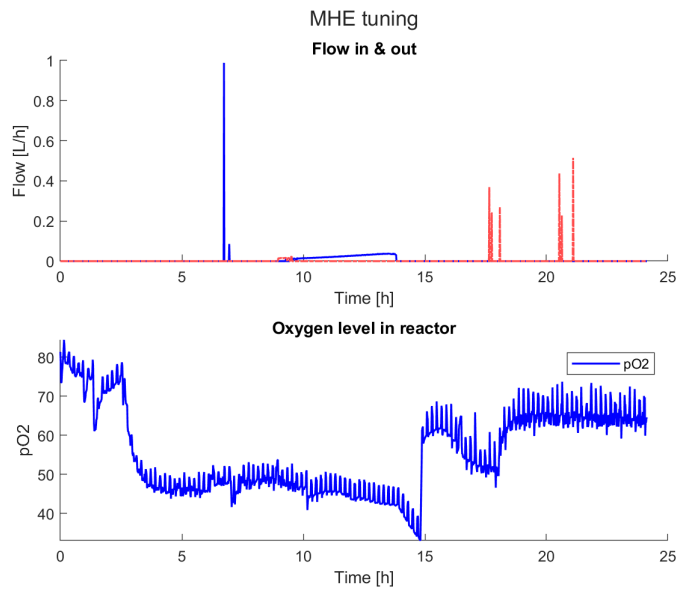
Figure B.7: The figure shows the experimental input profile used for parameter estimation and MHE tuning and the measured pO2 level. The batch phase is 6 hours, following an increase in feed, until around 11 hours when the feed is kept constant at 0.0178 L/h. The volume is not kept constant at 1.5 L due to issues with outlet

(a) First real-time MHE run.

(b) Second real-time MHE run.
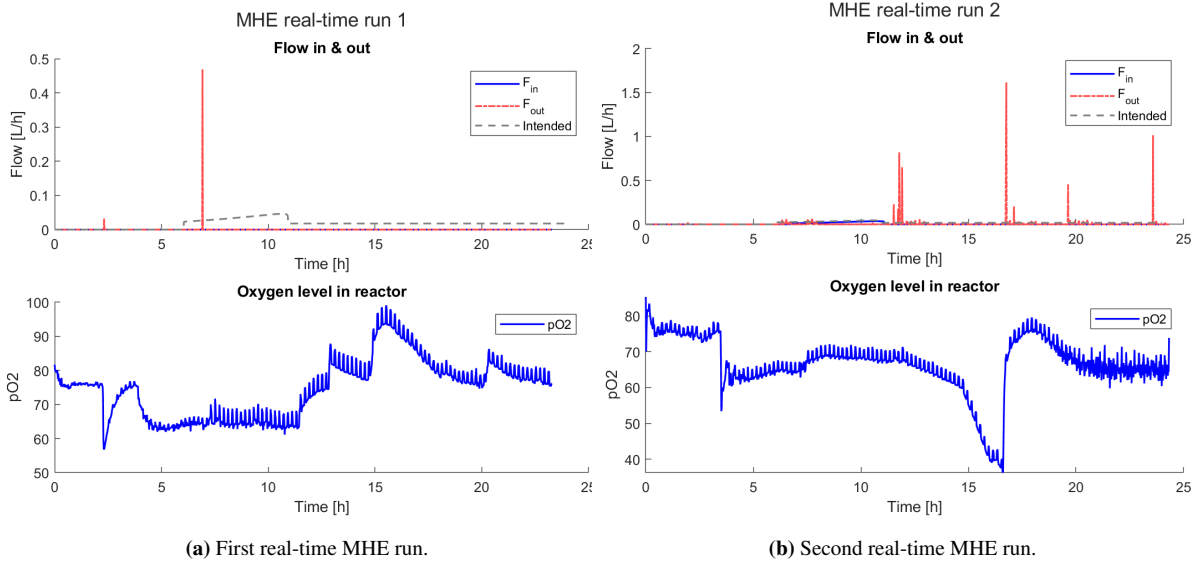
Figure B.8: The figure shows inlet and outlet flow, and O2 for first and second real-time MHE run.

B.10 Media compososition

2TY

volume of solution	1000	mL	H20	
component	conc.	unit	measure	unit
Tryptone	16	g/L	16	g
Yeat Extract	10	g/L	10	g
NaCl	5	g/L	5	g

~

Sterilization   Autoclave

Store   Room temp

* Always check if contaminated before use

Tetracycline

volume of solution	100	mL	EtOH	
component	conc.	unit	measure	unit
Tethracyclin	10	g/L	1	g

This has to be diluted in EtOH 70%

Sterilization   Filtration

Store   at -20°C

Biotin Solution

volume of solution	100	mL		
component	conc.	unit	measure	unit
Biotin	0.2	g/L	0.02	g
NaOH 1M		up to	7	pH
water		up to	100	mL

Sterilization   Filtration

Store   at -20°C

Essential vitamin

Trace elements Solution

volume of solution	100	mL	H2O	
component	conc.	unit	measure	unit
FeSO4 x7 H2O	16.4	g/L	1.64	g
MnSO4 x H2O	10	g/L	1	g
ZnSO4 x7 H2O	1	g/L	0.1	g
CuSO4 x5 H2O	0.31	g/L	0.031	g
NiCl2 x6 H2O	0.02	g/L	0.002	g

pH with HCl   1

First correct the pH and then bring it to volume.

Sterilization   Filtration   Falcons 5O ML

Store   at 4°C

important minerals needed in small amoiunts

Ca-stock 1000X

volume of solution	100	mL	H2O	
component	conc.	unit	measure	unit
CaCl2 x2 H2O	13.25	g/L	1.325	g

Sterilization   Filtration

important role in signal transduction pathways

Mg-stock 1000X

volume of solution	100	mL	H2O	
component	conc.	unit	measure	unit
MgSO4 x7	250	g/L	25	g

Sterilization   Filtration

growth and metabolic functions

IPTG stock 1M

volume of solution	100	mL	H2O	
component	conc.	unit	measure	unit
IPTG	2.38	g/L	0.238	g

Sterilization   Filtration   Falcons 5O mL

Store   at -20°C

inducer

Please note that after filtration the biotin,
Pks and IPTG could be aliquoted in smaller volumes for an easier handling

PKs solution (Protocatechuic acid or 3,4-dihydroxybenzoic acid)

volume of solution	50	mL		
component	conc.	unit	measure	unit
PKs	30	g/L	1.5	g
NaOH 1M		up to	50	mL

Sterilization   Filtration

Store   at -20°C

Agar plates

volume of solution	2500	mL	H2O	~20 plates
component	conc.	unit	measure	unit
2TY		g/L	2	L	
Agar solution	15	g/L	0.5	L	premade 15%

Sterilization   Already sterile

Store

For streaking on agar plate

Batch/Fed-Batch Fermentation Broth

total volume	1500	mL

Fermentation Broth Composition

CGXII	75	%	1125	
C-source	20	%	300	15 g/L glucose
Inoculum	5	%	75	

Anthibiothics (yes/no)

Kanamycin	no
Spectinomycin	no
Tetracycline	no
IPTG	no

Please note that all the cells with this colour are calculators!

