Real time estimations of $k_L a$ and viscosity in a fermentation pilot plant

DEPARTMENT OF CHEMICAL ENGINEERING | LUND UNIVERSITY
ARVID NORBERG | MASTER THESIS 2022

Real time estimations of k_La and viscosity in a fermentation pilot plant

by

Arvid Norberg

Department of Chemical Engineering, Lund University
in collaboration with Novozymes A/S, Bagsværd

June 2022

Supervisor **Ph.D Niklas Andersson**
Co-supervisor **Ph.D Mads Orla Kaiser-Albaek**
Examiner **Professor Bernt Nilsson**

Postal adress
Box 124
221 00 Lund, Sweden
Web address
http://www.lth.se/chemeng/

Visiting adress
Kemicentrum
Naturvetarvägen 14
223 62 Lund, Sweden

Telephone
+46 46-222 82 85
+46 46-222 00 00

Printed in Sweden by Media-Tryck
Lund 2022

Acknowledgements

This Master Thesis concludes my Master of Science in Engineering, Biotechnology, at Lund University. The thesis was done at the Department of Chemical Engineering at Lund University, in collaboration with Novozymes A/S at their Fermentation Pilot Plant in Bagsværd, Denmark.

I want to begin by thanking Novozymes and Mads Orla Kaiser-Albaek for giving me the opportunity to write my MSc thesis in collaboration with them. Mads has been of great help throughout the thesis project and we have had many rewarding discussions. His curiosity and knowledge have really inspired me to stay curious as well. He was always there and supported me and I am so happy that he wanted to be my supervisor during this thesis project. Without Mads, this project would not have been possible. I must also thank Daniel Sahlin who helped me connect with Mads in the first place.

I cannot begin to express my gratitude to the colleagues at Fermentation Pilot Plant who have really welcomed me as part of their team. They assisted me with everything from introduction to their MATLAB code to giving me advice on my defence presentation. They also gave me ideas on what areas I could explore further, and input on how I could interpret my results. Finally, they inspired me to stay ambitious and curious, and they gave me the joy and comfort needed to enjoy my time at Novozymes. A special thank you to, Benny Cassells, Simoneta Caño de las Heras, Thomas Rydal, Lisa Mears, Anna Johansson, Timo Hagermann, Casper Hermansen, Gisela Nadal Rey, Morten Skov Jacobsen, Mads Fog, Solveig Marie Norup Damgaard, Michael Lynge Nielsen, Carl Wiklund, and Mette Cebolinho Johansen.

Additionally, I want to thank some of the technicians who made the days in the laboratory more pleasant and who helped me with the collection of samples and analysis in the laboratory. Thank you, Nicki Veith Sørensen, Gitte Wilkens Rasmussen, Christina Ewe, and Ebba de Neergaard Harrison. Thank you also to programmer Henning Fugmann who I shared an office with the first month when I worked in MATLAB. He was very helpful with all the code related questions that I had.

I would also like to recognize the invaluable assistance of Niklas Andersson and Bernt Nilsson at the Department of Chemical Engineering at Lund University. They were very helpful, especially during the last month of writing. Niklas helped me with the statistical analyses and encouraged me to stay curious. He also helped me to direct my focus to not include too much into the thesis, which was really appreciated. It was also very pleasant to have Bernt with his positive energy encouraging me when I felt like I would not finish in time. I would also like to express my gratitude to Daniel Espinoza for proof reading my report and to my classmate Hannes Guth who opposed on my report and gave me relevant feedback that I could implement in the thesis.

Finally, to my many friends and family, thank you for your words of encouragements or listening ear. Your support was worth more than I can express on paper.

Abstract

Fermentation is a process that utilizes live microorganisms to produce different biological products such as enzymes. The oxygen mass transfer rate is often the limiting factor in aerobic fermentations and thereby an important process parameter. The initial aim of this MSc thesis was therefore to develop a soft sensor for the volumetric oxygen mass transfer coefficient, k_La, in order to monitor it online in real time. However, the oscillation in the k_La estimation had to be reduced to get smoother plots by first filtering the process data. The soft sensor was finished at an early stage and the scope of the thesis project was therefore expanded.

Previous studies have shown that k_La is dependent on the viscosity of the fermentation broth and that higher viscosities leads to lower values of k_La. It was therefore of interest to monitor the viscosity online as well, and to explore if it would be possible to use the viscosity for implementations of new control concepts. The thesis further explored the relationship between the k_La and the viscosity, and investigateted if it would be possible to describe the viscosity as a function of the k_La using rearranged empirical correlations. The viscosity estimation was validated versus offline measurements and new parameter sets for the correlation were also created and evaluated. The investigation resulted in another soft sensor which estimated the viscosity online with a standard deviation of 0.09 Pa s. It should be noted that the dataset generated in this thesis was not ideal for parameter estimation due to low variance in the process variables. It should therefore be possible to improve the estimation further, for example using previous datasets with larger variance in the process variables.

The substrate is fed in pulses instead of a continuous stream since previous studies have shown that this reduces the viscosity. The third objective of the thesis was therefore to explore how the viscosity was influenced by the changes is paus time between the pulses, i.e., the cycle time, and whether it would be possible to base new control concepts on this relationship. The final objective was to see if it was possible to correlate biomass concentration with rheology parameters including viscosity to explore the possibility of estimating the biomass from online process data. Both of these objectives were investigated but did not yield promising results since no clear correlations could be observed in either case. The relationship between viscosity and cycle time therefore needs to be investigated further before implementation of new control concepts. Further investigation of the relationship between biomass and rheology parameters is also needed to get a better insight in the processes. This could result in relevant insight in how the rheology of a specific host organism and process is correlated to the biomass. It does however seem unlikely that a general correlation could be found for any host organism and process.

Sammanfattning

Fermentering är en process som nyttjar levande mikroorganismer för att producera olika biologiska produkter som t.ex. enzymer. Överföringen av syre är ofta den begränsande faktorn vid aeroba fermentationer och är därmed en viktig processparameter. Det initiala syftet med det här examensarbetet var därför att utveckla en sk. soft senor för det volymetriska massöverföringstalet för syre, k_La, för att på så sätt övervaka denna online i realtid. Oscillationen i k_La- estimeringen måste dock reduceras för att få jämnare grafer genom att först filtrera processdatan. Soft sensorn färdigställdes i ett tidigt skede och således vidgades examensarbetets omfattning.

Tidigare studier har visat att k_La är beroende av fermenteringsvätskans viskositet och att högre viskositeter leder till lägre värden på k_La. Det var därför av intresse att även övervaka viskositeten online och att undersöka ifall det skulle vara möjligt att använda viskositeten för implementeringar av nya kontroll- koncept. Vidare så undersöktes sambandet mellan k_La och viskositeten och ifall det skulle vara möjligt att beskriva viskositeten som en funktion av k_La med hjälp av omvända empiriska korrelationer. Viskositetsuppskattningen validerades sedan mot offlinemätningar. Nya parameterar till viskositets- korrelationen skapades och utvärderades också. Undersökningen resulterade i ytterligare en soft sensor som uppskattade viskositeten online med en standardavvikelse på 0,09 Pa s. Det bör dock noteras att datan som genererades i detta examensarbete inte var perfekt att basera en parameteruppskattning på, eftersom det var låg varians i processvariablerna. Det borde därför vara möjligt att förbättra uppskattningen ytterligare, till exempel genom att använda tidigare genererad data med större varians i processvariablerna.

Substratet tillförs fermentringen i pulser istället för i en kontinuerlig ström eftersom tidigare studier visat att detta skulle minska viskositeten. Det tredje syftet med examensarbetet var därför att utforska hur viskositeten påverkas av paustiden mellan pulserna och om det skulle vara möjligt att basera nya kontrollkoncept på detta samband. Det slutliga målet var att se om det var möjligt att korrelera koncentrationen av biomassa med reologiparametrar inklusive viskositet, för att undersöka möjligheten att uppskatta biomassan från onlinedata. Båda dessa mål undersöktes men gav inga lovande resultat eftersom inga tydliga korrelationer kunde observeras i något av fallen. Förhållandet mellan viskositeten och paustiden behöver därför undersökas ytterligare innan nya kontrollkoncept kan implementeras. Ytterligare utredning av sambandet mellan biomassa och reologiparametrar behöver också göras för att få en bättre insikt i processerna. Detta skulle kunna resultera i relevanta insikter i hur reologin hos en specifik mikroorganism och process är korrelerad till biomassan. Det verkar dock osannolikt att en generell korrelation skulle kunna hittas för godtycklig mikroorganism och process.

Popular Science Summary

What if you could calculate different important process parameters in real time using data that are already being generated. Then you would not need to install new expensive intruments or waste time on labwork. It would also enable implementation of new control concepts. This MSc thesis project produced two new functions that does just that.

Enzymes are a part of the solution in the sustainable transition, but there are many different paramaters to account for in production. They are produced industrially by live microorganisms that are cultivated in large fermentation tanks, where they float around in a liquid with different nutrients etc. Microorganims need oxygen just as humans, and air is therefore sparged into the tanks. The oxygen is transported from the air, to the liquid, and then to the microorganisms. This transport is crucial for the productivity of the process. A function which estimates the oxygen transport in real time was therefore created using different process data, enabling online monitoring of this important process parameter.

However, the oxygen transport is affected by the thickness of the liquid. It is, for example, more difficult to sparge air through syrup than through water. Since you want to increase oxygen transport to maximize productivity, you also want the liquid to be as thin as possible. However, it is very difficult to measure the thickness of a liquid in real time. Different models which were compared to experimental data was therefore created, resulting in a function that actually can estimate the thickness of a liquid containing microorganisms. This important process parameter can therefore also be monitored online.

With these two new functions which estimate oxygen transport and liquid thickness, Novozymes now have two new tools which they can use to get a better understanding of their processes. However, more research is required before implementation of new control concepts.

Populärvetenskaplig sammanfattning

Tänk om man kunde räkna ut olika viktiga processparametrar i realtid med hjälp av data som ändå redan genererats. Då hade man sluppit installera dyra instrument eller ödsla en massa tid på att stå i labbet. Det hade också möjliggjort införandet av nya kontrollkoncept. Det här examensarbetet resulterade i två nya funktioner som gör just detta.

Enzymer är en del av lösningen i den hållbara omställningen, men det finns många olika paramterar att ta hänsyn till i tillverkningen. De produceras industriellt av mikroorganismer som odlas i stora jäskärl, där de flyter runt i en vätska med bl.a. olika näringsämnen. Mikroorganismer behöver syre precis som vi människor och därför bubblas luft in i jäskärlen. Syret transporteras då från luften till vätskan och sedan till mikroorganismerna. Den här transporten är helt avgörande för processens produktivitet. Det kontruerades därför en funktion som uppskattar syretransporten i realtid med hjälp av olika mätdata. På så vis kan man enkelt övervaka denna viktiga processparameter.

Syretransporten påverkas dock av hur trögflytande vätskan är. Det går t.ex. sämre att bubbla luft genom sirap än genom vatten. Eftersom man vill öka syretransporten för att maximera produktiviteten vill man samtidigt att vätskan ska vara så tunn som möjligt. Det är dock väldigt svårt att mäta hur trögflytande en vätska är i realtid. Därför konstruerades det olika modeller som jämfördes med expermintella data, vilket sedan resulterade i en funktion som i realtid faktiskt kan uppskatta hur trögflytande en vätska med mikroorganismer är. Således kan nu även denna viktiga processparameter enkelt övervakas.

Med hjälp av dessa två funktioner som uppskattar syretransport och hur trögflytande en vätska är, har Novozymes nu två nya verktyg som de kan använda för att få en bättre förståelse för sina processer. Mer forskning krävs dock innan nya kontrollkoncept kan införas.

Nomenclature

Roman letters

a	*Gas-liquid interfacial area per liquid volume.*	[m^2/m^3]
a, b, c, K	*Parameters for the k_La-viscosity correlations.*	
C	*Parameter for the k_La correlation without viscosity.*	
D_i	*Impeller diameter.*	[m]
D_T	*Tank diameter.*	[m]
$F(\alpha, P, N-P)$	*Fischer's F distribution with uncertainty $1-\alpha$, number of parameters P, and $N-P$ degrees of freedom.*	
F_{air}	*Air flow rate.*	[$m^3\ s^{-1}$]
F_{normal}	*Air flow rate at normal pressure and temperature.*	[$NL\ min^{-1}$]
F_{air}^{in}	*Inlet air flow rate.*	[$m^3\ h^{-1}$]
F_{air}^{out}	*Outlet air flow rate.*	[$m^3\ h^{-1}$]
g	*The acceleration of gravity*	[$m\ s^{-2}$]
H_O	*Henrys constant for oxygen in water.*	[$Pa\ m^3\ mol^{-1}$]
J	*Jacobian matrix.*	
J_O	*Overall molar flux of oxygen from the gas bubble to the bulk liquid.*	[$mol\ h^{-1}\ m^{-2}$]
$J_{O,G}$	*Molar flux of oxygen through the gas layer.*	[$mol\ h^{-1}\ m^{-2}$]
$J_{O,L}$	*Molar flux of oxygen through the liquid layer.*	[$mol\ h^{-1}\ m^{-2}$]
k_b, k_{pl}, k_{hb}	*Rheology consistency index for the Bingham, Power law and Herschel–Bulkley model, respectively.*	
k_G	*Mass transfer coefficient through the gas layer.*	[$mol\ Pa^{-1}\ h^{-1}\ m^2$]
K_L	*Overall mass transfer coefficient from the gas bubble to the bulk liquid.*	[$m\ h^{-1}$]
k_L	*Mass transfer coefficient through the liquid layer.*	[$m\ h^{-1}$]
k_La	*Volumetric oxygen mass transfer coefficient.*	[h^{-1}]

k_s	*Metzner and Otto's shear rate constant = 11 throughout the thesis.*	
m	*The weight of the fermentation broth.*	[kg]
$M(x,p_i)$	*Model response at any variable vector x with parameters p_i.*	
N	*Number of experimental measurements.*	
N_i	*Number of impellers.*	
n_{pl}, n_{hb}	*Rheology flow behaviour index for the Power law and Herschel–Bulkley model, respectively.*	
P	*Power consumption.*	[kW]
P	*Number of parameters.*	
P/V	*Specific power consumption.*	[kW m^{-3}]
p_{init}	*Initial total pressure in the fermentor.*	[Pa]
p_{normal}	*Normal pressure. Assumed to equal 101325.*	[Pa]
p_O	*Partial pressure of oxygen in the bulk gas phase.*	[Pa]
$p_{O,init}$	*Initial partial pressure of oxygen in the bulk gas phase.*	[Pa]
p_O^{in}	*Partial pressure of oxygen in the inlet bulk gas phase.*	[Pa]
p_O^{out}	*Partial pressure of oxygen in the outlet bulk gas phase.*	[Pa]
R	*Ideal gas constant. Approximated to 8.3145.*	[J mol^{-1} K^{-1}]
T	*Temperature in Kelvin inside the fermentor.*	[K]
T_{normal}	*Normal temperature, assumed to equal 298.*	[K]
T^{in}	*Inlet temperature of air flow in Kelvin.*	[K]
T^{out}	*Inlet temperature of air flow in Kelvin.*	[K]
u_g	*Superficial gas velocity.*	[m s^{-1}]
V	*Fermentation broth volume.*	[m^3]
X	*Biomass concentration.*	[g/L]
x_O	*Fraction of oxygen in the inlet air. Assumed to equal 0.2095.*	

Greek letters

α	*The likelihood that the true population parameter lies outside the condidence interval. For 95% confidence, α = 0.05.*	
α, β	*Parameters for the k_La correlation without viscosity.*	[Pa]
$\dot{\gamma}$	*Shear rate.*	[s^{-1}]
$\dot{\gamma}_{eff}$	*Metzner and Otto's effective shear rate.*	[s^{-1}]
μ	*Dynamic viscosity.*	[Pa s]
μ_{app}	*Apparent dynamic viscosity.*	[Pa s]
μ_{online}	*Online probe dynamic viscosity.*	
$\bar{\mu}_{online}$	*Mean online probe dynamic viscosity.*	
$\bar{\mu}_{offline}$	*Mean offline apparent dynamic viscosity.*	[Pa s]
μ_{scaled}	*Normalized and scaled online probe dynamic viscosity*	[Pa s]
π	*Mathematical constant Pi*	
ρ_{broth}	*Density of the fermentation broth. Assumed to equal 1050.*	[kg m^{-3}]
σ	*Residual mean square standard deviation based on N – P degrees of freedom*	
σ_p	*The standard deviation errors on p_i, calculated as the root of the pcov diagonal.*	
τ	*Shear stress.*	[Pa]
τ_b, τ_{pl}, τ_{hb}	*Shear stresses for the Bingham, Power law and Herschel–Bulkley model, respectively.*	[Pa]
τ_{eff}	*Effective shear stress, calculated using the effective shear rate and appropriate rheology model.*	[Pa]
τ_0	*Critical shear stress.*	[Pa]
$\tau_{0,b}$, $\tau_{0,hb}$	*Critical shear stresses for the Bingham and Herschel–Bulkley model, respectively.*	[Pa]

Abbrivations

CM	*Correlation matrix for the parameter estimation*	
CDW	*Cell dry weight.*	[g L^{-1}]
CWW	*Cell wet weight.*	[g L^{-1}]
DO	*Dissolved oxygen concentration in the bulk.*	[mol m^{-3}] or [%]
DO^{*}	*Saturation concentration of dissolved oxygen corresponding to the bulk gas phase.*	[mol m^{-3}]
DO^{*}_{in}	*Saturation concentration of dissolved oxygen corresponding to the bulk gas phase in the bottom of the fermentor at the inlet.*	[mol m^{-3}]
DO^{*}_{out}	*Saturation concentration of dissolved oxygen corresponding to the bulk gas phase in the top of the fermentor at the outlet.*	[mol m^{-3}]
DO^{*}_{init}	*Initial saturation concentration of dissolved oxygen corresponding to the bulk gas phase. Used as 100% reference.*	[mol m^{-3}]
DOT	*Dissolved oxygen tension*	
pcov	*The estimated covariance matrix generated using the curve_fit method from scipy*	
OTR	*Volumetric oxygen transfer rate.*	[mol h^{-1} m^{-3}]
OUR	*Volumetric oxygen uptake rate.*	[mol h^{-1} m^{-3}]
trf	*Trust region reflective non-linear regression method.*	

Table of Contents

1 Introduction

1.1 Overview

Enzymes are nature's own catalysts, and they are at work everywhere. Novozymes A/S is the world's largest producer of industrial enzymes, and they help companies develop more sustainable and environmentally friendly solutions. The Fermentation Pilot Plant in Bagsværd is a large facility responsible for process development for different microorganism strains entering production, where the goal is to develop and optimize the operating conditions for the specific strain and production site. Novozymes wishes to utilize models online during the process development, to provide an opportunity to monitor parameters that cannot be measured online, i.e., soft sensors, and to allow for implementation of new control concepts.

This MSc thesis explores the relationship between the volumetric oxygen mass transfer coefficient, k_La, and the viscosity of a fungal fermentation broth. Previous studies have shown that k_La is dependent on the geometry of the fermentor, the agitation intensity, the airflow but also the viscosity of the broth. The initial aim of the thesis was to implement a real time k_La calculation. This was achieved successfully in MATLAB at an early stage and the scope of the thesis was therefore expanded. Some of the fermentors have viscosity probes installed which measure viscosity online. However, these online probes are both expensive and sensitive to changes in agitator intensity, and they only show a relative viscosity. Today, Novozymes are only using the online measurements as an indication for the viscosity. This thesis explores the possibilities of implementing a soft sensor instead of online probes or offline measurements. The idea for the soft sensor is to investigate if it is possible to take advantage of the correlation where k_La is dependent on viscosity and instead estimate the viscosity as a function of k_La, which could then be observed online throughout the fermentation. Different sets of fitting parameters, old and new, were compared to see which set gave the best estimation of the viscosity. It was also of interest to see what effect different feed cycle times, i.e., changes of the substrate feed pulse-pause cycle times, had on viscosity, biomass concentration and carbon effiency. Finally, the thesis explores if it is possible to estimate the biomass concentration using the rheological parameters or the estimated viscosity.

1.2 Aims

The thesis had four distinguishable aims:

1. Create a soft sensor in python for the volumetric oxygen mass transfer coefficient, k_La.
2. Estimate viscosity using empirical correlations from the literature and compare with offline measurements to evaluate if the estimation would be useful. Also explore the possibility of deriving new sets of parameters to improve the empirical correlation. If possible, create a soft sensor for the viscosity as well.
3. Get a better understanding of the effect of substrate feed pulse-pause cycle time to lay the ground for future control concepts.
4. Explore if it is possible to estimate biomass as a function of the rheology parameters or viscosity.

1.3 Outline and scope of the thesis

This thesis was organized in the following way.

Chapter 2 provides relevant background to oxygen mass transfer and viscosity so that the reader acquires sufficient knowledge. Most of the calculations are presented here as well. To keep the scope of the thesis at a manageable level, none of the statistics used were addressed in the background. Finally, the relationship between oxygen mass transfer and viscosity is established.

Chapter 3 explores all the equipment used during this project. It also describes the experimental methods as well as how the different statistical methods are done. This chapter includes relevant statistics, for example that the curve_fit module from the Scipy package is used when performing non-linear regression analysis or how to calculate a confidence band. Calculations for the remaining variables in the different correlations not explored in the background are also included in this chapter.

Chapter 4 presents, compares and discusses the offline, online and estimated data. Plots are created to visualize the results and ease comprehension. Correlation matrices, confidence intervals and confidence bands are also explored.

Chapter 5 evaluates whether the four distinguishable aims were achieved or not. It also contains the final conclusions and recommendations for future work.

2 Background

Fermentation is an industrial process which utilizes host cells to produce a product. The best known examples would be insulin production or ethanol in the production of beverages, but the possibilities are endless. Usually, the host cells are either bacteria or yeast, but fungi can also be used. Novozymes primarily produces industrial enzymes, and they are experts on fungal fermentations. There is a range of different fermentors available, but this thesis will only focus on aerated and agitated fermentors used in at Novozymes's pilot plant in Bagsværd.

2.1 General process description

There are three main modes of operation for fermentors. The simplest is batch, where all the nutrients and the carbon source are added from the beginning. The fermentation then runs until the carbon source is depleted. The upsides are its simplicity and low risk of contamination due to the closed system. The disadvantages, on the other hand, are that it is inflexible, requiring long downtimes between the fermentations due to sterilization and not allowing for control of substrate concentration etc. To tackle the inflexibility of the batch operation mode, a continuous operation mode can be used instead. In this mode substrate is added and fermentation medium withdrawn throughout the fermentation. It allows for the implementation of different control concepts, and much lower down times, yielding higher productivity. However, continuous fermentation is much more sensitive with regards to the operation settings as well as contamination (Villadsen et al., 2011).

At Novozymes, fed-batch mode operation is used, since it tackles the disadvantages of batch and continuous operation. The process starts out as a batch, but at a given time, usually at a specific pH, the feeding is initialized. The fermentor is then controlled at a specific dissolved oxygen concentration, DO, setpoint profile. In reality, it is the dissolved oxygen tension, DOT, that is measured and used as control parameter, but it is just referred to as the dissolved oxygen concentration, DO, throughout this report. The fermentation is run at overpressure and with temperatures ranging from 30 to 40 ºC, depending on the process and microorganism used. Many variables are measured continuously, including pH, temperature, pressure, feed rate, normal air flow rate, agitation speed, agitation power consumption, OUR and CER. The process data is then filtered and stored into a database continuously, from where it can be retrieved as datapoints every 15 minutes. Different offline measurements are also stored in the database, for example the carbon efficiency which is the same thing as product yield per carbon source. The carbon efficiency is an especially good measurement when comparing the yield for different scales of fermentors, however, the specific calculations are not presented in this thesis on purpose due to proprietary reasons. The database is accessible through the internal Novozymes network and it is the filtered process data that was used for all data analysis performed during this thesis project.

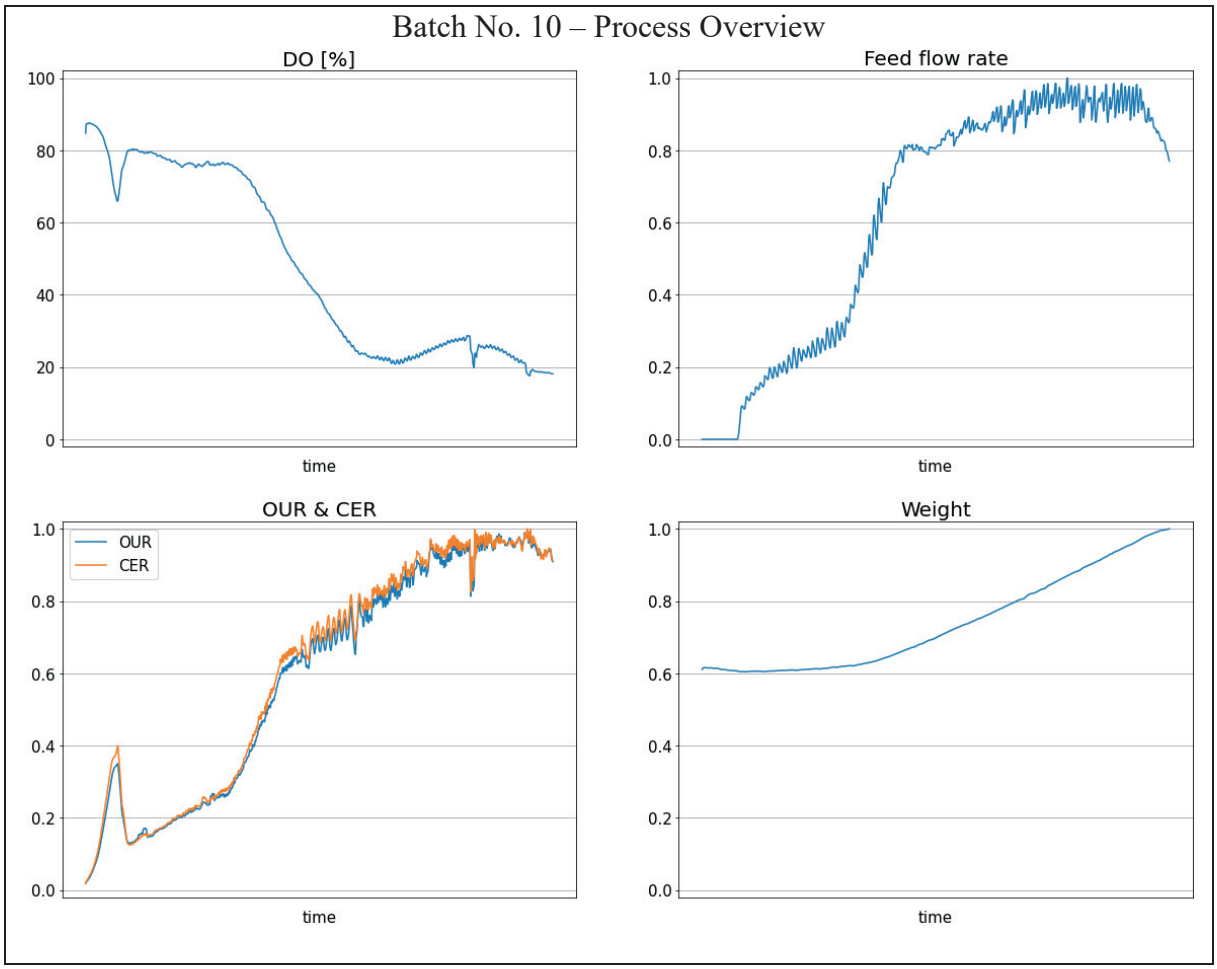

Figure 2.1. A process overview for Batch No. 10, shown as an example. The DO is controlled by a setpoint profile. OUR and CER are measured as soft sensors. The feed flow rate, OUR, CER and weight are shown as relative values on purpose due to proprietary reasons. The time units are removed on purpose as well, also due to proprietary reasons.

Another important parameter is the substrate feed pulse-pause cycle time, or cycle time for short. Instead of feeding the substrate with a constant flow, Novozymes implement a feeding regime that feeds the substrate in pulses instead, where the cycle time is the time between the feed pulses. The main advantage of feeding the fermentation in pulses is that it supposedly reduces the viscosity (Wenger et al., 2002). It is not known for certain why the viscosity is reduced, but it may be that the fungi cells get stressed leading to a limited growth of hyphal elements.

The investigations of this thesis are focused on the two filamentous fungi strains *Aspergillus niger* and *Aspergillus oryzae*. Due to growth of hyphal elements during fermentation, which tend to entangle leading to high viscosities, the oxygen mass transfer worsens. This is not ideal since the oxygen mass transfer is critical for the productivity (Albaek et al., 2011). However, even though filamentous fungi have issues with oxygen transfer, they are still suitable for industrial production due to an effective protein secretion machinery (Villadsen et al., 2011).

2.2 Mass transfer of oxygen

In an aerated fermentation, air is sparged into the fermentation broth to provide oxygen for the microorganisms. The oxygen gas is then transferred into the liquid phase where it is consumed by the live microorganisms. The transport phenomena when a compound is transported from gas to liquid can be described using two film theory, which was introduced a century ago by (Whitman, 1923). The following derivation (Eq. (2.1) to (2.7)) is based on the derivation presented by (Villadsen et al., 2011).

The molar fluxes $J_{O,G}$ and $J_{O,L}$ of oxygen through the gas and liquid layers, respectively, are calculated using the gas and liquid concentrations at the interface. Since these interface concentrations are difficult to measure, an overall flux J_O of oxygen from the gas bubble to the bulk can be used instead, as described by Eq. (2.1), where K_L is the overall mass transfer coefficient.

$$J_O = K_L(DO^* - DO) \tag{2.1}$$

DO^* is the saturation concentration corresponding to the bulk gas phase and DO is the dissolved oxygen concentration in the bulk liquid. The saturation concentration can be described by Henry's law using Eq. (2.2) , where H_O and p_O are Henry's constant and the partial pressure in the bulk gas phase for oxygen, respectively.

$$DO^* = \frac{p_O}{H_O} \tag{2.2}$$

Since H_O is temperature dependent, it can be calculated using the correlation from (Rettich et al., 2000)(Eq. (2.3)), where T is the temperature in K. The correlation is valid in temperatures ranging from 283.17 to 328.14 K and it is assumed that the solubility of oxygen in the fermentation broth resembles that of water.

$$ln\left(\frac{H_O}{p_{normal}}\right) = 14.835722 + \frac{5.837177 \times 10^3}{T} - \frac{1.085201 \times 10^6}{T^2} \tag{2.3}$$

A temperature between 30 and 40 ºC with 1 bar overpressure yields a H_O in water of approximately 850 bar L/mol. At steady state, $J_{O,G} = J_{O,L} = J_O$ which gives Eq. (2.4), where k_G and k_L are mass transfer coefficients for the molar fluxes through the gas and liquid layers, respectively. Since Henry's constant for oxygen in water is large, it can be assumed that $K_L \approx k_L$.

$$\frac{1}{K_L} = \frac{1}{H_O k_G} + \frac{1}{k_L} \tag{2.4}$$

The volumetric oxygen transfer rate, OTR, is the product of the overall flux J_O and the gas-liquid interfacial area per liquid volume, a (Eq. (2.5)).

$$OTR = J_O a \tag{2.5}$$

The interfacial area a is also difficult to calculate, which is why it is combined with k_L to form the volumetric mass transfer coefficient $k_L a$. This value is a measurement of the mass transfer of oxygen in a fermentation. Eq. (2.1) to (2.5) thus gives the commonly used expression Eq. (2.6).

$$k_{\mathrm{L}}a = \frac{\mathrm{OTR}}{(\mathrm{DO}^* - \mathrm{DO})} \tag{2.6}$$

A higher $k_{\mathrm{L}}a$ value indicates a better mass transfer and thus a higher productivity. It is therefore a common measure of the productivity of a fermentation. In large fermentors, however, the saturation concentration differs through the fermentor due to the added pressure caused by a higher liquid column. To account for this, a logarithmic mean value for the oxygen mass transfer driving force can be used instead (Eq. (2.7)).

$$k_{\mathrm{L}}a = \mathrm{OTR}\,\frac{ln\left(\frac{\mathrm{DO}^*_{\mathrm{in}} - \mathrm{DO}}{\mathrm{DO}^*_{\mathrm{out}} - \mathrm{DO}}\right)}{(\mathrm{DO}^*_{\mathrm{in}} - \mathrm{DO}) - (\mathrm{DO}^*_{\mathrm{out}} - \mathrm{DO})} \tag{2.7}$$

2.2.1 Experimental methods

There are different ways to calculate the $k_{\mathrm{L}}a$ value, and the choice is dependent on the utilities available. Today, most aerated fermentors, including the ones at Novozymes, are equipped with both probes for measuring the dissolved oxygen concentration in the broth and analysis tools that measure the composition of the exhaust gas. It is thereby possible to calculate the volumetric oxygen transfer rate, OTR as Eq. (2.8)(Villadsen et al., 2011).

$$\mathrm{OTR} = \frac{1}{V}\left(\frac{p_{\mathrm{O}}^{\mathrm{in}} F_{\mathrm{air}}^{\mathrm{in}}}{RT^{\mathrm{in}}} - \frac{p_{\mathrm{O}}^{\mathrm{out}} F_{\mathrm{air}}^{\mathrm{out}}}{RT^{\mathrm{out}}}\right) \tag{2.8}$$

R is the ideal gas constant, F_{air} is the air flow rate, T is the temperature in K and V is the broth volume. At steady state, the OTR is equal to the OUR, and $k_{\mathrm{L}}a$ can therefore be calculated in real time using Eq. (2.8) (Villadsen et al., 2011).

Other methods include the dynamic method, sulphite method and hydrogen peroxide method (Villadsen et al., 2011). However, they are not suitable when calculating $k_{\mathrm{L}}a$ as a soft sensor online since they either require that the supply oxygen is cut (the dynamic method) or the addition of chemicals (sulphite and hydrogen methods). Since Novozymes also have the appropriate equipment needed, the direct method is used throughout the thesis project.

2.2.2 Empirical correlations

A range of empirical correlations for estimating $k_{\mathrm{L}}a$ have previously been created, and most of these can be written as Eq. (2.9) (Villadsen et al., 2011). However, the viscosity can increase with more than a factor of a hundred during a fermentation, and intuitively this should also influence the oxygen transfer. Various literature data also suggests the inclusion of the viscosity in the model, as seen in Eq. (2.10)(Albaek, 2012).

$$k_{\mathrm{L}}a = C\left(\frac{P}{V}\right)^{\alpha} {u_{\mathrm{g}}}^{\beta} \tag{2.9}$$

$$k_{\mathrm{L}}a = K\left(\frac{P}{V}\right)^{a} {u_{\mathrm{g}}}^{b} \mu^{c} \tag{2.10}$$

K, a, b and *c* are parameters obtained experimentally, P/V is the specific power consumption, u_{g} is the superficial gas velocity and μ is the dynamic viscosity of the fermentation broth.