Batch/Fed-Batch Reactor - CGXII

volume of solution	1125	mL	H2O

component	conc.	unit	measure	unit
(NH4)2SO4	20	g/L	30	g
Urea	5	g/L	7.5	g
KH2PO4	1	g/L	1.5	g
K2PO4	1	g/L	1.5	g
Ca-stock 1000X	1	mL/L	1.5	mL
Mg-stock 1000X	1	mL/L	1.5	mL

pH correction with KOH   7

First correct the pH and then bring it to volume.

Sterilization   at 121°C for 20 minutes

For Broth: **Biotin** and **Trace element** solution

are added with Inoculum and C-source at experiment start

For Washing: No extra additives

FEEDING SOLUTION Batch/Fed-Batch Reactor - CGXII (modified)

volume of solution	1500	mL	H2O

component	conc.	unit	measure	unit
Glucose	200	g/L	300	g
(NH4)2SO4	**50**	g/L	75	g
Urea	5	g/L	7.5	g
KH2PO4	1	g/L	1.5	g
K2PO4	1	g/L	1.5	g
MgSO4	10	g/L	15	g
Ca-stock 1000X	1	mL/L	1.5	mL
Mg-stock 1000X	1	mL/L	1.5	mL
Trace Element solution	1	mL/L	1.5	mL
Biotin solution	1	mL/L	1.5	mL

pH correction with KOH   7

First correct the pH and then bring it to volume.

Sterilization   Filter with large filter under vaccume

Carbon source 15 g/L for broth

volume of solution	300	mL	H2O

component	conc.	unit	measure	unit
Glucose	75	g/L	22.5	g
Concentration in broth	15.00	g/L		

Carbon source *in broth media* should normally be 10-20 g/L

!! DONT MIX ALL SUGAR AND WATER AT ONCE

!Use glass beaker and low heat with stirrer

92

B.10.1 Purpose chemicals and media

Table B.2: The table presents all compounds used in the bioreactor experiments. Their name, chemical formula, the media or solution, and a brief explanation of its role is presented (Madigan et al., 2018b). CGXII* : Enriched CGXII

Name	Formula	Media	Role
Glucose	$C_6H_{12}O_6$	Feeding, CGXII	Carbon source
Ammonium sulfate	$(NH_4)_2SO_4$	CGXII	Nitrogen source, sythesis of amino acids
Urea	CH_4N_2O	CGXII	pH buffer and nitrogen source
Monopotassium phosphate	KH_2PO_4	CGXII	Potassium and phosphate source, osmoregulation, pH homeostasis, regulation of protein synthesis, enzyme activation,
Dipotassium hydrogen phosphate	K_2HPO_4	CGXII	membrane potential adjustment and electrical signaling
Magnesium sulfate	$MgSO_4$	CGXII*	-"-
Calcium chloride	$CaCl_2$	CGXII	Ionic balance
Biotin	$C_{10}H_{16}N_2O_3S$	CGXII	Ionic balance, maintenance of cell structure, motility, transport and cell differentiation processes Essential cofactor for enzymes
Ferrous sulphate heptahydrate	$FeSO_4 \cdot x\,7\,H_2O$	CGXII	Trace Element
Manganese sulfate monohydrate	$MnSO_4 \cdot x\,H_2O$	CGXII	Trace Element
Zinc sulphate heptahydrate	$ZnSO_4 \cdot x\,7\,H_2O$	CGXII	Trace Element
Copper(II)sulfate pentahydrate	$CuSO_4 \cdot x\,5\,H_2O$	CGXII	Trace Element
Nickel chloride hexahydrate	$NiCl_2 \cdot x\,6\,H_2O$	CGXII	Trace Element
Tryptone		2TY	Source of amino acids (including essential ones)
Yeast Extract		2TY	Nitrogen source, promotes cell growth
Sodium Chloride	NaCl	2TY	Allows for adjustments in osmotic pressure

Parameters for MPC

Model parameters used for the MPC are from Tuveri et al. (2021) for Fed-Batch cultivation of *C.glutamicum*, and listed in Table C.1.

Table C.1: Value of model parameters in equation (2.1) with the unit and standard deviations taken from Tuveri et al. (2021).

Parameter	Description	Value	Unit
μ_{max}	Maximum growth rate	0.19445	$[\text{h}^{-1}]$
K_S	Monod growth constant	0.007	$[\text{g} \cdot \text{L}^{-1}]$
k_d	Death rate constant	0.006	$[\text{h}^{-1}]$
Y_{XS}	S from X yield	0.42042	$[\text{g} \cdot \text{g}^{-1}]$
$Y_{X\text{CO}_2}$	CO_2 from X yield	0.54308	$[\text{g} \cdot \text{g}^{-1}]$

Appendix D

Protocols for bioreactor

D.1   Protocol for Infors reactor setup and dismantling

	NTNU	**PROCEDURE**	**Last revision :** **21/Dec/2021**
Faculty of Natural Sciences and Technology		**Title: Protocol for Infors reactor setup and dismantling**	**Page 1 of 15**
Department: IKP		**Procedure number: 1**	
Revision number: 1		**Developed by: Dr. Pedro Antonio Lira Parada (Researcher)**	
		Approved by: Dr. Nadav Bar (P.I.)	
		Maintained by: Dr. Christopher Sørmo	

1. PURPOSE

The purpose of the following protocol applies to the Infors Bioreactor systems, the bioreactors are useful to conduct cell cultures monitor, control and optimize the bioprocess. The system can measure pH, dissolved oxygen (pO_2), biomass, temperature, and implement control structures with the stirrer, Antifoam probe, and it can sample systematically sample with the incorporation of the Numera system.

Bioreactors are the heart of the industrial biotechnology, and the operation is omnipresent in the brewing processes, pharma industry, wastewater treatment, biofuel production, and cell cultures. Different reactor configurations and control strategies have been explored in cell cultures, for instance: work in the area of monitoring, simulation and control in penicillin production [1,2], analysis in mammalian cell cultures [3], non-linear control strategies based on exact linearization [4], model predictive control in bioprocess [5,6], integration of process engineering, fermentation, enzyme and metabolic engineering in ethanol optimization [7], in baker's yeast production [8] amino acid synthesis [9], and analysis of CO_2 and byproduct concentration levels [10–12]. Bioreactors require adequate feeding strategies and a control scheme to obtain a reasonable trade-off between biomass accumulation, yields and productivity [13]. The process require fine-tuning of parameters, conditions and adequate control strategies [14,15]. Traditional bioreactor models usually consider sets of kinetic equations that consider microorganism growth and concentrations in their specific media [16].