2.3 Rheology

To understand what viscosity is and how it is connected to the rheological characteristics of a fluid, some concepts need to be addressed. A fluid can be described as a series of layers that moves on top of each other. The shear rate, $\dot{\gamma}$, is the velocity gradient and it is proportional to the shear stress. The shear stress, τ, is the force per area i.e., the pressure required to maintain the momentum of the layer furthest from the wall and it is usually expressed as a function of the shear rate. The viscosity, μ, is a measure of how resistant a fluid is to deformation and is defined as the quotient of shear stress and shear rate (Eq. (2.11)) (Bird et al., 2001).

$$\mu = \frac{\tau(\dot{\gamma})}{\dot{\gamma}} \tag{2.11}$$

A fluid behaves in different ways depending on its rheological characteristics. Water, for example, is a Newtonian fluid, meaning that the shear stress is linearly proportional to the shear rate resulting in a constant viscosity. However, nature is seldom ideal, and most fluids show non-Newtonian behaviour which can be classified into three major groups.

Shear thinning fluids, also known as pseudoplastic fluids, have a decrease in viscosity with increased shear rate. Contrarily, shear thickening fluids, also known as dilatant fluids, have an increase in viscosity with increased shear rate. There are also Bingham plastic fluids which will not flow until a critical shear stress, τ_0, is exceeded. Three common examples for pseudoplastic, dilatant and Bingham plastic fluids are paint, a mixture of potato starch and water, and toothpaste, respectively. The rheological characteristics of a fluid are usually visualized in a rheogram, where the shear stress is plotted as a function of the shear rate. (Bird et al., 2001). A typical rheogram is shown in Figure 2.2.

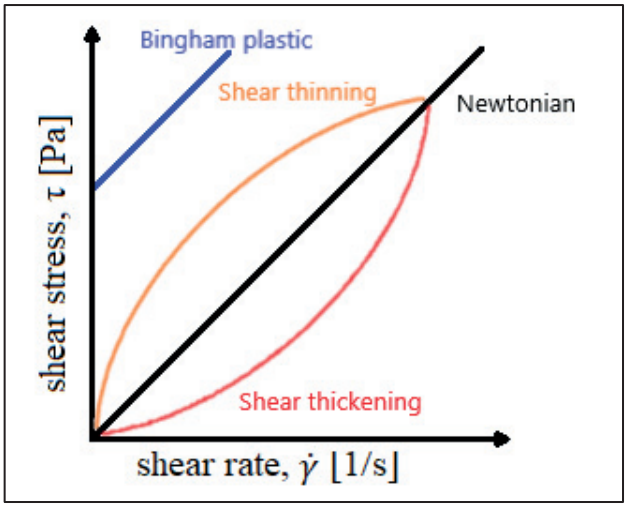

Figure 2.2. Typical rheogram for Newtonian, shear thinning, shear thickening and Bingham plastic fluids. The figure is inspired by (Bird et al., 2001).

Most non-Newtonian fluids are shear thinning, and most fungal fermentation broths are no exceptions. This does not make the estimations of the viscosity easier since the shear rate in the fermentor needs to be accounted for (Oniscu et al., 2003). Fluids are usually described flowing over a surface, and the shear rate is thereby related to the flowrate. In a fermentor, however, the shear rate is connected to the agitation intensity, but it will differ throughout the fermentor.

One way of estimating the effective shear rate, $\dot{\gamma}_{\text{eff}}$, is by using the approach of Metzner and Otto, where the effective shear rate in laminar flow is proportional to the impeller speed, N_{i}, as seen in Eq. (2.12).

$$\dot{\gamma}_{\text{eff}} = k_{\text{s}} N_{\text{i}} \tag{2.12}$$

The shear rate constant $k_{\text{s}} = 11$ is used throughout the thesis for all of the fermentors at Novozymes in accordance with (Albaek, 2012). It is not ideal to assume that the viscosity is constant throughout the fermentor, but the alternative would be computational fluid dynamics, CFD, which necessarily is not better since it is difficult, expensive and time consuming. It is important to remember what the objective of the viscosity estimation is, and since it is only vital to know the trends, the exact value is of less importance. Metzner and Otto's approach is therefore deemed sufficient.

2.3.1 Rheological measurements

Offline

The rheological characteristics of a fluid can be measured in a laboratory using a rheometer. A common method, and the one used at Novozymes, is a controlled rate rotational rheometer, but many other experimental methods are also available (Wazer et al., 1963). The shear stress is measured at varying shear rates and a fit is then created expressing the shear stress as a function of the shear rate. There are several empirical models to use when fitting the rheometer data, and the more complex they get, the more parameters are used. Simple correlations are the Bingham, Ostwald–de Waele, and Herschel–Bulkley model, and the parameters are obtained by fitting the shear stresses vs the shear rates. (Bird et al., 2001). The Bingham model can be seen in Eq. (2.13) where $\tau_{0,b}$ is the critical shear stress and k_{b} is the consistency index.

$$\tau_{\text{b}} = \tau_{0,\text{b}} + k_{\text{b}} \dot{\gamma} \tag{2.13}$$

The Ostwald–de Waele model, also known as the power law model, can be seen in Eq. (2.14), where k_{pl} is the consistency index and n_{pl} is the flow behaviour index.

$$\tau_{\text{pl}} = k_{\text{pl}} \dot{\gamma}^{n_{\text{pl}}} \tag{2.14}$$

Finally, the Herschel–Bulkley model, which is a combination of the Bingham and power law model, can be seen in Eq. (2.15). $\tau_{0,\text{hb}}$, k_{hb} and n_{hb} are the critical shear stress, consistency index and flow behaviour index, respectively.

$$\tau_{\text{hb}} = \tau_{0,\text{hb}} + k_{\text{hb}} \dot{\gamma}^{n_{\text{hb}}} \tag{2.15}$$

If a fluid has a rheological characteristic which fits the power law and/or the Herschel–Bulkley model, the flow behaviour index indicates whether the fluid is shear thinning or shear thickening. Per definition, the fluid is shear thinning if $0 < n < 1$ and shear thickening if $n > 1$.

The effective shear stress τ_{eff} is then calculated using the rheological parameters and appropriate rheological model, substituting for Metzner and Otto's effective shear rate $\dot{\gamma}_{\text{eff}}$. The apparent viscosity μ_{app} in the fermentation broth is then finally calculated using Eq. (2.11), substituting for $\dot{\gamma}_{\text{eff}}$ and τ_{eff} (Villadsen et al., 2011). As discussed earlier, the use of the effective shear rate

when calculating the effective shear rate and apparent viscosity is a rough estimate, since both the shear rate and viscosity should differ in the fermentor. The highest shear rates would be found around the impeller whereas the shear rates will be lowest further away from the impeller and close to the wall. If the broth is shear thinning, the viscosity will be lowest where the shear stress is highest and if it is shear thickening, the opposite is true.

Online

Novozymes have installed vibrating viscosity probes in some, but not all, of the fermentors. These probes measure the viscosity inside the tanks continuously. The solid probe (Figure 2.3) is submerged in the fermentation broth and set to vibrate microscopically, or resonate. By quickly turning back and forth at a specific frequency, the resonating probe dissipates energy as waves through the liquid. The dissipated energy is measured as the solid stainless-steel probe shears through the broth while resonating. The more viscous the broth is, the more energy is lost, resulting in a higher reading. However, these probes are expensive and sensitive to differences in shear rate, i.e., changes in agitation intensity. (Hydramotion, 2022).

Figure 2.3. Example of a vibrating viscosity probe installed in different tanks at Novozymes. The figure was retrieved from (Hydramotion, 2022).

2.4 Viscosity and the volumetric mass transfer coefficient

As previously mentioned, the viscosity is an important variable when estimating the k_La. At the same time, the viscosity measurement is either time consuming or expensive and not prevalent in all fermentors. It was therefore explored if it would be possible to rearrange the k_La correlation in Eq. (2.10) and instead estimate the viscosity as a function of the calculated k_La, as seen in Eq. (2.16).

$$\mu = \left(\frac{k_La}{K}\left(\frac{P}{V}\right)^{-a} {u_g}^{-b}\right)^{1/c} \qquad (2.16)$$

Two sets of the parameters *K, a, b, and c* used by (Albaek, 2012) in the k_La correlation (Eq. (2.10)) were explored, both of which estimated the k_La to an accuracy of ±30 %. These two sets were especially of interest since (Albaek, 2012) used the same fermentors as in this MSc thesis project.

3 Material and methods

3.1 Fermentors

The fermentors at Novozymes were equipped with two different types of impellers. The smaller reactors had either a Rushton turbine impeller or a Hayward Tyler B2 impeller. The larger fermentors had three Rushton turbine impellers. The same shear rate constant of k_s=11 could be used independently on the impeller, and it was therefore used for all fermentors, in accordance with (Albaek, 2012). Each fermentor was also equipped with four equally spaced baffles and a standard ring-type sparger. A schematic figure of a fermentor can be seen in Figure 3.1. An overview of some of the relevant geometries of the different fermentor scales can be seen in Table 3.1, where D_T, D_i and N_i is the tank diameter, impeller diameter and number of impellers, respectively.

Table 3.1. Overview of some relevant geometries for the two fermentors scales.

	550 L	*2500 L*
D_T	0.68 m	1.08 m
D_i	0.33 m	0.54 m
N_i	1	3

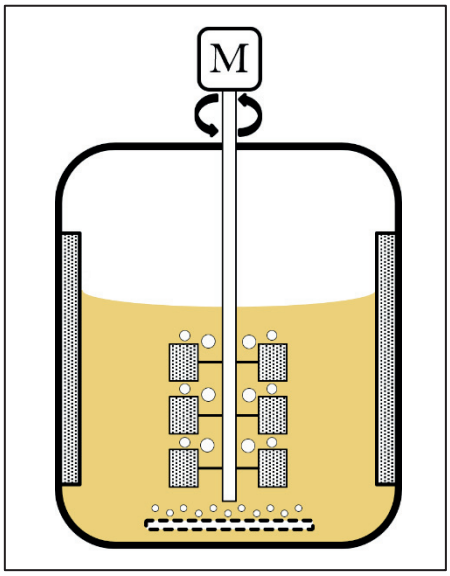

Figure 3.1. Schematic figure of a fermentor. Impellers, baffles and a sparger are visible.

3.2 Fermentation media

Generally, the fermentation media is complex waterbased solution which contains salts, a nitrogen source, and a carbon source. It is also common that the media contains amino acids, vitamins, and growth factors etc. (Villadsen et al., 2011). However, due to proprietary reasons, the fermentation media used in this thesis project cannot be expanded on. This was not a problem since the differences in fermentation media were outside of the scope of this thesis project.

3.3 Overview of batches

In total, 117 samples from 15 batches were observed. The viscosity and k_La estimations were done in a busy pilot plant, on batches which ran with different settings and purposes. This was not ideal, but the aim of the report was not to derive the best possible new sets of parameters; rather the aim was to see if the correlations worked and if they gave useful results. An overview of the batches and some of the settings used can be seen in Table 3.2.

Table 3.2. Overview of the batches observed and some relevant settings. The lowercase letters attached to the Recipe No. represents differences in run setting but with the same media recipe, whereas a different number represent a different recipe all together. The fermentation time, the specific host organism strain, the DO setpoint profile, and the recipe is not expanded on due to proprietary reasons.

Batch	Host organism	Scale	Cycle time	No. of samples	DO setpoint profile	Recipe	Additional comments
No. 1	*A. niger, strain 1*	550 L	3 min	6	No. 1	No. 1a	Online viscosity probe available.
No. 2	*A. niger, strain 1*	550 L	3 min	6	No. 1	No. 1b	Online viscosity probe available.
No. 3	*A. niger, strain 1*	550 L	3 min	6	No. 2	No. 1c	Online viscosity probe available.
No. 4	*A. niger, strain 1*	550 L	3 min	6	No. 2	No. 2	Online viscosity probe available.
No. 5	*A. oryzae, strain 1*	550 L	8 min	9	No. 3	No. 3a	
No. 6	*A. oryzae, strain 1*	550 L	8 min	9	No. 3	No. 3b	
No. 7	*A. oryzae, strain 1*	550 L	5min, 8min	9	No. 4	No. 4	
No. 8	*A. oryzae, strain 2*	2 500 L	8 min	8	No. 5	No. 5a	
No. 9	*A. oryzae, strain 2*	2 500 L	8 min	8	No. 5	No. 5b	
No. 10	*A. oryzae, strain 2*	2 500 L	8 min	3	No. 5	No. 5c	
No. 11	*A. oryzae, strain 2*	2 500 L	8 min	3	No. 5	No. 5d	
No. 12	*A. oryzae, strain 2*	550 L	8 min	11	No. 5	No. 5b	Prolonged fermentation time.
No. 13	*A. oryzae, strain 2*	550 L	8 min, 1 min	11	No. 5	No. 5b	Prolonged fermentation time.
No. 14	*A. oryzae, strain 2*	550 L	8 min, 4 min	11	No. 5	No. 5b	Prolonged fermentation time.
No. 15	*A. oryzae, strain 2*	550 L	8 min	11	No. 5	No. 5b	Prolonged fermentation time. Lower pH and T.

3.4 Offline rheology

The rheology measurements were conducted using a “vane-and-cup” geometry that Novozymes typically uses, see Figure 3.2. The rheometer used was provided by TA Instruments, model AR-G2. The cup was cylindrical with a 15 mm radius which contained approximately 30 mL of fermentation broth, just enough to cover the vane. The vane consisted of four symmetrical blades, each with the dimension H = 42 mm, W = 14 mm. The gap between the vane and the bottom of the cup was 4000 µm and it was recalibrated with each new sample. The

measurements were performed in steady state flow at a constant temperature equal to the fermentation temperature. The shear stresses were measured at varying shear rates, where each sample used 19 evenly distributed shear rates from 10 to 100 s^{-1}. The power law model, Herschel–Bulkley model and Bingham model were then compared to see which model best described the rheological characterization. The apparent viscosities μ_{app} were then calculated using the effective sheer rates $\dot{\gamma}_{eff}$ from Eq. (2.12).

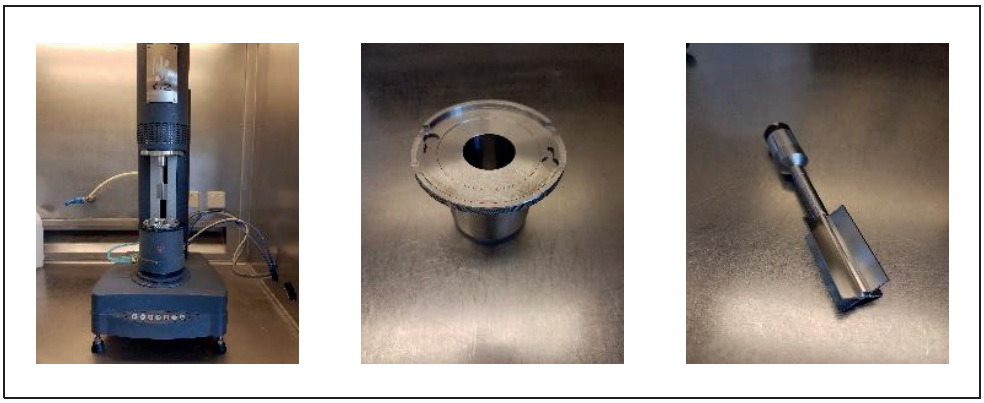

Figure 3.2. *Left:* AR-G2 rheometer from TA instruments, mounted with a vane-and-cup. *Midde:* the cup with the specified geometry. *Right:* the vane with the specified geometry.

3.4.1 Estimation of rheology parameters

The rheology data was loaded into python where the three different rheology models were compared. The curve_fit function from pythons science library, scipy, was used to fit the shear rate vs the shear stress data for each sample, using the trust region reflective, trf, method (Virtanen et al., 2020). The boundaries used when fitting the data against the rheology models can be seen in Table 3.3.

Table 3.3. Boundaries for the parameters when fitting the different rheology models.

Model	*k – lower*	*k – upper*	*n – lower*	*n – upper*	τ_0 *– lower*	τ_0 *– upper*
Bingham	1e-9	50			0.0001	100
Power law	1e-9	50	1e-9	5		
Herschel–Bulkley	1e-9	50	1e-9	5	0.0001	100

3.5 Mean filtering process data

The process data used were retrieved from Novozymes PI-database, which stores process data continuously. The datapoints used were filtered and reduced to 15 minutes intervals. Since the dissolved oxygen oscillated considerably due to the feeding regime, it had to be filtered further. By using two hours moving average it was possible to smoothen the data, as seen in the comparison of the data going into the k_La calculation (see Figure 3.3). This mean filtering was especially appropriate for the parameter estimation, but also for the viscosity estimations since the viscosity shouldn't oscillate with different dissolved oxygen concentrations. The mean filtering was therefore used on all the data from the PI-database before doing any calculations.

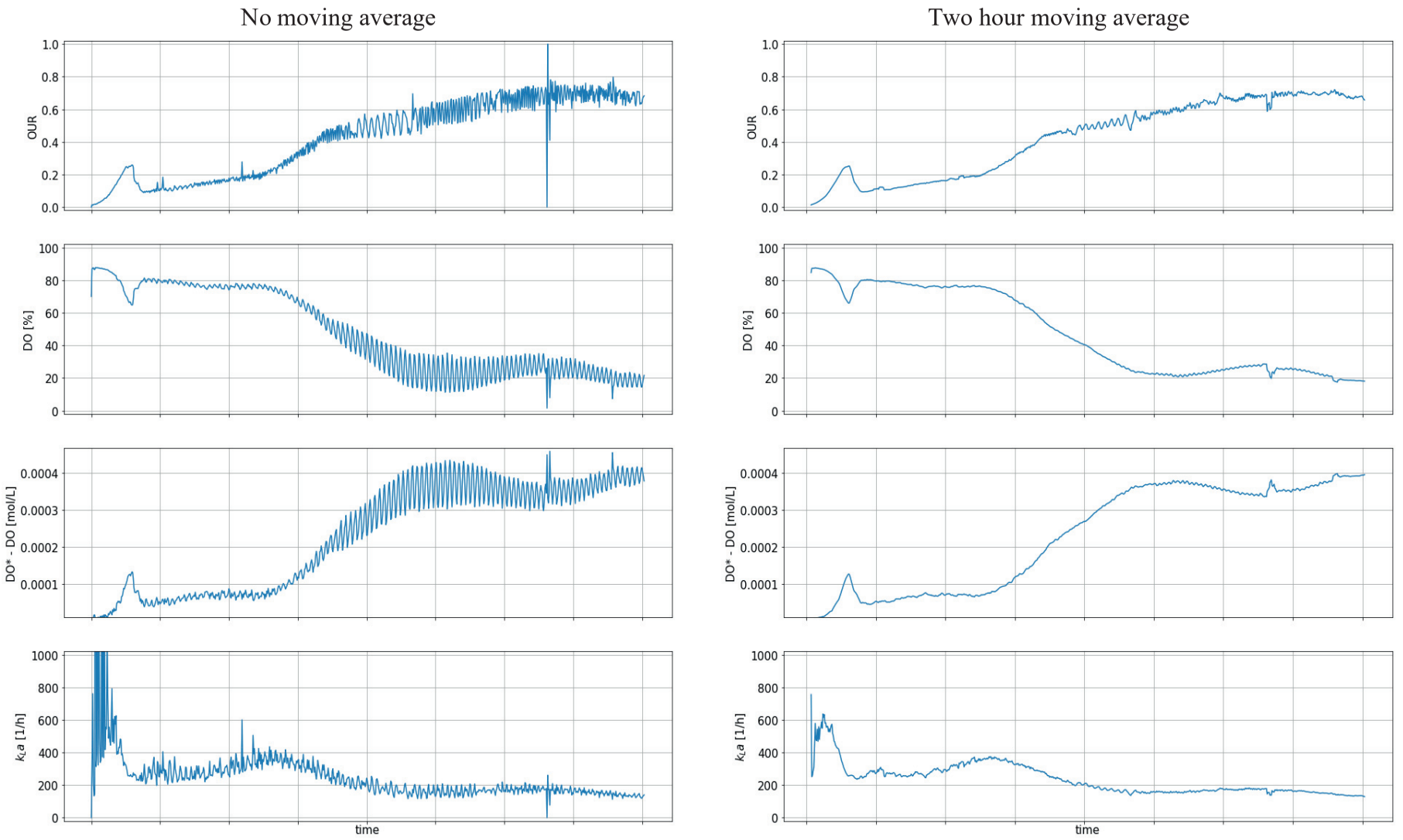

Figure 3.3. Comparison of no moving average vs 2 hours moving average for batch 10. The plots show the ingoing variables in the k_La calculation (Eq. (2.6)). Note that the OUR is shown as relative values and that the time axis labels are removed on purpose due to proprietary reasons.

3.6 Calculation of process variables

DO is retrieved in percentage and it must therefore be transformed into mol/L to use in the k_La calculations (Eq. (2.1) to (2.7)). This is done by first determining what a DO of 100% corresponds to, i.e., what the initial saturated concentration DO^*_{init} is in mol/m^3. This is possible if the initial temperature and pressure used when calibrating the DO^*_{init} (Eq. (2.2)) are known. The initial temperature is used to estimate Henry's constant H_O (Eq. (2.3)) and the initial pressure is used to calculate the initial partial pressure of oxygen (Eq. (3.1)). X_O is the fraction of oxygen in the inlet air and is assumed to equal 0.2095 since this is the fraction of oxygen in normal air. P_{init} is the total pressure in the fermentor at the time when DO^*_{init} is calibrated.

$$p_{O,init} = x_O p_{init} \tag{3.1}$$

DO in percentage is then scaled with DO^*_{init} to yield a DO concentration in mol/m^3 (Eq. (3.2)).

$$DO = \frac{DO\ [\%]}{100} DO^*_{init} \tag{3.2}$$

The broth volume, V, is calculated as the quotient of weight and broth density as Eq. (3.3) where m is the weight in kg, and ρ_{broth} is the broth density. The density is assumed to be constant at 1050 kg/m^3 in accordance with Novozymes.

$$V = \frac{m}{\rho_{broth}} \tag{3.3}$$

It is assumed that the total energy dissipated in the fermentation broth can be simplified to the measured power consumption, P, which does not take the power losses due to bearings, seal, and gearbox into account. This assumption is deemed reasonable at the scale used in this thesis, however, at larger scales it would not apply. The specific power consumption is therefore just calculated as *P/V,* with the unit kW/m^3.

The air flow rate, F, is retrieved as normal L/min, which mean that it is retrieved in L/min at normal pressure and temperature. Since the pressure and temperature are different inside the fermenter, F must be transformed into m^3/s before calculating the superficial gas velocity, u_g. This is done using Eq. (3.4), where F_{normal}, p_{normal} and T_{normal} is the normal air flow rate, pressure in Pa and temperature in K, respectively. P is the pressure in Pa in the fermentor and T is the temperature in the fermentor in K.

$$F_{air} = \frac{F_{normal}}{60} \frac{p_{normal}}{p} \frac{T}{T_{normal}} \frac{1}{1000} \tag{3.4}$$

Finally, u_g is calculated using Eq. (3.5), where D_T is the inner diameter of the fermentor.

$$u_g = F_{air} \frac{4}{\pi {D_T}^2} \tag{3.5}$$

3.7 Online viscosity

The online viscosity probe measurements were retrieved and mean filtered as the other process data. However, since the online viscosity probe data was relative and uncalibrated, it had to be transformed in some manner to compare it with the offline viscosity measurements or the viscosity correlation. The online viscosity probe measurements were therefore first normalized by

dividing with the mean and then scaled to the offline measurements, see Eq. (3.6). Since none of the online viscosity probes had been calibrated, each probe was scaled with the offline viscosity measurements for the corresponding batch.

$$\mu_{\text{scaled}} = \frac{\mu_{\text{online}}}{\bar{\mu}_{\text{online}}} \times \bar{\mu}_{\text{offline}} \tag{3.6}$$

3.8 Soft sensors

The relevant process data was loaded into Jupyther, where the calculations were conducted, and the plots created. The volumetric oxygen mass transfer coefficient, k_La, was calculated using the direct method (Eq. (2.6) and (2.8)) while assuming that OUR = OTR. The change in the saturated oxygen concentration due to the added pressure from the liquid column in the fermentor was not accounted for.

3.8.1 Estimation of new correlation parameters

After loading the relevant process parameters and offline measurements into Jupyther, the curve_fit function from scipy using the trf method was used to estimate two new sets of parameters in the k_La-viscosity correlation (Eq. (2.10) and (2.16)) (Virtanen et al., 2020). The two new parameter sets were estimated using the exact same data, only differentiated by which variable the correlation was fitted against. One set was derived by fitting relevant data against the measured k_La (Eq. (2.10)) whereas another set was derived by fitting relevant data against the offline measured μ_{app} (Eq. (2.16)). The same boundaries were used for both parameter estimations and they can be seen in Table 3.4.

Table 3.4. Boundaries for the parameter estimation when estimating k_La and apparent viscosity using the k_La-viscosity correlations (Eq. (2.10) and (2.16)).

	K	*a*	*b*	*c*
Upper	∞	∞	∞	1e–9
Lower	1e–9	1e–9	1e–9	–∞

3.8.2 Statistics

The goodness of fit was evaluated in accordance with (Bates & Watts, 1988)(Eq.(3.7)), where σ is the residual mean square standard deviation based on $N - P$ degrees of freedom. N is the number of measurements, P is the number of parameters, $M(x, p_i)$ is the model response at any variable vector x with parameters p_i, and y is the vector with experimental measurements.

$$\sigma = \sqrt{\frac{\sum(y - M(x, p_i))^2}{N - P}} \tag{3.7}$$

A 1 – α confidence band for the viscosity correlation with α = 0.05 was created in accordance with (Bates & Watts, 1988)(Eq. (3.8)), where J is the Jacobian matrix of $M(x, p_i)$ in a single point, and F is Fischer's F distribution with uncertainty $1 - \alpha$, P and $N - P$ degrees of freedom.

$$confidence\ band = M(x, p_i) \pm \sigma\sqrt{x(J^TJ)^{-1}x}\sqrt{PF(\alpha, P, N - P)} \tag{3.8}$$

The confidence band is used to describe the uncertainty of the model response and is based on the uncertainty of the parameters from the parameter estimation. It is not a predictor for the certainty of future measurements, i.e., it only describes the uncertainty for the dataset used in the estimation (Bates & Watts, 1988). However, it still gives an indication for the expected variance and could therefore be used on future estimations as well.

Correlation matrices CM, with row and column index i and j, were also created for the two new parameter sets in accordance with (Bates & Watts, 1988) and (Virtanen et al., 2020)(Eq.(3.9)). *pcov* is the estimated covariance matrix of parameters p_i generated from the curve_fit method. σ_p is the standard deviation errors on p_i, calculated as the square root of the *pcov* diagonal.

$$\mathrm{CM_{ij}} = \frac{pcov_{ij}}{\sigma_{p,i}\sigma_{p,j}} \tag{3.9}$$

The correlation matrix portrays the correlation between each possible pair of correlation coefficients for the different parameters and it is a common way to evaluate the reliability of a parameter estimation. Small correlation coefficients indicate a high reliability and vice versa. The matrix is a powerful tool which identifies patterns in a large dataset (Bates & Watts, 1988).

3.9 Biomass

The biomass measurements were conducted in a series of steps. First, empty glass tubes were weighed. Approximately 5 ml of fermentation broth was then added to the tubes, and they were weighed again. Then, the tubes were centrifuged for 20 minutes at 2800 g. After that, the supernatants were removed. The cell pellets were then washed with 5 ml of milli-Q water. A vortex mixer was used to dissolve the pellets and additionally 20 minutes of centrifugation at 2800 g followed. The supernatant was then removed, and the tubes were put in the oven at 105 ºC overnight. Finally, the tubes were weighed one last time. The biomass concentration, *X*, was then calculated using Eq. (3.10).

$$X = \frac{\mathrm{CDW}}{\mathrm{CWW}}\rho_{\mathrm{broth}} \tag{3.10}$$

CDW and CWW is the cell dry weight and cell wet weight, respectively. The density, ρ_{broth}, is assumed to have a constant value of 1050 g/L. Triplicates were used for every sample. An overview of the equipment used for the biomass measurements can be seen in Figure 3.4.

Figure 3.4. *Left:* XP204 Scale from Mettler Toledo. *Midde:* Heareus Multifuge 3SR+ centrifuge from Thermo Scientific. *Right:* Memmert oven from Buch & Holm.

4 Results and discussion

4.1 Calculations of the volumetric mass transfer coefficient

An overview of all variables going into the k_La calculation can be seen in Figure 4.1. The markers represent the process data used in the parameter estimation. It's visible that the OUR and DO profiles differ for the different batches, resulting in varying k_La values. The DO is high in the initial phase of the fermentations which gives very small driving forces DO^* – DO. These small driving forces thus results in very high k_La values in the beginning of the batches.

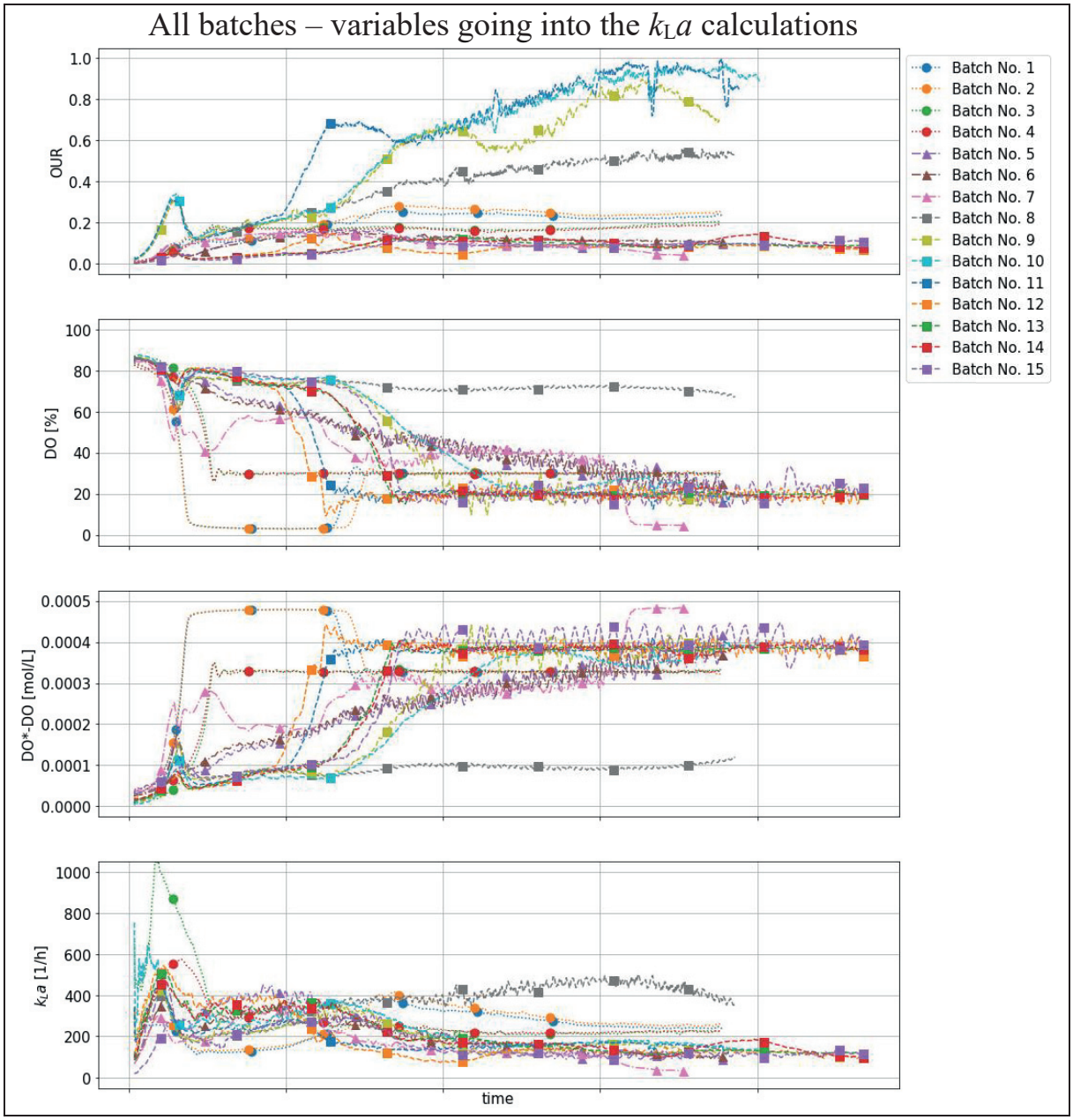

Figure 4.1. Variables from all batches going into the k_La calculation (Eq. (2.6)). The lines represent all of the process data, and the markers represent the 111 samples going into the parameter estimation. The data are differiented by color, line style and marker. Dotted lines with circle markers represent *A. niger, strain 1,* dash-dotted lines with triangle markers represent *A. oryzae, strain 1*, and dashed lines with square markers represent *A. oryzae, strain 2*. Note that the OUR is shown as relative values and that time axis labels are removed on purpose due to proprietary reasons.

Eq. (2.1) to (2.6) as well as the assumption that OUR = OTR are only valid in steady state, and since the DO varies most in the beginning, the early k_La values should be treated with suspicion. In the very beginning, the DO is also close to DO* resulting in miniscule driving forces which in turn causes really high estimates of the k_La. When the DO stabilizes, the driving forces increases resulting in lower and reasonable k_La values ranging from 50 to 500 h^{-1}. The use of the mean filter should resolve the issue caused by initial low driving forces but it has not been investigated further if the early k_La estimations really are inaccurate or not.

The use of the logmean driving force (Eq. (2.7)) instead of DO^* – DO was also evaluated, but the difference caused by the added liquid column pressure was less than 10% and thus considered negligible at the scale observed in this thesis project. It is therefore not included in the report either. However, it would be important to take the added liquid pressure into consideration when calculating the k_La using the direct method at production scale.

4.2 Rheology measurements

Initially, the R-squared value was used as a measure of how well each rheology model fitted the rheology data. A comparison of the resulting R-squared values for the different models for all samples can be seen in Figure 4.2.

A general trend was visible were the R-squared values were worst in the beginning of a batch. It should be noted that all the noisy samples with poor R-squared values were from batches that had *A. oryzae, strain No. 2* as its host organism, i.e., batch No. 8 to 15. It was observed that samples from these batches were grainier compared to the other batches, and although it was not investigated, the hypothesis was that the graininess caused the noise. However, this was not necessarily an effect of the host organism used. Instead, it could be because of a difference in the complex cultivation media, for example. The fact that the noisiness was worst in the beginning also reinforces the hypothesis that the noisiness was a result of a grainy cultivation media. If true, the grainy fraction of the broth decreased as the fermentation proceeded, resulting in R-squared values which increased with time.