In the following, we start by describing the minimum requirements for mounting and dismantling the bioreactor. Following that, we describe the add-on hardware, and then proceed to describe the software, and their connection with the hardware as a complete system.

- The present protocol has as main purpose the safe and **correct** operation of the bioreactor setup and their measurement devices.

- This protocol **needs to be printed** and followed thoroughly during the microbial cell culture.

2. SCOPE

NTNU	Title: Protocol for Infors reactor setup and dismantling
Faculty of Natural Sciences and Technology	
Prosedure number: 1 **Revision no. 1**	**Page 2 of 15**

Correct assembly and connection of the sterile bioreactors to the tower/base, and dismantling operation of the bioreactors

3. TARGET GROUP

- Process system engineering group.
- Biosystems feedback control real-time fermentation laboratory.
- Master students, Ph.Ds., Researchers.

4. DEFINITIONS AND ABBREVIATIONS

5. EQUIPMENT

The minimal set-up configuration consists of a bioreactor with its standard instrumentation that is required for a cell culture set-up. Figure 1 depicts the current instrumentation for one bioreactor in the laboratory set-up. The system has *in-situ* sensors, non-invasive sensors, and the technology to take samples online, measure *at-line*, and off-line monitoring can be conducted with this configuration. The set-up has *at-line* monitoring of sugars with the HPLC and, it is capable to communicate and transfer the information *via* OPC using Lucullus and rest-API (Eve) with Matlab for control implementation. Communication protocols allow to monitor the hardware and define set-points into the bioprocess.
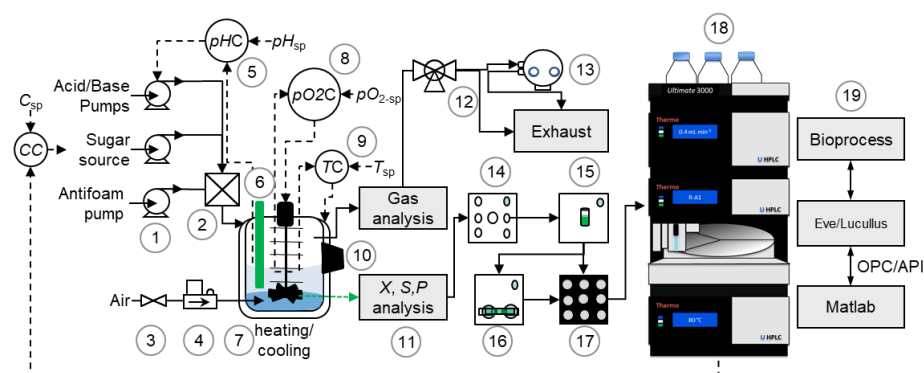
Figure 1. Scheme of the laboratory set-up for the microbial cell culture experiments. The system has 1: a set of acid/Base, sugar source and Antifoam pumps, 2: an inlet liquid flow at the top of the bioreactor, 3: air valve, 4: air mass flow controller, 5: pH Controller (pHC) with a defined set-point (pHsp) that trigger the acid/base pumps, 6: *in-situ* probes including near-infrared (NIR) measurement with an Optek OD probe or the Back Scaterring Buglab probe 7: heating and cooling jacket, 8: pO_2 controller that

	Title: Protocol for Infors reactor setup and dismantling
NTNU **Faculty of Natural Sciences and Technology**	
	Page 3 of 15
Prosedure number: 1	**Revision no. 1**	

has a defined set-point (pO$_2$-sp) that controls the stirring of the Rushton type impeller, 9: temperature controller (TC) with a defined Temperature set-point (T_{sp}), 10: *in-situ* non-invasive measurement devices, gas multiplexer, 13: blue in one sensor that monitors the O$_2$ and CO$_2$ concentration in the gas phase. The down process consists of a Numera system that takes sample from the liquid phase and it consists of a 14: liquid multiplexer that samples from the different bioreactors, 15: a dilution module, 16: filtration module, 17: an auto-sampler that preserves the analytes into a vial. The filtered sample can also be *at-line* monitored in the 18: high performance liquid chromatography (HPLC). The bioreactor setup can be monitored with 19: Eve/Lucullus, and the information can be transferred to Matlab *via* OPC and API.

1. A set of acid/Base, sugar source and Antifoam pumps
2. An inlet liquid flow at the top of the bioreactor
3. Air valve, pressure reducer and air/gas filters
4. Air mass flow controller
5. pH Controller (pHC) with a defined set-point (pHsp) that trigger the acid/base pumps
6. *in-situ* probes including near-infrared (NIR) measurement with an Optek OD probe or the Back Scaterring Buglab probe
7. heating and cooling jacket, and glass reactor vessel
8. pO$_2$ controller that has a defined set-point (pO$_2$-sp) that controls the stirring of the Rushton type impeller
9. Temperature controller (TC) with a defined Temperature set-point (T_{sp})
10. *In-situ* non-invasive measurement devices, that can be adapted from outside of the glass jacket.
11. Sampling tube in the bioreactor
12. Gas multiplexer for systematic monitoring of the gas composition.
13. Blue in one sensor that monitors the O$_2$ and CO$_2$ concentration in the gas phase.
14. The down process consists of a Numera system that takes sample from the liquid phase and it consists of a liquid multiplexer that samples from the different bioreactors.
15. A dilution module
16. A Filtration module
17. An auto-sampler that preserves the analytes into a vial, and the sample can also be *at-line* monitored (after filtration with the HPLC) or offline analyzed.
18. The high-performance liquid chromatography (HPLC).
19. The bioreactor setup can be monitored with Eve/Lucullus, and the information can be transferred to Matlab *via* OPC and API.