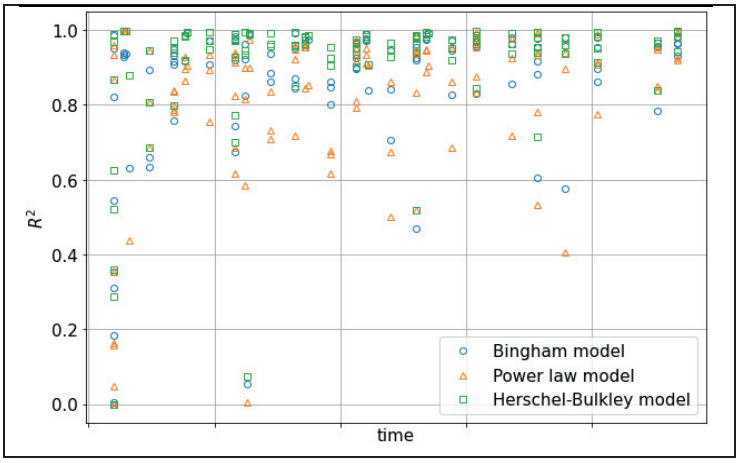

Figure 4.2. Comparison of the R-squared values for different rheology models as a function of time. Note that the time axis labels are removed on purpose due to proprietary reasons.

However, it was difficult to decide on how to interpret the R-squared results. Initially, a threshold was used and every sample which had an R-squared value of under 0.5 was excluded from

the parameter estimation. However, the R-squared value is not a good measure of goodness of fit for non-linear models, since the general mathematic framework for calculating the value only works for linear regression models. (Spiess & Neumeyer, 2010). Therefore, the idea to filter the data using the R-squared values was abandoned. Instead, each rheology sample was observed individually, specifically looking at how good the estimation of the effective shear stress, τ_{eff}, was with regards to surrounding measurements.

A comparison of the rheology models for three different samples from Batch No. 10 can be seen in Figure 4.3. As described in the Material and Methods section, the shear stress was measured at 19 evenly spaced shear rates creating a dataset of measurements for every sample. Sample 1 shows some variance between the models in the resulting effective shear stress, sample 2 shows similar results but with poor fits, and sample 3 shows very similar results and with good fits. Sample 2 was thereby deleted from the data set whereas sample 1 and 3 were kept. The same rationale was used for all samples.

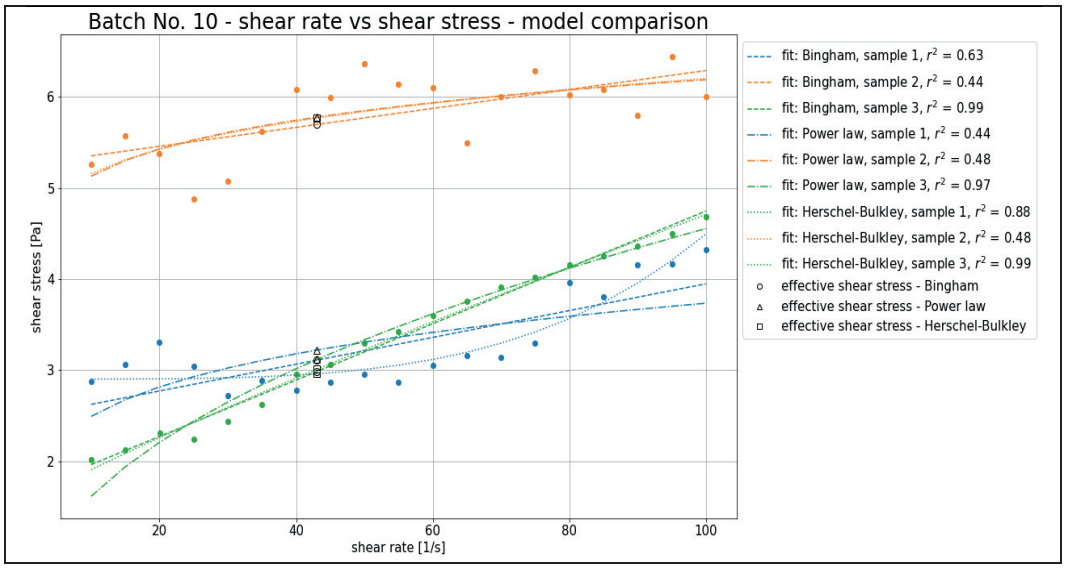

Figure 4.3. Comparison of the Bingham, Power law and Herschel–Bulkley models. The effective shear stress was calculated by inserting the effective shear rate in the appropriate model function (Eq. (2.13), (2.14) or (2.15)).

After manual inspection of all rheology measurements for all batches, it was clear that the effective shear stresses had the most satisfactory results using the Herschel–Bulkley model. This was expected since the Herschel–Bulkley model has more parameters than the other two models. Thus, the Herschel–Bulkley model was chosen as the primary rheology model in the following parameter estimation. It resulted in fewer deleted samples from the data set, which was satisfactory since it yielded a larger data set for the parameter estimation. An important note, however, was that the difference between the effective shear stresses using the different rheology models were very small for most of the samples. The difference was only significant for the samples that had noisy measurements.

In total, six samples were excluded from the data set before continuing with the parameter estimation resulting in a total of 111 samples. The excluded samples were all from an early part of batches that had *A. oryzae strain No. 2* as its host organism, which was in line with the hypothesis that the noisiness was caused by the graininess of the cultivation media. (Table 4.1).

Table 4.1. Excluded samples due to poor fitting of the rheology measurements.

	Batch No. 9	*Batch No. 10*	*Batch No. 11*	*Batch No. 13*	*Batch No. 15*
Sample No.	2	2	1, 2	4	4

A plot of the rheological parameters and viscosity vs time for all 111 samples shows an alarming result related to the flow behavior index n. As seen in Figure 4.4, approximately half of the n_{hb} values were larger than 1 indicating that the broth showed shear thickening behaviour. This is not in line with the expected shear thinning flow behaviour examined in most filamentous fermentation broths and they should therefore be treated with suspicion. At the same time, each sample was inspected manually and the model fitted the experimental data well. The shear thickening bevaiour may be caused by human error when conducting the rheology measurements, but most probably because of the graininess in many of the samples. It should be noted that these high n_{hb} values were prevalent in all batches but it was not investigated why so many of the samples showed shear thickening flow behaviour.

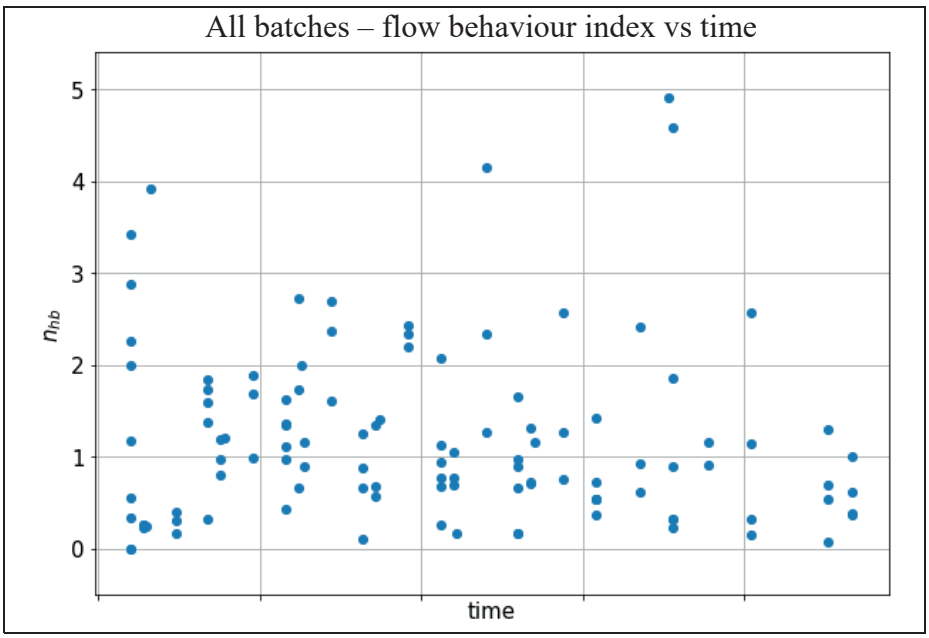

Figure 4.4. The flow behaviour index calculated using the Herschel–Bulkley model vs time for all 111 samples. Note that the time axis labels are removed on purpose due to proprietary reasons.

4.3 Parameter estimation

When evaluating the parameter estimation, it was useful to plot an overview of all the batches and how much each variable in the viscosity correlation varied, see Figure 4.5. As previously discussed, the early k_La values need to be treated with suspicion, but otherwise the k_La values seems reasonable ranging from 50 to 500 h^{-1}. The specific power consumption, P/V, also seemed reasonable with values below 20 kW/m3. The P/V starts at low values and then increases rapidly as the agitation intensities ramp up. The agitation intensities were then constant but since the fermentations ran at fed-batch mode, it was expected that the P/V would decrease as the volume increased, which is also seen in the figure. The superficial gas velocity, u_g, can be seen distributed at three different levels depending on the process settings and fermentor geometry for the different batches. These levels were approximately 0.007, 0.01 and 0.013 m/s,

respectively. Batch No. 8 to 11, for example, ran in the larger 2500 L fermentors which is part of the explanation why the u_g is larger for these batches. Finally, the offline viscosity was the variable that varied the most, with the lowest and highest values differing with a factor of approximately 100. The offline viscosities are presented in the log-scale to ease comparison between the early measurements. Except for some outliers in Batch No. 12, the apparent viscosities ranged from approximately 0.01 to 0.5 Pa s, which also seems very reasonable. The most viscous samples are more than 100 times viscous than water, but this is expected for fungal fermentation broths (Oniscu et al., 2003).

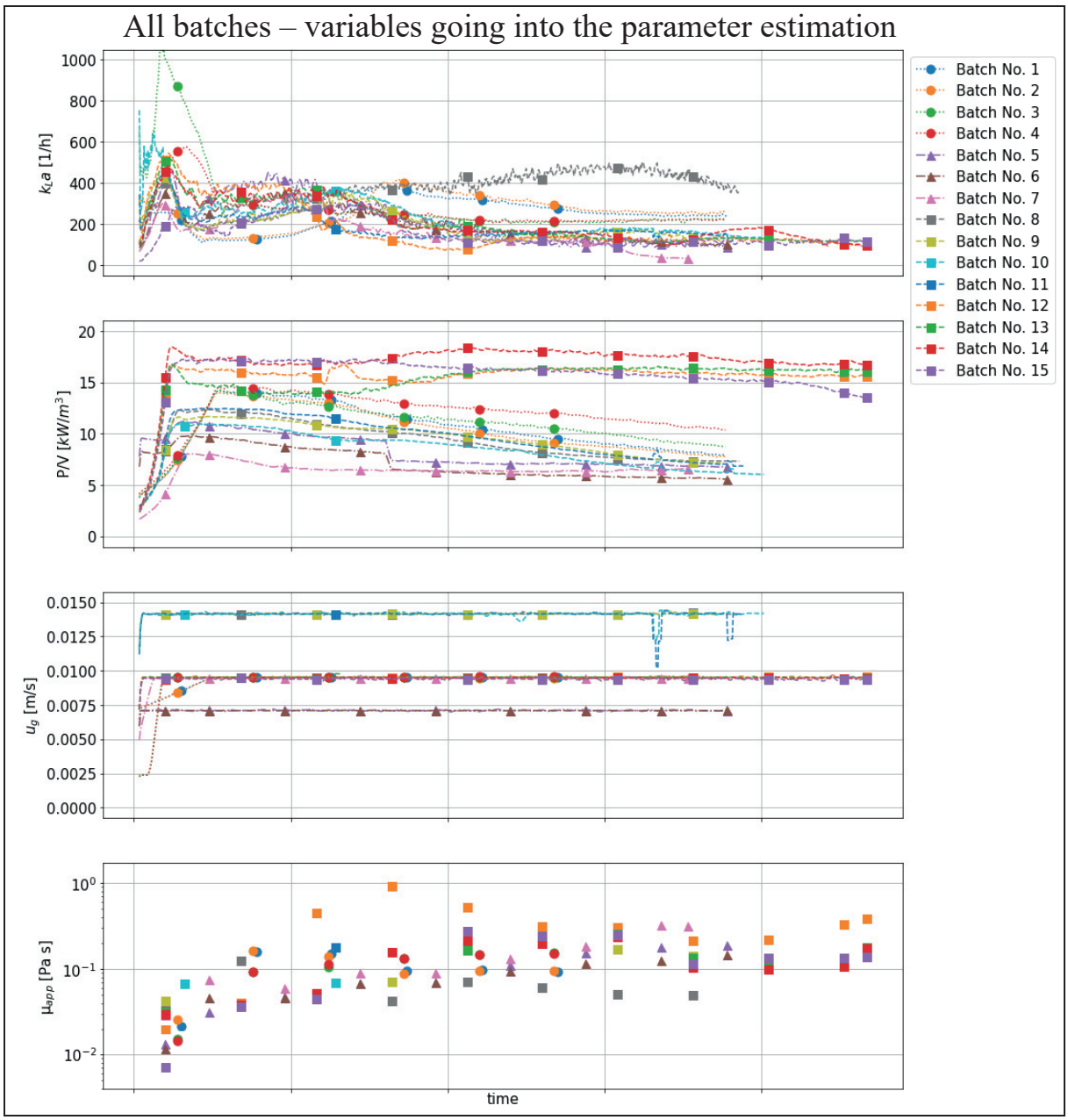

Figure 4.5. Variables from all batches going into the viscosity correlation (Eq. (2.16)). The lines represent all of the process data, and the markers represent the 111 samples going into the parameter estimation. The data are differienting by color, line style and marker. Dotted lines with circle markers represent *A. niger, strain 1,* dash-dotted lines with triangle markers represent *A. oryzae, strain 1*, and dashed lines with square markers represent *A. oryzae, strain 2*. The μ_{app}-axis have been log-transformed to ease interpretation. Note that the time axis labels are removed on purpose due to proprietary reasons.

The resulting parameter estimations can be seen in Table 4.2, where parameter set No. 1 and 2 were taken from (Albaek, 2012) and parameter set No. 3 and 4 were created using the data for the 15 batches, excluding the deleted samples in Table 4.1. The only difference between the two newly created parameter sets was how they were generated, i.e., the same data was used. Parameter set No. 3 was generated by curve-fitting of k_La measurements vs the k_La correlation (Eq.), the same approach used by (Albaek, 2012) when parameter sets No. 1 and 2 were generated. The result using set No. 3 was a model which predicted k_La more accurately than it predicted μ_{app}. Contrarily, parameter set No. 4 was generated by curve-fitting μ_{app} vs the μ_{app} correlation (Eq. (2.16)). The result using set No. 4 was a model which predicted μ_{app} more accurately than it predicted k_La. This is an astounding result, since it was expected that the generated parameter sets would be the same or at least nearly equal. The cause of this difference was not investigated, but it may be because the different models weight the variables differently thus resulting different parameter estimations. This shows that it really matters which variable the data is fitted to.

Table 4.2. Four sets of parameters for the k_La-viscosity correlations (Eq. (2.10) and (2.16)). Parameter set No. 1 and 2 are taken from previous work by (Albaek, 2012). Parameter set No. 3 and 4 were created from the dataset of 111 samples collected from the 15 batches. Set No. 3 is derived by curve-fitting measured k_La vs the k_La correlation, and parameter set No.4 is derived by curve-fitting measured μ_{app} vs the viscosity correlation.

Parameter set	*K*	*a*	*b*	*c*
set No. 1	32	0.52	0.15	–0.50
set No. 2	63	0.41	0.16	–0.39
set No. 3	1530	0.00	0.62	–0.40
set No. 4	20	0.80	0.44	–1.14

It should be noted however, that parameter a is equal to 0.00 in parameter set No. 3, which is unreasonable. As previously mentioned, parameter set No. 3 was created by fitting k_La measurements vs the k_La correlation (Eq. (2.10)). Thus, a = 0.00 might indicate that the specific power consumption, P/V, in this dataset did not vary enough to impact k_La, but as discussed in the background, previous correlations show that the k_La is in fact dependent on P/V (Albaek, 2012). It is also a result of using bounds on the parameter estimation, where there lower limit of parameter a was delibrelately set to be 1e-9, as seen in Table 3.4. This indicates that the resulting parameter estimation of a probably is a suboptima. However, if P/V had been varied deliberately, it would be expected to see an effect on k_La and thus a parameter a value above zero. Parameter set No. 3 may work when describing this dataset, but caution is advised when working with other datasets with varying P/V.

A way of comparing the goodness of fit for a non-linear model is to calculate the residual mean square standard deviation, σ (Eq. (3.7)). This was done for the four different parameter sets and their estimations of k_La and μ_{app}, as seen in Table 4.3. It was especially interesting to compare the standard deviation for parameter set No. 3 and 4, since the same data was used when the parameter sets were generated. The best k_La estimation was observed using parameter set No. 3, whereas the best μ_{app} estimation was observed using parameter set No. 4, indicated by the smallest σ in each case. A visual comparison of the four parameter sets for k_La and μ_{app} can also be seen in Figure 4.6, where A1 to A4 compares measured k_La vs estimated k_La and B1 to B4 compares measured μ_{app} vs estimated μ_{app}. The visual comparison show similar results as Table 4.3, where the best correlation for viscosity can be seen in B4.

Table 4.3. Residual mean square standard deviations for the four different parameter sets and their estimations of k_La and μ_{app}. The standard deviations were calculated using Eq. (3.7).