	Title: Protocol for Infors reactor setup and dismantling	
NTNU **Faculty of Natural Sciences and Technology**		
Prosedure number: 1	Revision no. 1	**Page 4 of 15**

The vessel / core bioreactor unit

This protocol entails the correct operation of the LabFors V from INFORS, a natural solution because of the complete integral system that INFORS delivers, its industrial experience, the service, and the contribution of INFORS to iFermenter project (6 bioreactors+2 bought by LabNorway). The vessel includes a cover with ports for pH sensor, Antifoam sensor, Exit Gas cooler, Feed inlet, pO_2 sensor, inoculation port, sample port (to be connected to the add-on NUMERA), and other available ports that will be used for add-on probes (OD and gas analyzer). The fermentation vessel has a capacity of 3 L, a good choice for laboratory study, because continuous samples of 2 mL can be drawn every 10 minutes without compromising the dynamics or affecting the process.
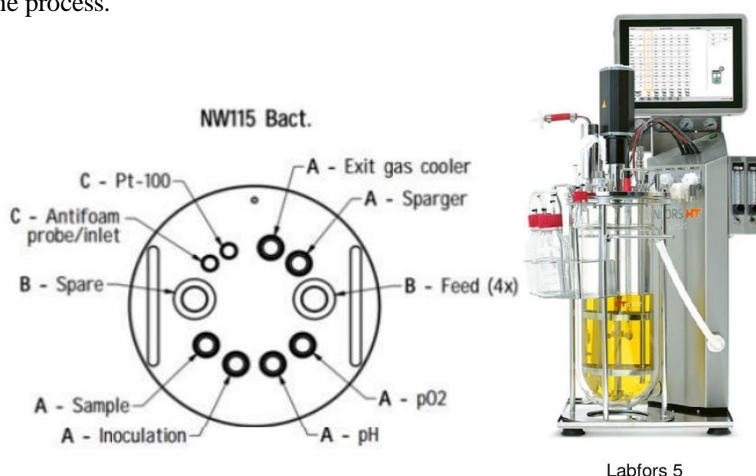
Figure 2. Scheme of the top plate of the Infors-Labfors 5 bioreactor set-up, available ports, and their use during a fermentation. At the right, the bioreactor and its tower is presented with the touchscreen and the basic instrumentation for a microbial fermentation procedure.

6. METHOD

Protocol for Infors reactor set-up and dismantling

6.1 Confirm you have booked the equipment in Bookitlab

6.2 Verify that the reactor is clean and available for the microbial cell culture

Make sure the following elements are clean, not bent, and unclogged. In case, something is not correct refer to Christopher Sørmo about its state.

- Probes: Tª, OD, DO, pH

![NTNU logo]	**NTNU**	**Title: Protocol for Infors reactor setup and dismantling**	
	Faculty of Natural Sciences and Technology		
			Page 5 of 15
Prosedure number: 1		**Revision no. 1**	

- Sampler tube

- Antifoam tube

- Rotor/stirrer (When needed, replace the glycerin of the rotor)

- Metal baffles

Figure 3. Hamilton VisiFerm DO Arc 325 mm that measures the oxygen partial pressure (pO$_2$) in the medium

Figure 4. The Hamilton pH probe monitors the pH in the medium. pH quantification allows to control the supply of acid and base into the bioreactor
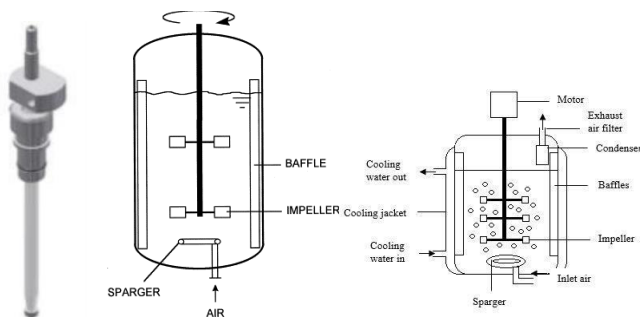
Figure 5. Probe for supplying AF into the system, and schematics that shows the baffles, impeller, the sparge tube for gas supply (air, oxygen or nitrogen)

6.3 Assemble the bioreactor:

Pour the media in the vessel (only the CGXII). Composition of the CGXII can be found in the following references XXX, or in the spreadsheet XXX

	NTNU	Title: Protocol for Infors reactor setup and dismantling		
	Faculty of Natural Sciences and Technology			
Prosedure number: 1		**Revision no. 1**		**Page 6 of 15**

1. Close the lid taking care not to cause damage to the probes and keeping the lid always vertical.
2. Be sure the rubber ring is between the vessel and the lid, and that all the ports had their corresponding O-ring.
3. Screw the lid

Picture(S) instead

NTNU Faculty of Natural Sciences and Technology	Title: Protocol for Infors reactor setup and dismantling		
			Page 7 of 15
Prosedure number: 1	Revision no. 1		

6.4 Acid, base and antifoam bottles:

- Fit the empty bottles for (200 mL of phosphoric acid 10% or Potassium hydroxide 4M) on the rack or 50 mL of antifoam
- Make sure that the hoses are not dirty or clogged.
- Make sure that the filters are not damaged
- When needed, a fourth bottles can be added for the feed
- **Note: we do not autoclave the strong acid and base due to previous experience. Instead the acid & base are added into the empty bottles later after the sterilization.**

Picture(S) instead

6.5 Gas condenser:

- Make sure that the hose is not dirty or clogged.
- Make sure that the filter is not damaged
- Cover with aluminum-foil the tubes and filter

Picture(S) instead

6.6 Autoclave:

(Each round takes 3 hours approx.. in the small autoclave and 2 in the bigger one):

Before autoclaving close with metal clamps the hoses of:

1. super-safe-sampler (if used)
2. air,
3. acid,
4. base,
5. antifoam
6. and feed

1. Before autoclaving cover with aluminum foil the: rotor, super-safe-sampler, all the filters

	NTNU	Title: Protocol for Infors reactor setup and dismantling		
⬜	**NTNU** Faculty of Natural Sciences and Technology			
				Page 8 of 15
Prosedure number: 1		**Revision no. 1**		

2. Before autoclaving open a bit the bottles for acid, base and antifoam
   a. Remove acid and base from bottles
   b. Autoclave with AF detached from reactor.

3. It is needed an escape point for the gas while keeping the sterile conditions. We can use a big filter provisionally placed in a 12mm port on the lid.

4. Make sure that the open tube of the external glass jacket remains open during the autoclavation process.

5. It case of the small autoclave, be sure the autoclave has enough water (it must cover the resistance)

6. Introduce in the autoclave the closed bioreactor, the acid/base/antifoam bottles and the gas collector

7. Close the autoclave and press OK. Wait till it is running. If you see vapor coming out stop it and close the lid of the autoclave properly.