Parameter set	σ_{k_La}	$\sigma_{\mu_{app}}$
set No. 1	113 h^{-1}	0.20 Pa s
set No. 2	105 h^{-1}	0.67 Pa s
set No. 3	91 h^{-1}	1.27 Pa s
set No. 4	604 h^{-1}	0.09 Pa s

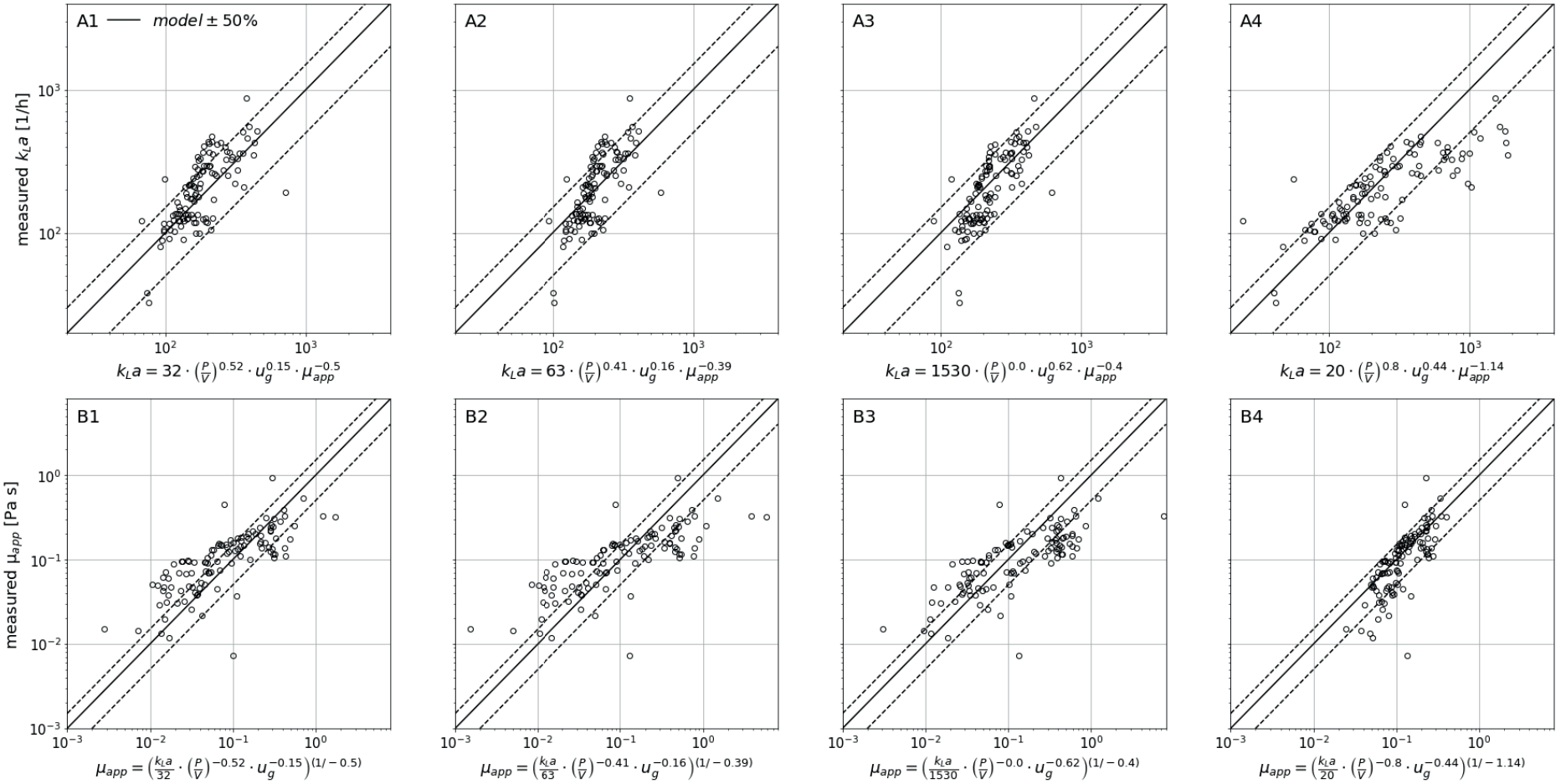

Figure 4.6. Comparison of the four different parameter sets for the correlations in Eq. (2.10) and (2.16). Error lines of ± 50% are added for reference. A1 to A4 show the measured k_La vs the estimated k_La, using parameter sets No. 1 to 4, respectively. B1 to B4 show the measured μ_{app} vs the estimated μ_{app}, also using parameter sets No. 1 to 4, respectively. Parameter set No. 1 and 2 were taken from the literature (Albaek, 2012), parameter set No. 3 was derived by curve-fitting measured k_La vs the k_La correlation (Eq. (2.10)), and parameter set No.4 was derived by curve-fitting measured μ_{app} vs the μ_{app} correlation (Eq. (2.16)). The standard deviations for the k_La estimations, i.e., for A1 to A4, were 113, 105, 91 and 604 h^{-1}, respectively. The standard deviations for the μ_{app} estimations, i.e., for B1 to B4 were 0.20, 0.67, 1.27 and 0.09 Pa s, respectively.

4.3.1 Statistics

Different things were considered when evaluating the results of the parameter estimation. It was of interest so see how robust the parameter estimation was, and one way to test this was to change the initial guesses for the parameters going into the curve_fit function. At first, no guess was specified resulting in the estimated parameter sets No. 3 and 4. Then, the parameters from the previous work by (Albaek, 2012) was used as initial guesses; i.e., parameter set No. 1 and 2. Both of them also resulted in the same estimated parameter sets No. 3 and 4. Finally, parameter set No. 3 was tried as the initial guess for No. 4 and vice versa. The parameter estimations remained the same indicating that the curve fit converged satisfactorily. To conclude, all sets of different initial parameters resulted in the same parameter estimations No. 3 and 4. The choice of initial guesses was thus deemed irrelevant. However, as previously mentioned, it should be noted that parameter *a* in set No. 3 is estimated to equal 0.00 which is on the boundary and this is therefor a suboptima. It was not explored whether a different non-linear regression model would yield different results. It should also be noted that no cross-validations were done for the parameter estimations. This would have been beneficial to further evaluate the robustness of the models.

The correlation matrices for model No.3 and 4 can be seen in Table 4.4. The parameters were surprisingly uncorrelated which indicates that the models are quite reliable. This is also in line with the result regarding the initial guesses, i.e., the model converges, and a true minimum is found. This is an improvement compared to the correlation matrix presented by (Albaek, 2012), where the parameters showed signs of large correlation. There was one exception, however, and that was the correlation coefficient for pair *K*-*b*, which had a value of over 0.9 for both parameter sets indicating very high correlation. This may be caused by the relatively low variance in the superficial gas velosity, u_g.

Table 4.4. Correlation matrices for parameter set No. 3 and 4, calculated using Eq. (3.9). Each value in the matrix ranges from –1 to 1, where 1 represents perfect correlation, 0 represent no correlation at all and –1 represents perfect inverse correlation

set No. 3	*K*	*a*	*b*	*c*
K	1	–0.30	0.92	0.03
a	–0.30	1	0.07	–0.17
b	0.92	0.07	1	–0.18
c	0.03	–0.17	–0.18	1

set No. 4	*K*	*a*	*b*	*c*
K	1	–0.27	0.97	0.21
a	–0.27	1	–0.06	–0.17
b	0.97	–0.06	1	0.08
c	0.21	–0.17	0.08	1

To conclude, the dataset generated during this theses was not ideal for a parameter estimation. However, it really matters if the parameters are estimated by curve-fitting k_La vs the k_La correlation, or by curve-fitting μ_{app} vs the μ_{app} correlation. The recommendation to Novozymes is therefore to look at the dataset used by (Albaek, 2012) and estimate new parameters by curve-fitting μ_{app} vs the μ_{app} correlation.

4.4 Viscosity estimation

The apparent viscosities were estimated using the viscosity correlation in Eq. (2.16). The four different parameter sets presented in Table 4.2 were used when creating the plots. Good and bad estimates, respectively can be seen in Figure 4.7.

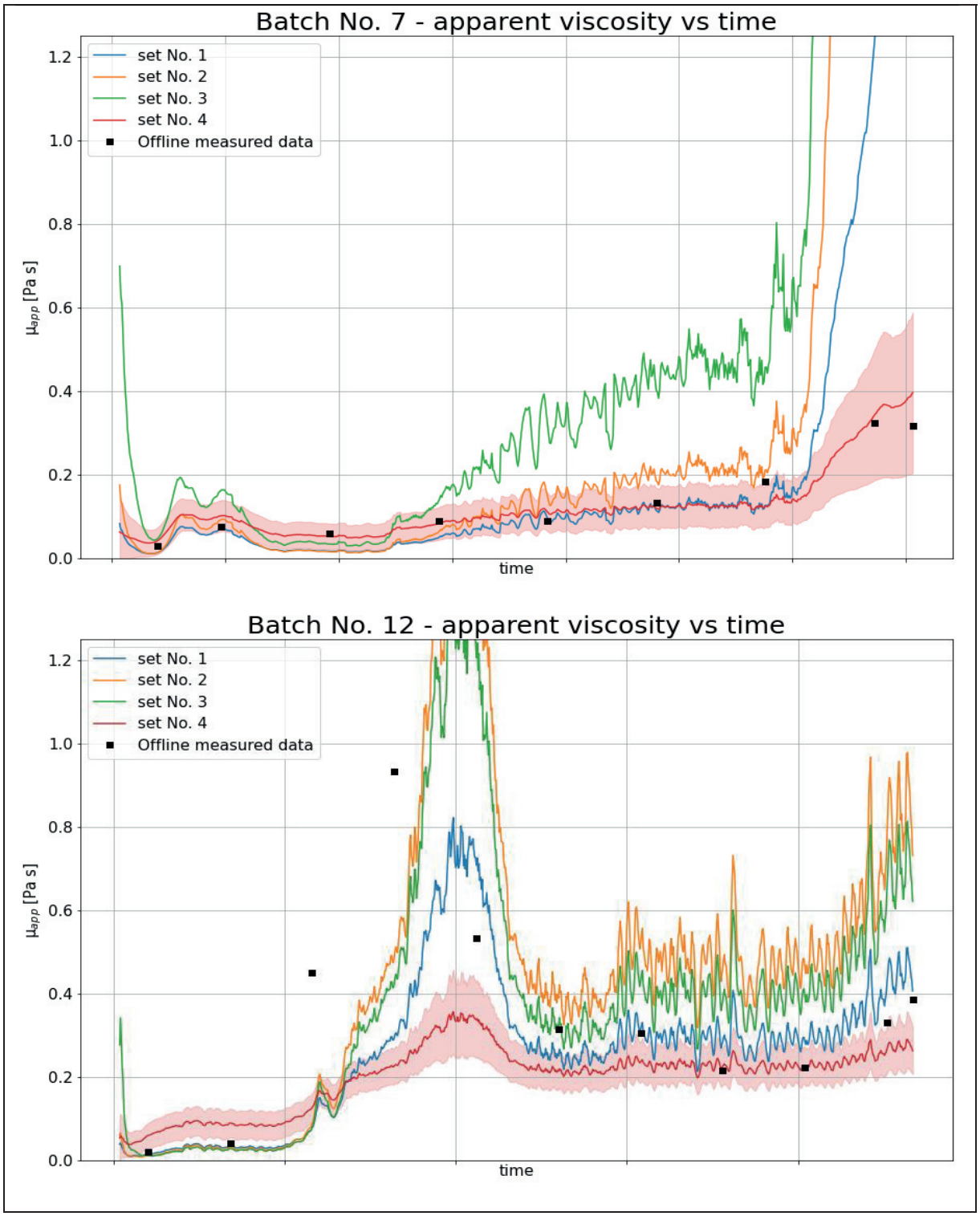

Figure 4.7. Plots of viscosity estimations using the different sets of parameters. The upper and bottom figures are examples of relatively good and bad fits to the experimental data, respectively. A 95% confidence band for a non-linear model was also created for set No. 4 using Eq. (3.8) in accordance with (Bates & Watts, 1988). Note that the time axis labels are removed on purpose due to proprietary reasons.

The newly estimated parameter set No. 4 was deemed the best one and it is probably because this parameter set was created using the same data that it estimated. Parameter set No. 1 also showed quite promising results. However, when comparing the data set generated during this thesis with the data set generated by (Albaek, 2012), the latter is more diverse, with greater variance in power input and superficial gas velocity for example. However, parameter set No. 1 and 2 were created to fit k_La and as seen when comparing parameter set No. 3 and 4, this does not achieve the best possible parameter set for estimating viscosity. A final recommendation would therefore be to look back at the dataset generated by (Albaek, 2012) and see if it is possible to create a new parameter set by fitting the data to the apparent viscosity instead. If Novozymes wishes to implement a soft sensor for the viscosity before creating a new parameter set, parameter set No. 4 should be used, since this is the only parameter set created to fit the apparent viscosity and it shows promising results. It also has a confidence band which could be used to highlight the uncertainty of the estimation.

It was also of interest to see how the estimated viscosity compared with the online viscosity. But first, a comparison of the online probe viscosities and offline measured viscosities was done, as seen in Figure 4.8. Only Batch No. 1 to 4 had viscosity probes installed in the fermentors, so the evaluation of the viscosity probe data could therefore only be done on these four batches. Batch No. 1 and 2 ran on almost the same settings, and the same was true for Batch No. 3 and 4. It was also visible that these two groups of batches had almost identical offline viscosity results. When comparing the online viscosities, Batch No. 3 and 4 were very similar whereas Batch No. 1 and 2 almost differed by a factor of two, even though capturing the same trends. Batch No. 1 also had some visible spikes, but they were not a result of spikes in the impeller intensity or any other measured process parameter. However, the spikes were not further explored and therefore just considered as noise in the data. It is important to note that the online probe data only gave a relative viscosity, and that the probes were not calibrated either. Therefore, the probes did not necessarily give the same results even though they were identical.

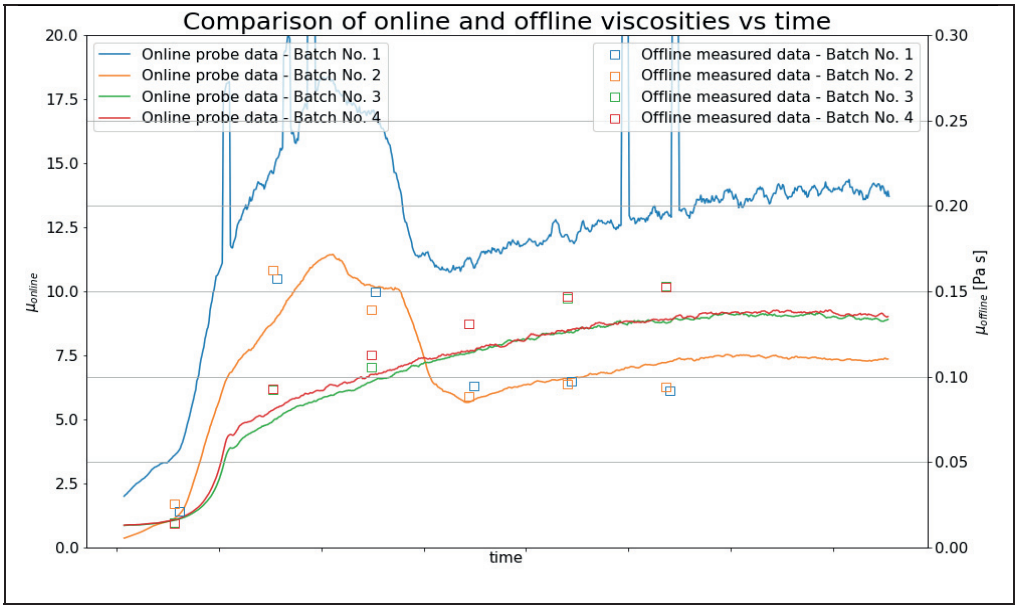

Figure 4.8. Comparison of online and offline viscosities vs time. Only batch No. 1 to 4 had online viscosity probes installed in the fermentors. Note that the time axis labels are removed on purpose due to proprietary reasons.

It was clearly visible that the viscosity probes did not give reliable data to use for implementing different control concepts. However, it was possible to compare the online viscosities with the estimated viscosities after scaling each batch's online viscosity probe data to the experimental apparent viscosities (Eq. (3.6)). As seen in Table 4.5, the viscosity probe measurements for Batches No. 2 to 4 where much more similar in scale compared to the probe in Batch No. 1.

Table 4.5. Viscosity probe scale factor. Calculated as the mean online viscosity divided by the mean offline viscosity, as in Eq. (3.6).

	Batch No. 1	*Batch No. 2*	*Batch No. 3*	*Batch No. 4*
Scale factor	125:1	69:1	64:1	65:1

As seen in Figure 4.9, the scaled probe data follows the same trend as the offline measured data. However, it has not been investigated if the scale factors are altered with different agitation intensities, different host organisms or any other difference in process settings, but the scale factors cannot be assumed to be constant either. To conclude, the online viscosity probes are worthless if they are to be used to implement any control concepts, since they would need to be scaled to offline apparent viscosity data first. The disadvantages with price and sensitivity are also still prevalent whereas the estimated soft sensor viscosity shows promising results, is easily implemented and free of charge. The recommendation to Novozymes is therefore to continue investigating the viscosity estimation and not invest in more online viscosity probes.

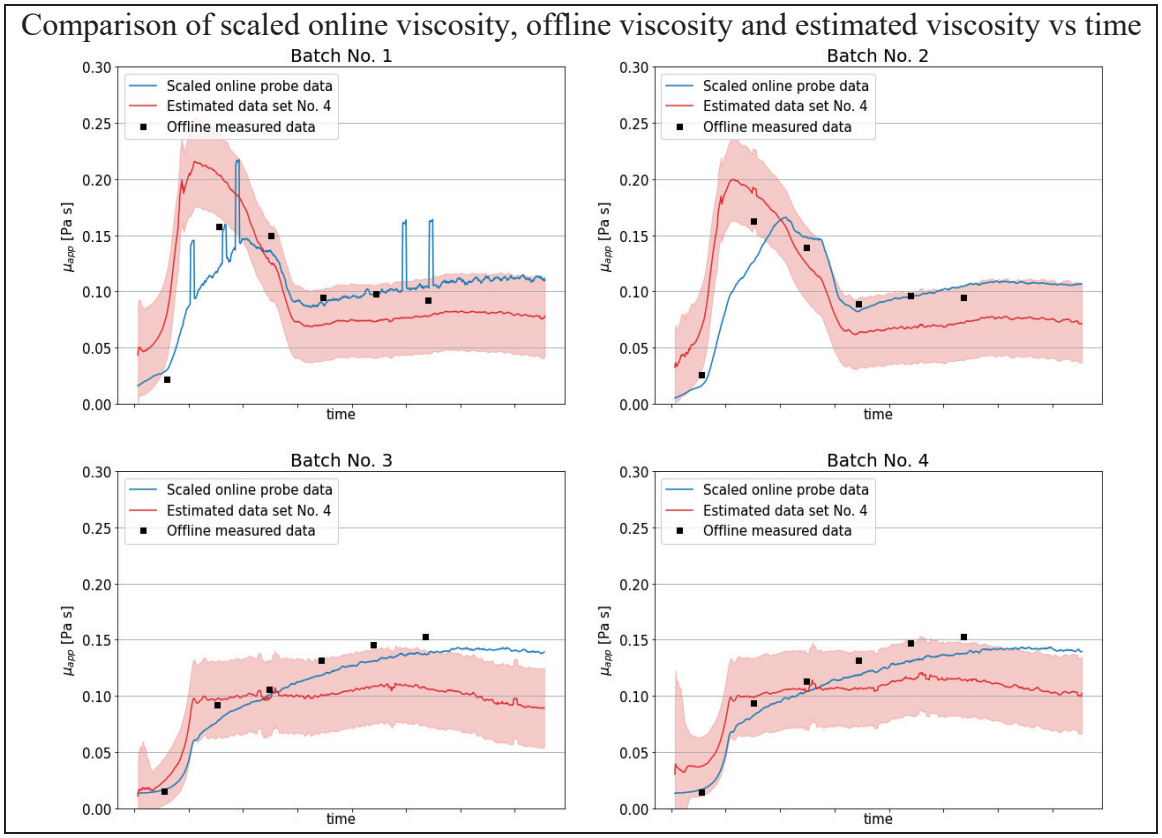

Figure 4.9. Scaled online probe data vs estimated viscosity using parameter set No. 4. with a 95% confidence band. Offline viscosity data was also included. Note that the time axis labels are removed on purpose due to proprietary reasons.

4.5 Cycle time influence on carbon efficiency, biomass and viscosity

Three batches were observed. Batch No. 9 was run at a larger scale, but aside from that, the settings for Batch No. 9, 13 and 14 were almost identical. The only difference between the batches was the change of cycle time; at a specific point in time, the cycle times for Batch No.13 and 14 were changed from 8 minutes to 1 and 4 minutes, respectively. The effect that the cycle time change had on carbon efficiency, the biomass, and the viscosity is seen in Figure 4.10.

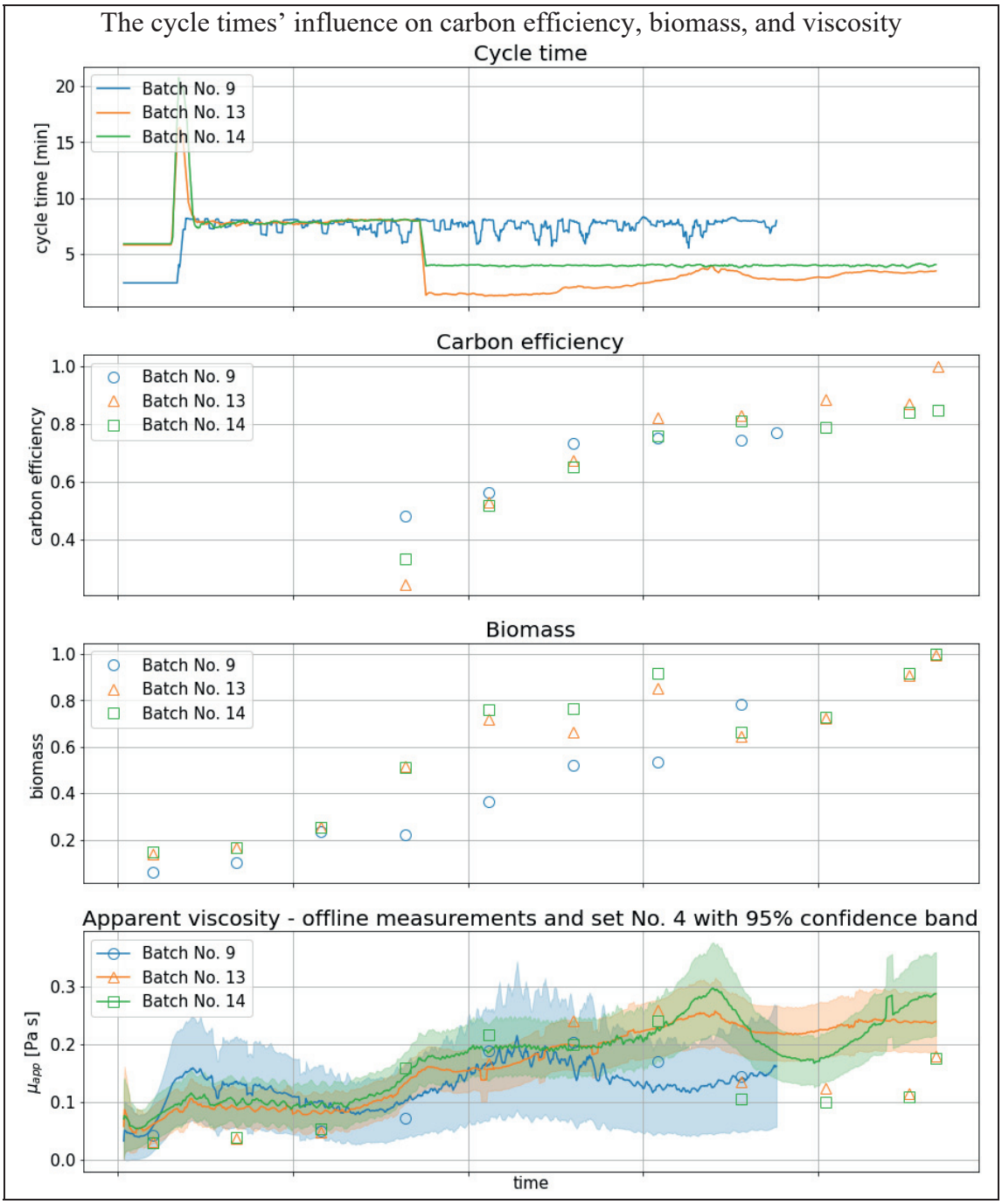

Figure 4.10. Influence of the change in cycle time on carbon efficiency, biomass concentration, and estimated as well as offline apparent viscosity. The apparent viscosity was estimated using parameter set No. 4 and Eq. (2.16). The carbon efficiency and biomass are only shown as relative values due to proprietary reasons. Note that the time axis labels are removed on purpose, also due to proprietary reasons.

The result that the change in cycle time has on the viscosity estimation is difficult to comprehend by just looking at plots in Figure 4.10. When looking at the specific moment for the change in cycle time, it might indicate that the slope of the viscosity flattens out a bit for Batch No.13 and 14 because of a shorter cycle time, but it is uncertain if this change in the slope was caused by the change cycle time or if it's just by coincidence. It was also difficult to draw relevant conclusions when looking at the cycle times' effect on biomass concentration. Batch No. 13 and 14 showed almost identical results whereas Batch No. 9 had lower concentrations almost throughout. Thus, the differences in biomass concentrations were not necessarily because of the difference in cycle times. Sample No. 4 for each batch, for example, are generated before the change in cycle times but still show a huge difference between Batch No. 9 and Batch No. 13 and 14.

Finally, when looking at the carbon efficiency, the effect of changing cycle times was not obvious either. The final sample might have shown some difference between Batch No. 13 and 14, where a shorter cycle time resulted in a higher carbon efficiency, but the general effect is not obvious. To conclude, a more thorough investigation with a larger dataset is needed to draw relevant conclusions on the effect cycle time has on viscosity, biomass and carbon efficiency.

4.6 Biomass results and correlation with rheology parameters.

4.6.1 Biomass measurements

A summary of the biomass measurements can be seen in Figure 4.11, and the biomass measurements are reasonable. The biomass increased with time, which was expected. The standard deviations for the replicates were small, resulting in neat confidence intervals.

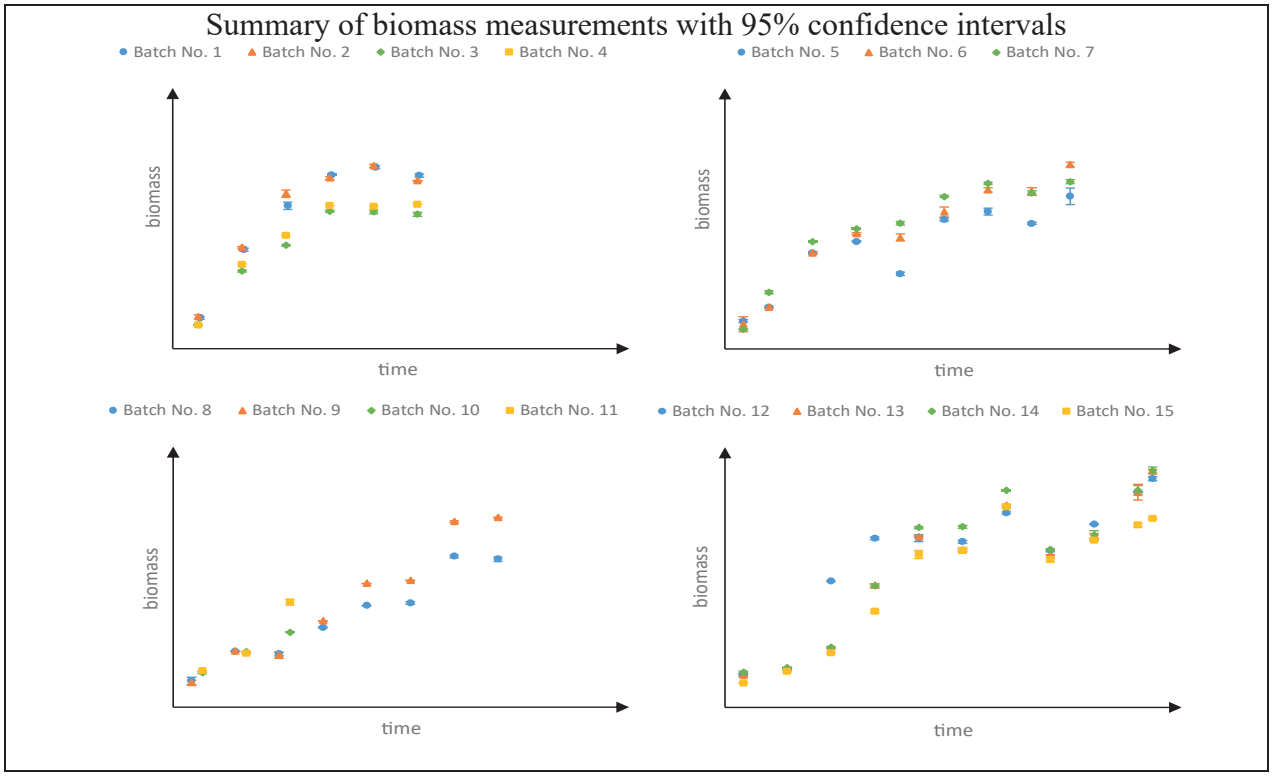

Figure 4.11. Summary of all the biomass measurements with 95% confidence intervals. Note that both the time and biomass axis labels are removed on purpose due to proprietary reasons.

As seen in Figure 4.7, the viscosity increases dramatically for Batch No. 12 and this dramatic change can also be seen in the biomass measurements in Figure 4.11. It is only possible to speculate why this sample is an outlier, but it is not caused by a difference in the media recipe. The difference is probably caused by an earlier initiation of the substrate feed. A visual timeline of the biomass experimental method can be seen in Figure 4.12, where some samples are shown before centrifugation, after washing and centrifugation, and after drying.

Figure 4.12. *Left:* Broth before centrifugation. *Middle:* Wet cells after washing and centrifugation. *Right:* Dry cells after drying in the oven.

4.6.2 Biomass and rheology

The left plot in Figure 4.13 was created to explore whether it was possible to model the rheology parameters as a function of the biomass concentration and it shows the different rheology parameters for the Herschel–Bulkley model for all the 111 samples. It was also observed how the rheology parameters changed with time, as seen the right plot (Figure 4.13). Since the biomass concentration increases with time, as seen in Figure 4.11, the similarities between the plots in were expected. Similar plots were also created for the Bingham and Power law models, however, these were not included in the report. It should be noted that the y-axis for k_{hb}, $\tau_{0,hb}$ and μ_{app} have been displayed in the logarithmic scale to get a better visualization.

There is some indication that the flow behaviour index n_{hb} decreases with the biomass, but the variance is still too large to get any meaningful correlations. The variance for the consistency index k_{hb}, and the critical shear stress $\tau_{0,hb}$ is even larger with values differing by factors of 10. $\tau_{0,hb}$ also shows some outliers where the values lay on the specified boundary as seen in Table *3.3*. The biomass vs viscosity seemed to have the best correlation; however, it was still not enough to yield any meaningful correlations for estimating biomass as a function of the apparent viscosity. It was also explored whether it was possible to observe a correlation between the rheology parameters and the biomass, differentiating for the three host organisms used. However, no obvious correlation could be seen for any of the host organisms either. Rheology vs biomass plots for the different strains and host organisms used in the thesis can be seen in the appendix (Figure 7.1, Figure 7.2, and Figure 7.3).

In conclusion, it was not possible to derive any meaningful correlations from the dataset analyzed in this thesis. It might be possible for other host organisms or if the dataset was analyzed more thoroughly, but the idea to have a model which predicts the biomass concentration from the viscosity regardless of the host organism used seems far-fetched.

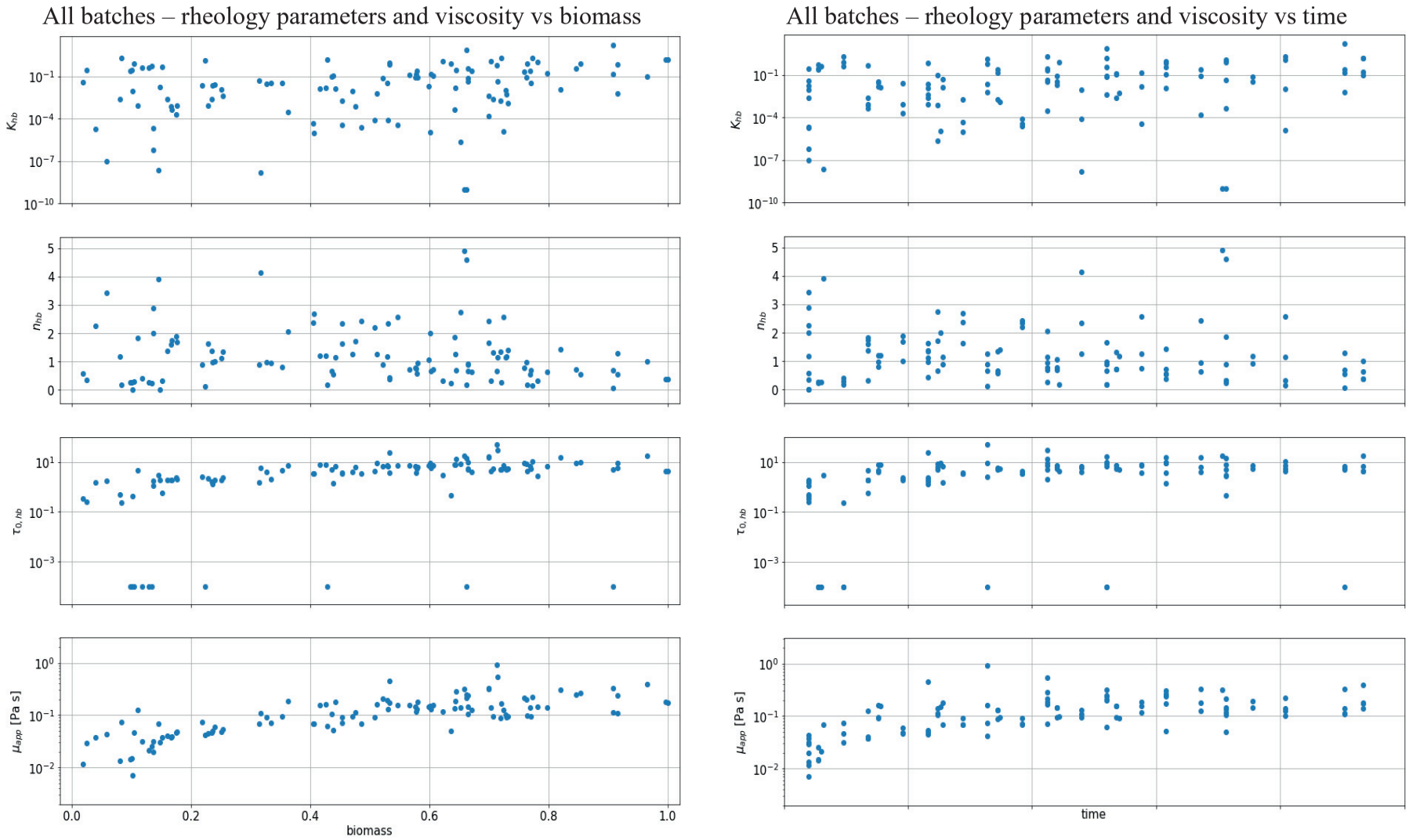

Figure 4.13. *Left:* Rheology parameters using the Herschel–Bulkley model vs biomass as well as viscosity vs biomass. *Right:* Rheology parameters using the Herschel–Bulkley model vs time as well as viscosity vs time. All 111 samples are used in both plots. The y-axes for k_{hb}, $\tau_{0,hb}$ and μ_{app} were log-transformed in both plots for better visualization. Note that the biomass measurements only show relative values and that the time axis labels are removed on purpose due to proprietary reasons.