8. When the autoclavation is finished the lid could be opened again when the internal Tª is around 65ºC.

9. When needed, calibrate pumps for the next day.

Picture(S) instead

6.7 Autoclave:

- In case of using the small autoclave to sterilize the bioreactor, you would have to attach the gas collector and remove the big nose.
- Pour in the vessel with sterile funnels the following:

1. carbon source,
2. trace elements,
3. biotin,
4. and when needed antibiotics
5. and IPTG

- Move the bioreactor to the control tower
- Remove all the aluminum foils

Picture(S) instead

	NTNU	**Title: Protocol for Infors reactor setup and dismantling**	
	Faculty of Natural Sciences and Technology		
Prosedure number: 1	**Revision no. 1**		**Page 9 of 15**

6.8 How to set-up the bioreactor with the tower:

- Connect OD, DO and pH probes to the tower
- Place the Tª probe in the right place.
- The antifoam red cable goes in the antifoam needle, the antifoam black cable goes to the lid
- Connect the air hose to the air supplier of the tower
- Connect the hose from the gas collector to the gas output (which is connected to the gas analyzer)
- Turn on the gas analyzer (needs almost 1 hour to warn-up)
- Connect the cold-water input and output hoses to the gas collector
- Connect the warn-water input, output and overflow hoses to the external glass jacket of the vessel
- Connect the motor of the stirrer
- Remove all the metal clamps

Picture(S) instead

6.9 Prepare the fermentation method in Lucullus and/or EVE and/or PC and confirm in the Infors towers:

- Tª= 30ºC (± 0.5ºC)
- pH= 7.0 (± 0.1)
- pO_2= 30% (controlled by the stirrer speed)
- Initial stirrer speed of 200 rpm (set min to 100 rpm and max to 1000 rpm)
- Air supply of 2 NL

When needed, feeding controlled by the DO (pO2 from 30% to above 60% for the first time).
Start all the controllers

Picture(S) instead

	NTNU	**Title: Protocol for Infors reactor setup and dismantling**	
	Faculty of Natural Sciences and Technology		
Prosedure number: 1	**Revision no. 1**		**Page 10 of 15**

6.10 Calibration

- Set speed to 200 rpm and the air flow to 2 L/min or to the specific air flow required by the operation
- Set the OD probe to 0. Go to "calibrate"
- Set the DO probe to 100%.

How: set the air flow to 2 NL (done before), set the stirrer speed to 800 rpm, wait 10-20 min., once the pO2 measurement is stable you can go to "calibrate" and set the value to 100%

- Press the two buttons of the gas analyzer at the same time for few second to initiate its calibration (only if the green like blinks)
- If the pH is not 7 the system will adjust it automatically or you can speed-up the process by controlling the pumps manually

Picture(S) instead

6.11 Prepare inoculum:

- Measure the OD of the preculture in complex media (Day 2)
- Calculate the amount of cell you would need to inoculate your bioreactor with OD600=1 (use C1 * V1 = C2 * V2)
- Collect the needed volume of cells in falcon tubes.
- Centrifuge the cells at 5000 rpm, 10 min and at room temperature (be sure that the centrifuge is balanced)
- Discard the supernatant
- Resuspend the cells in CGXII buffer (WITHOUT carbon source)
- Centrifuge the cells at 5000 rpm, 10 min and at room temperature (be sure that the centrifuge is balanced)
- Discard the supernatant
- Collect all the cells in 50 mL of CGXII buffer (without carbon source). This is your <u>inoculum</u>.

Picture(S) instead

NTNU	Title: Protocol for Infors reactor setup and dismantling		
Faculty of Natural Sciences and Technology			
Prosedure number: 1	**Revision no. 1**		**Page 11 of 15**

6.12 Inoculate the bioreactor:

- Place the inoculum in a 60mL syringe under the clean-bench.
- Cover the top of the syringe with sterile aluminum foil
- With a flame open a lid-port in the bioreactor
- Remove the aluminum foil of the syringe and inject all the cells into the bioreactor
- Close the lid-port and turn off the flame.
- Click Start in the PC
- Check the pO_2 level, if it drops the cells are growing
- Manually pump 2 drops of antifoam in the media
- Take the T_0 sample using the super-safe-sampler

Picture(S) instead

6.13 Sampling

The sampling frequency depends on your experiment

1. strain,
2. type
3. and concentration of the carbon source

C. glutamicum WT with 1% glucose should take 8-10 hours to finish the growth

- Use the super-safe sampler for collecting manually samples.

Take 3 mL every time. 1 mL will be used for OD600 measurements. The other 2 mL will be filtered (0.2 µm pore diameter filters) and stored at 4ºC for further HPLC measurements

- Use NUMERA for automatic sampling.
- Sample for OD measurement – set up NUMERA for collecting 1:10 diluted sample
- Sample for HPLC measurement – set up NUMERA for collecting 1:10 diluted and filtered sample

<u>Option 1</u> – the fermentation is over (typically when the operator defines it)

- Stop the fermentation

- Remove all the connections

- Autoclave the bioreactor

- Clean the bioreactor (this can be done also next day depending of the number of bioreactors)

- Remove the lid and discard the dead cells through the sink

- **ALWAYS KEEP THE METAL LID VERTICAL**

- Clean everything properly with deionized water.

- Check that all the hoses are OK. Renew them once in a while

<u>Option 2</u> – feeding-phase (typically when the sugar is depleted or not consumed anymore)

- The feeding will be controlled by a control structure

This point can be used for step response or single shot of sugar (Data for MPC, modelling and PID)

7. SAFETY EQUIPMENT

Use gloves, lab glasses and lab coat during the operation

NTNU	Title: Protocol for Infors reactor setup and dismantling	
Faculty of Natural Sciences and Technology		
		Page 13 of 15
Prosedure number: 1	**Revision no. 1**	

8. REFERENCES

1.   Golabgir, A.; Herwig, C. Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production. *Chemie Ing. Tech.* **2016**, *88*, 764–776.

2.   Goldrick, S.; Tefan, A.; Lovett, D.; Montague, G.; Lennox, B. The development of an industrial-scale fed-batch fermentation simulation. *J. Biotechnol.* **2015**, *193*, 70–82.

3.   Xing, Z.; Kenty, B.M.; Li, Z.J.; Lee, S.S. Scale-up analysis for a CHO cell culture process in large-scale bioreactors. *Biotechnol. Bioeng.* **2009**, *103*, 733–746.

4.   Henson, M.A.; Seborg, D.E. Nonlinear control strategies for continuous fermenters. *Chem. Eng. Sci.* **1992**, *47*, 821–835, doi:10.1016/0009-2509(92)80270-M.

5.   Macharia, M.A.; Tay, M.E. Model predictive control of fermentation in biofuel production 2013.

6.   Sommeregger, W.; Sissolak, B.; Kandra, K.; von Stosch, M.; Mayer, M.; Striedner, G. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnol. J.* **2017**, *12*, 1600546, doi:10.1002/biot.201600546.

7.   Hahn-Hägerdal, B.; Galbe, M.; Gorwa-Grauslund, M.F.; Lidén, G.; Zacchi, G. Bio-ethanol – the fuel of tomorrow from the residues of today. *Trends Biotechnol.* **2006**, *24*, 549–556.

8.   Frisch, M. ~J.; Trucks, G. ~W.; Schlegel, H. ~B.; Scuseria, G. ~E.; Robb, M. ~A.; Cheeseman, J. ~R.; Montgomery; Jr., A., J.; Vreven, T.; et al. {G}aussian 03 2004.