5 Overall conclusion and suggestion for future work

Overall, this thesis project has been a great achievement. Novozymes can now continuously monitor the k_La as a soft sensor online at their pilot production site in Bagsværd, Denmark. The process data should however first be filtered to reduce oscillation and get smoother plots. It was also possible to derive a useful viscosity estimation using the online calculated k_La and parameter set No. 4, with a standard deviation of 0.09 Pa s, thereby creating yet another soft senor for Novozymes to implement. It was concluded that it matters if the parameters are estimated by curve-fitting k_La vs the k_La correlation, or by curve-fitting μ_{app} vs the μ_{app} correlation. It was also concluded that the dataset generated in this thesis was not suitable for a parameter estimation due to low variance in the variables. The parameter set used in the viscosity estimation could thus be improved if a dataset with a larger variance in P/V, u_g and k_La was used. Instead of generating a new dataset, Novozymes is recommended to look back at the work previously done by (Albaek, 2012) and use his dataset to do a new parameter estimation by curve-fitting μ_{app} vs the μ_{app} correlation.

Nevertheless, by assuming that the offline measured value of the apparent viscosity is true, the resulting viscosity estimation still estimates the apparent viscosity better than the current online viscosity probes, since they only give a relative viscosity. Novozymes is therefore recommended to continue working on perfecting the viscosity estimation and not invest in new online viscosity probes.

The influence that the substrate feed pulse-pause cycle time had on viscosity, biomass and carbon efficiency was not completely clear. Further investigation is therefore still needed before implementing new control concepts using the cycle time and estimated viscosity.

The final objective to explore whether it was possible to correlate the biomass with rheology parameters was achieved. However, the observation did not yield promising results since no clear correlation could be observed. It is still an interesting idea and further investigations could result in relevant insight in how the rheology of a specific host organism and process is correlated to the biomass. However, it seems unlikely that a general correlation could be found for any host organism and process.

6 References

Albaek, M. O. (2012). *Evaluation of the efficiency of alternative enzyme production technologies.* DTU Chemical Engineering].

Albaek, M. O., Gernaey, K. V., Hansen, M. S., & Stocks, S. M. (2011). Modeling enzyme production with Aspergillus oryzae in pilot scale vessels with different agitation, aeration, and agitator types. *Biotechnol Bioeng*, *108*(8), 1828-1840. https://doi.org/10.1002/bit.23121

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. Wiley.

Bird, R. B., Stewart, W. E., & Lightfoot, E. N. (2001). *Transport Phenomena* (2nd ed.). John Wily & Sons Inc.

Hydramotion. (2022). *Why in-line measurement?* Retrieved 2022-06-01 from https://hydramotion.com/en/technical

Oniscu, C., Galaction, A.-I., & Cascaval, D. (2003). Rheology of fermentation broths. 1: Rheological behaviours and influencing factors. *Revue Roumaine de Chimie*, *48*(2), 91-110.

Rettich, T. R., Battino, R., & Emmerich, W. (2000). Solubility of gases in liquids. 22. High-precision determination of Henry's law consants of oxygen in liquid water from T = 274 K to T = 328 K. *The Journal of Chemical Thermodynamics*, *32*, 1145-1156. https://doi.org/10.1006/jcht.1999.0581

Spiess, A. N., & Neumeyer, N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol*, *10*, 6. https://doi.org/10.1186/1471-2210-10-6

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction Engineering Principles* (3rd ed.). https://doi.org/10.1007/978-1-4419-9688-6

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. v. d., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . Contributors, S. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261-272. https://doi.org/10.1038/s41592-019-0686-2

Wazer, J. R. V., Lyons, J. W., Kim, K. Y., & Colwell, R. E. (1963). *Viscosiy and Flow Measurement. A laboratory handbook of rheology*. Interscience Publishers.

Wenger, K. S., Caicedo, M. A., Stock, S. M., Bhargava, S., & Marten, M. R. (2002). *Fermentation with Cyclic Pulse-Pause Feeding* (US Patent No. 2005/0064536). Novozymes. https://www.freepatentsonline.com/y2005/0064536.html

Whitman, W. G. (1923). A preliminary experimental confirmation of the two-film theory of gas absorption. *Chemical and Metallurgical Engineering*, *29*(4), 146–148.

7 Appendix

7.1 Additional figures

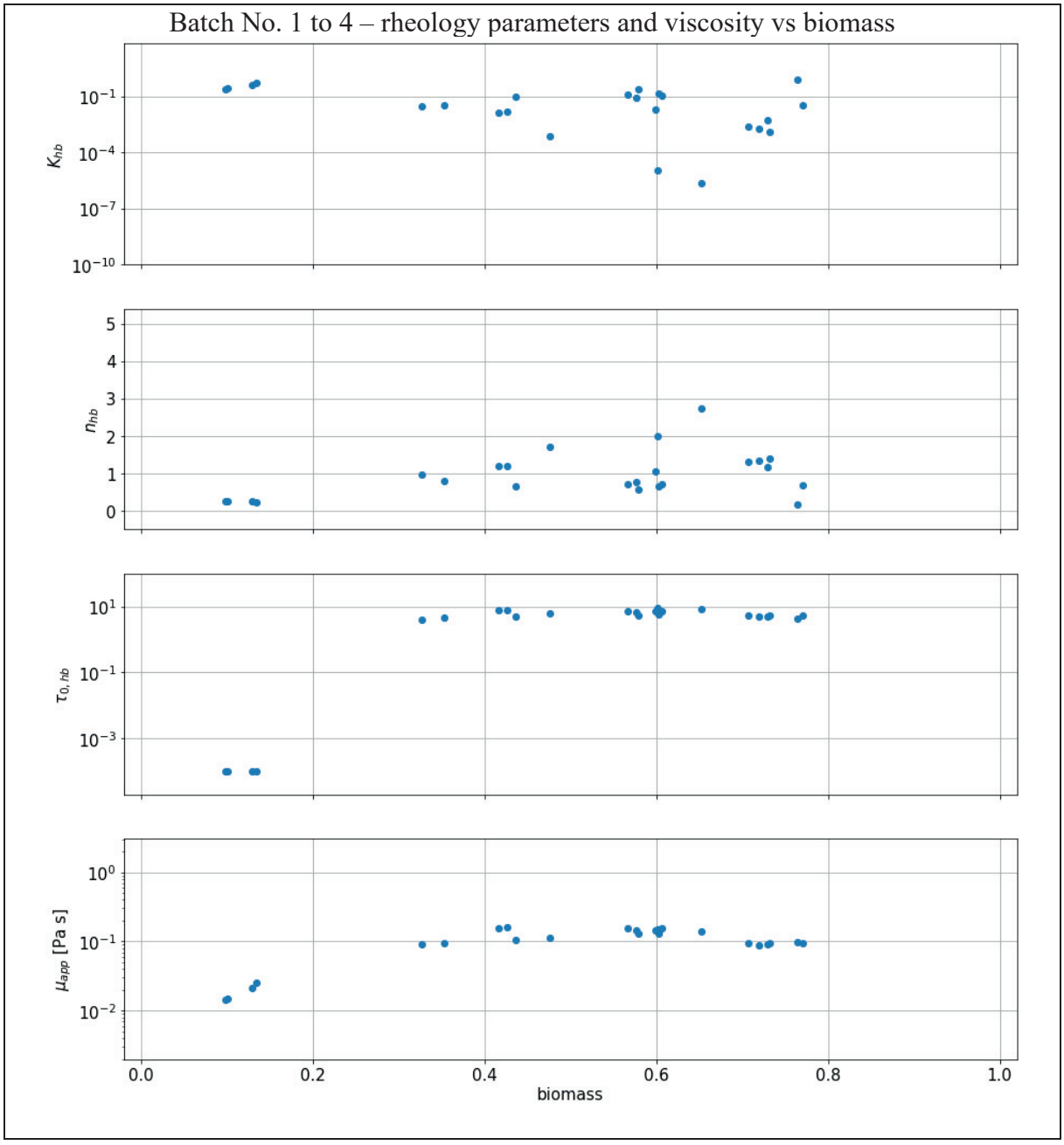

Figure 7.1. Rheology parameters using the Herschel–Bulkley model vs biomass as well as viscosity vs biomass for the samples from Batch No. 1-4. The host organism was *A. niger, strain 1*. The y-axes for k_{hb}, $\tau_{0,hb}$ and μ_{app} were log-transformed for better visualization. Note that the biomass measurements only show relative values on purpose due to proprietary reasons.

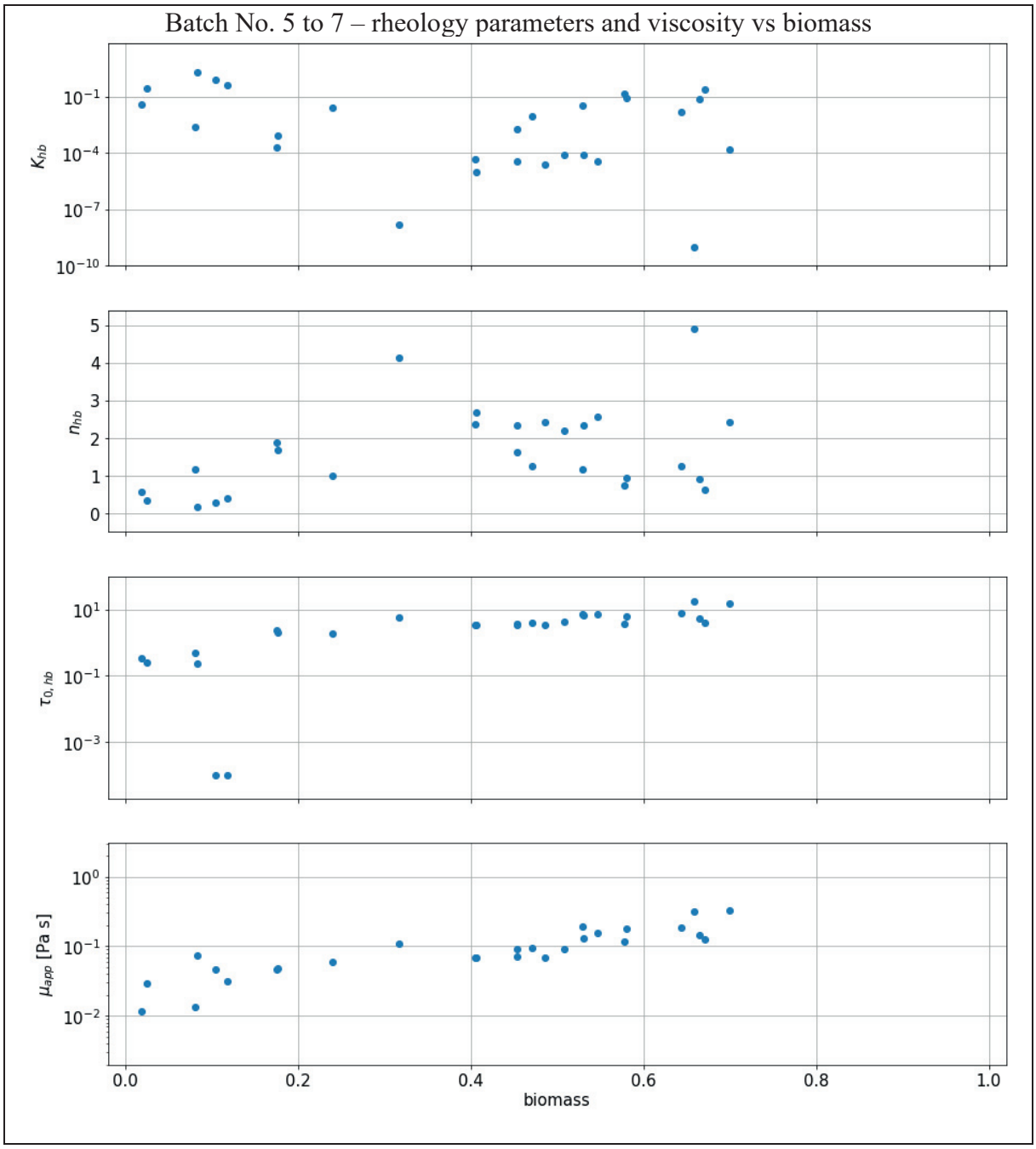

Figure 7.2. Rheology parameters using the Herschel–Bulkley model vs biomass as well as viscosity vs biomass for the samples from Batch No. 5-7. The host organism was *A. oryzae, strain 1.* The y-axes for k_{hb}, $\tau_{0,hb}$ and μ_{app} were log-transformed for better visualization. Note that the biomass measurements only show relative values on purpose due to proprietary reasons.

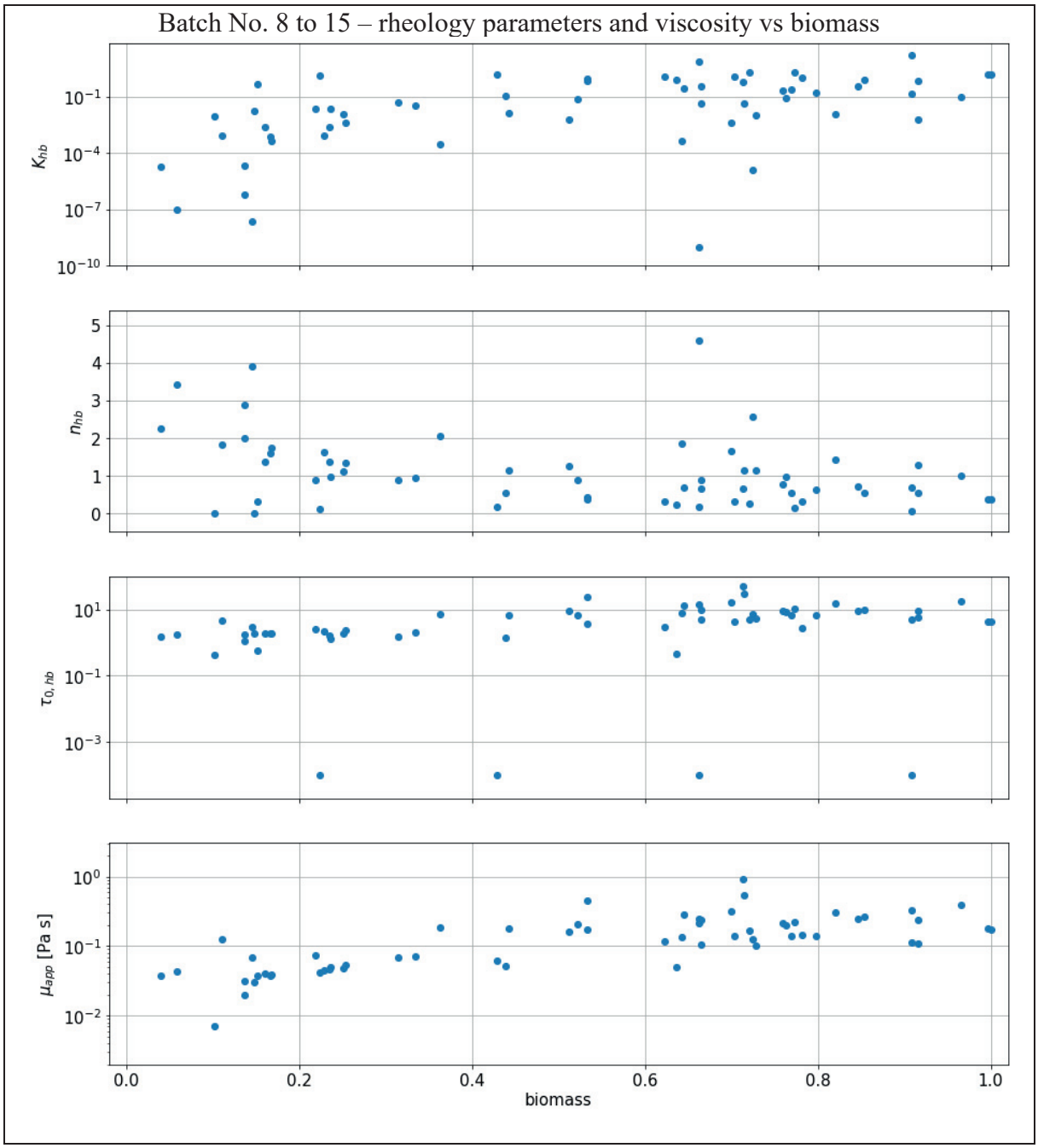

Figure *7.3.* Rheology parameters using the Herschel–Bulkley model vs biomass as well as viscosity vs biomass for the samples from Batch No. 8-15. The host organism was *A. oryzae, strain 2.* The y-axes for k_{hb}, $\tau_{0,hb}$ and μ_{app} were log-transformed for better visualization. Note that the biomass measurements only show relative values on purpose due to proprietary reasons.

Printed by Media-Tryck, Lund 2022 NORDIC SWAN ECOLABEL 3041 0903

LUND
UNIVERSITY

LUND UNIVERSITY
Faculty of Engineering
Department of Chemical Engineering