9.   Leuchtenberger, W.; Huthmacher, K.; Drauz, K. Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnol.* **2005**, *69*, 1–8, doi:10.1007/s00253-005-0155-y.

10.   Kiefer, P.; Heinzle, E.; Wittmann, C. Influence of glucose, fructose and sucrose as carbon sources on kinetics and stoichiometry of lysine production by \textit{Corynebacterium glutamicum}. *J. Ind. Microbiol. Biotechnol.* **2002**, *28*, 338–343.

11.   Paczia, N.; Nilgen, A.; Lehmann, T.; Gätgens, J.; Wiechert, W.; Noack, S. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microb. Cell Fact.* **2012**, *11*, 122.

12.   Sinner, P.; Stiegler, M.; Herwig, C.; Kager, J. Noninvasive online monitoring of \textit{Corynebacterium glutamicum} fed-batch bioprocesses subject to spent sulfite liquor raw material uncertainty. *Bioresour. Technol.* **2020**, e124395, doi:10.1016/j.biortech.2020.124395.

13.   Mears, L.; Stocks, S.M.; Sin, G.; Gernaey, K. V A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *J. Biotechnol.* **2017**, *245*, 34–46, doi:10.1016/j.jbiotec.2017.01.008.

14.   Bornscheuer, U.T.; Huisman, G.W.; Kazlauskas, R.J.; Lutz, S.; Moore, J.C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185, doi:10.1038/nature11117.

15.   Stanbury, P.F.; Whitaker, A.; Hall, S.J. *Principles of fermentation technology*; 3rd ed.; Butterworth-Heinemann: Oxford, 2016;

16.   I.J.Dunn E.Heinzle, J.I.J.E.P. *Biological Reaction Engineering: Dynamic*

Appendix E

Code

Figure E.1 shows the file directory of the zip-file in the Inspera attachment delivery. All code has been written in Matlab. A tenfold more code has been written in the development, however, the attached code to the Inspera delivery shows the essence of the work done, and most importantly the code used to produce results in Chapter 5 and plots in the appendix. All code that simulates (MPC, MHE, parameter estimation) needs access to main folder (colorHex, system parameters). In addition most code that plots or simulates offline needs access to the "Results" folder. The necessary paths are added provided in the top of the files, however they must be rewritten to match your root directory. For the files in the top layer in Figure E.1, the first file, system_params.m, contains all common constants and parameters for the models and simulations. colorHex.m contains colors used for plotting. The ODEsolver is used for integration in the files where the nominal model (Equation (2.1)) is integrated. A file directory for the parameter estimation folder is presented in Figure E.2 and the MHE folder presented in Figure E.3. These two folders are included as they contain many files with different purposeas. The remaining folders are not included as they should be self-explanatory and of less importance.

The following tree shows the structure of the laboratory code for real-time MHE. main_meas_v1_MHE_A2.m is the main file for real-time MHE it contains and uses all the functions listed below it. In the list, functions between GetParValues and restart are functions related to communication with the reactor through Chromeleon, retrieving and saving measurements. The functions between model_DYCOPS and sortResultsMHE_DYCOPS are related to the state estimation using MHE. Note that this function needs feed values, as seen in Figure E.3 showing dependency to the file "Values_feed_09_03_23.mat".

```
/
└── main_meas_v1_MHE_A2.m
    ├── GetParValues
    ├── findReactorTags
    ├── GetInstData
    ├── Analysis_instrumentation
    ├── GetHPLCFromFile
    ├── HPLC_data
    ├── SaveVar2Memory
    ├── plotting_VarMem
    ├── ReadSugarValuesFromChromeleonFile
    ├── FindNewChromeleonFiles_HPLC
    ├── restart
    ├── model_DYCOPS
    ├── ODE_solver
    ├── Process_model
    ├── MHEprops_DYCOPS
    ├── makePlantIntegrator_DYCOPS
    ├── arrivalCostUpdate_DYCOPS
    ├── MHEmodel_DYCOPS
    ├── moving_horizon_estimation_DYCOPS
    └── sortResultsMHE_DYCOPS
```
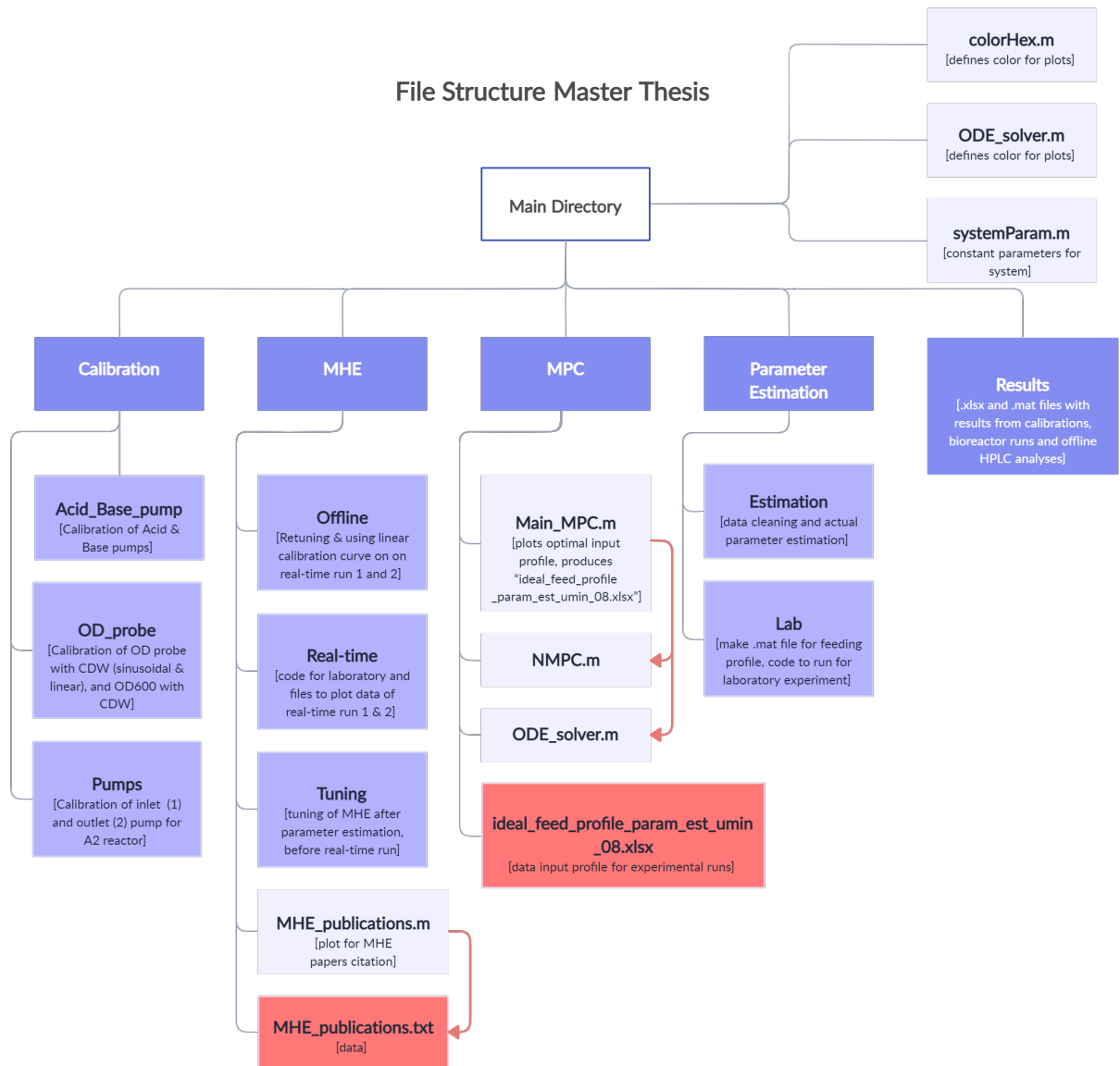
Figure E.1: File directory for the zip file in the Inspera delivery. The dark purple boxes are the folders in the main directory, the slightly lighter purple boxes are folders within the top layer folders, connected to the relevant folder with grey lines. The light purple boxes are matlab (.m) files. Pink boxes are data files - for example .m, .xsxl or .txt files. Dependencies are shown by pink arrows; from a file A to the file B that it needs access to, to run properly. All code that simulates (MPC, MHE, parameter estimation) needs access to main folder (colorHex, system parameters). In addition most code that plots or simulates offline needs access to the "Results" folder. The necessary paths are added provided in the top of the files, however they must be rewritten to match your root directory.
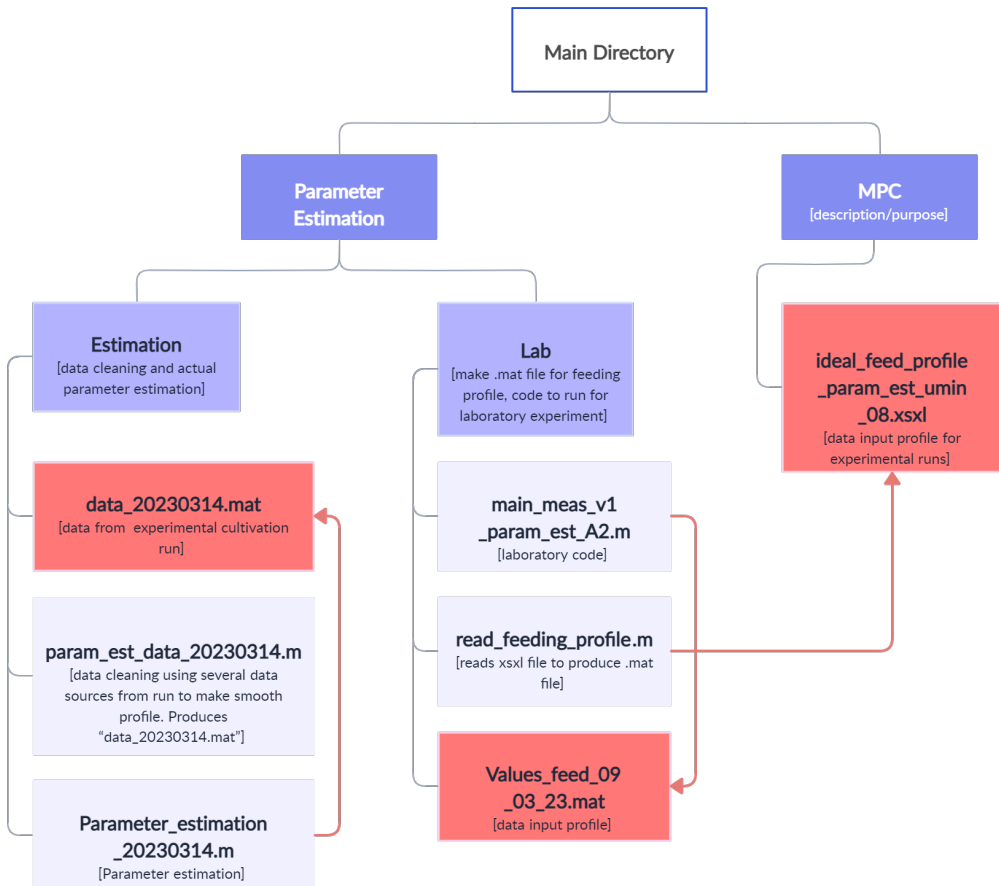
Figure E.2: File directory of Parameter Estimation folder. The overall structure is shown in Figure E.1. Dark purple and purple boxes are folders, while light purple boxes are matlab files, pink boxes are data files. Dependencies are shown with pink arrows (A is dependent on -> B).
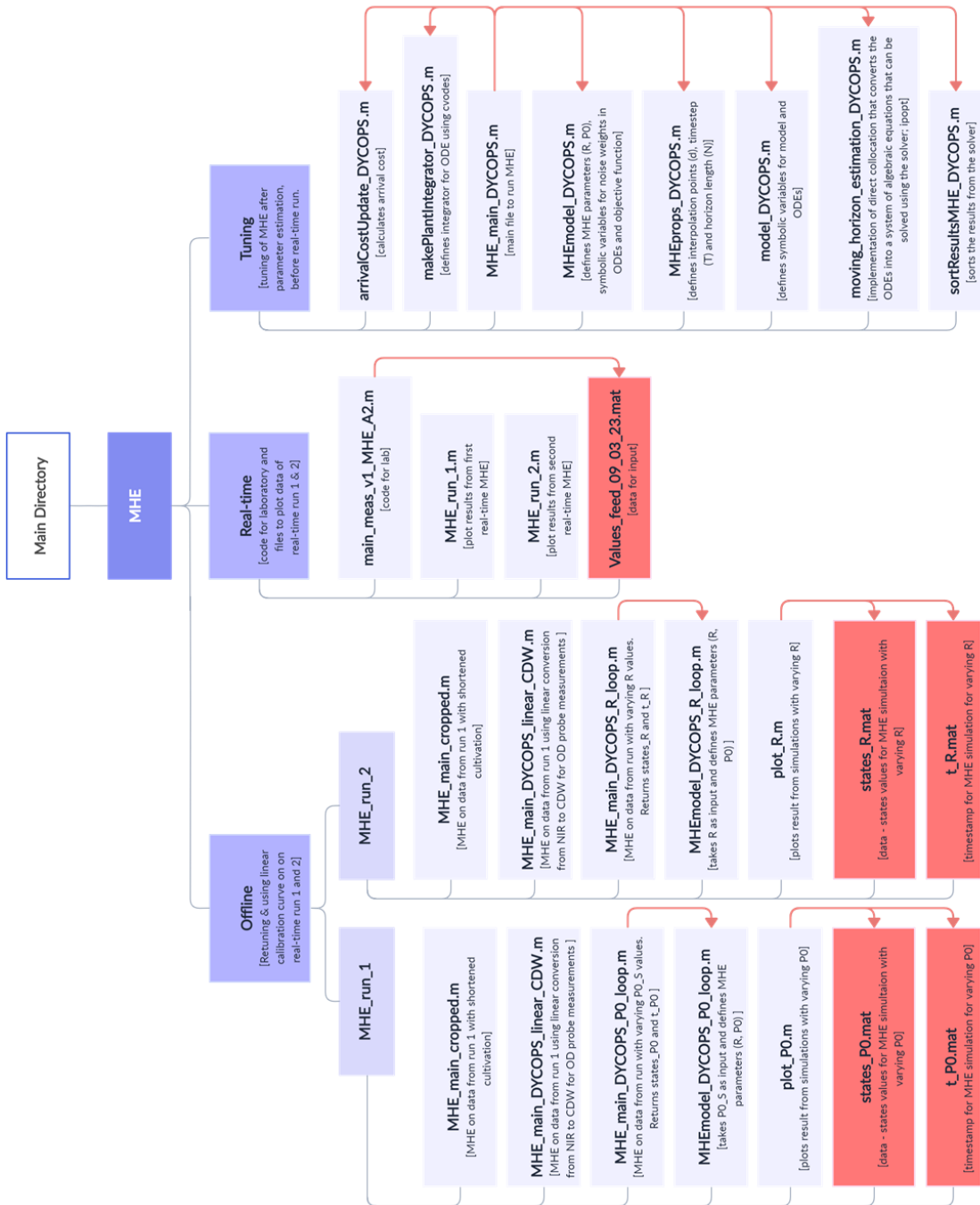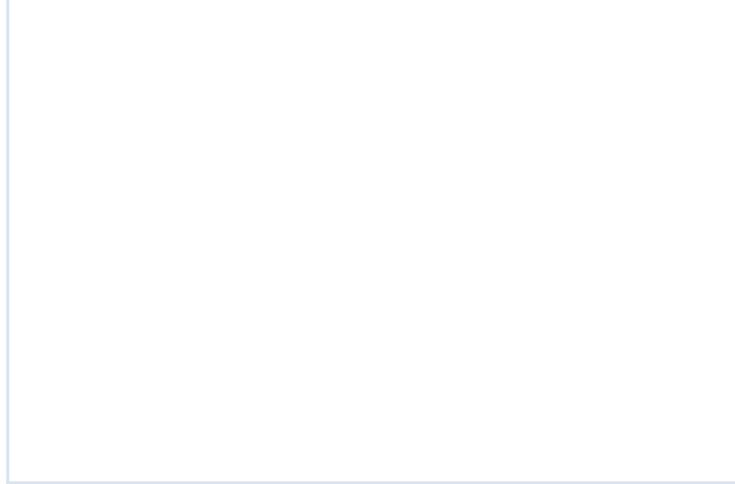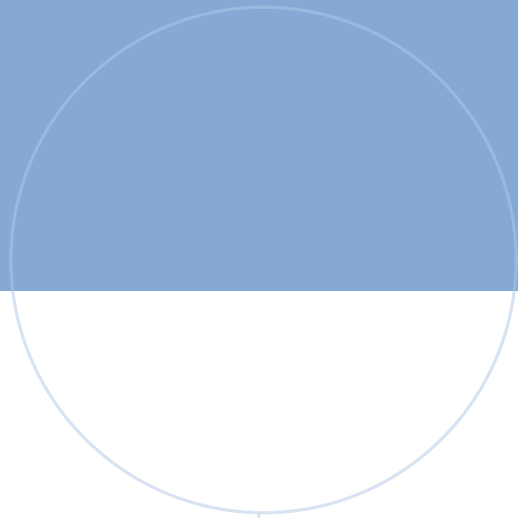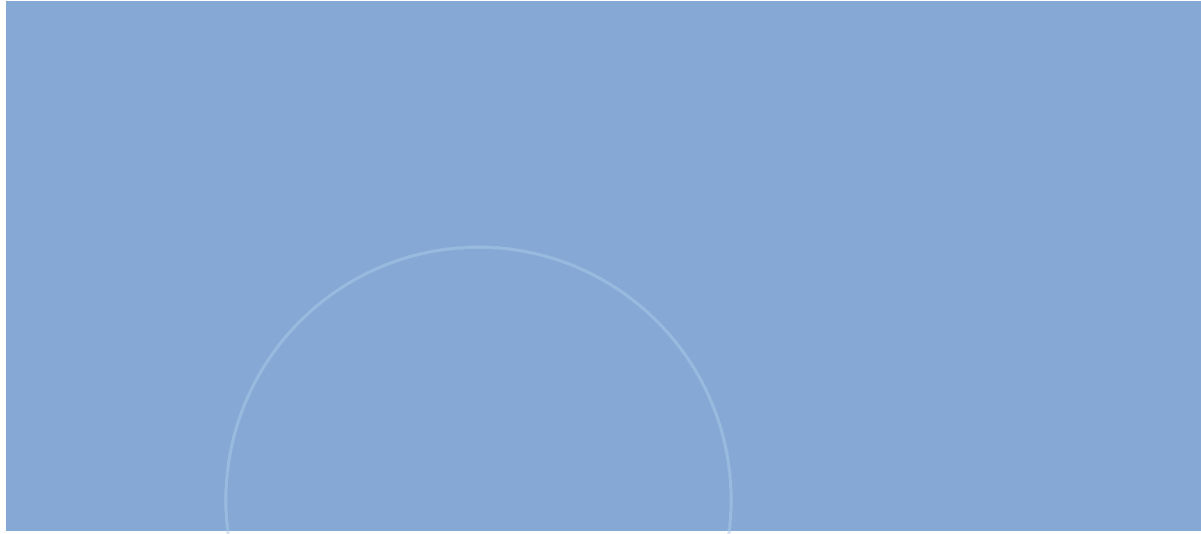
Figure E.3: File directory of MHE folder. The overall structure is shown in Figure E.1. Dark purple and purple boxes are folders, while light purple boxes are matlab files, pink boxes are data files. Dependencies are shown with pink arrows (A is dependent on -> B). The folders "MHE_run_1" and "MHE_run_2" also include the same files as in "Tuning" where all files with "MHE_main" have the same dependencies as "MHE_main_DYCOPS.m".