Online biomass estimation via a robust soft sensor

MASTER-THESIS

in partial fulfilment of the requirements for the degree of

Diplom-Ingenieur

in the subject of

Chemical and Process Engineering

by

Felix Pilz B.Sc.
Matr. Nr. 1126475

Institute of Chemical, Environmental and Bioscience Engineering

Supervisor: Univ.-Prof. Dipl.-Ing. Dr. techn. Christoph Herwig
Dipl.-Ing. Dr. techn. Paul Kroll

Vienna, 08 January 2019

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Author

Vienna, 08 January 2019

Acknowledgement

Inspired by the laboratory practice of process engineering in 2014, I knew since then that the combination of biology and technology would fascinate me for many years to come. My first experience was during bachelor's degree under the guidance of Dr. Nicole Mahler and the “BioVT Tempel” has not let me go since then. In my search for an exciting master’s thesis, I was thankfully contacted by Dr. Paul Kroll who suggested this topic to me. In this context, I would like to thank my two supervisors Prof. Christoph Herwig and Dr. Paul Kroll for the opportunity to perform this thesis in the interdisciplinary field of bioprocess technology and for the many great hours with your excellent research groups. I have learned a lot from you for my personal future. A special thanks goes to Paul, who supported me whenever I ran into a troublesome spot or had questions and steered me in the right the direction whenever he thought I needed it.

Another special thanks go to Dipl.-Ing Kager and his brother Pirmin for the energetic support and assistance. I would also thank the members of the Research Alliance group who supported and advised me especially in my work, on how to perform a successful cultivation.

Finally, I must express my very profound gratitude to my parents for providing me with tireless support and continuous encouragement throughout my years of study and through the process of writing this thesis. This accomplishment would not have been possible without them. A heartfelt thank-you.

1 Abstract

Bioprocesses require efficient monitoring and control of the respective process parameters. In order to achieve defined states to produce a product in high quantities and required quality, the respective time-dependent biomass concentration as a catalytic converter is a key variable. However, several offline measurements such as the optical density or dry weight measurements exist but repeatedly do not satisfy the requirements for modern bioprocess development. Therefore, hard sensor types based on permittivity measurements and dielectric spectroscopy were used to determine the biomass online. However, these measurements come with different challenges and disadvantages.

An alternative to these hard-type sensors are soft-type sensors, which indirectly estimate the biomass produced by existing measurements with a bioreaction system in real-time. Therefore, the known inlet, outlet and feed streams were quantified to calculate rates. These rates are used to develop a balance system according the law of mass conservation for carbon, nitrogen and degree of reduction.

This overdetermined model system of equations can be solved with calculated rates or reconciled rates for the best possible estimation of the biomass. The estimated biomass amount is used for different specific substrate uptake rates during the experiment. For a trustworthy estimate of the biomass, a statistical test has been introduced which can distinguish whether a systematic error exists or whether the deviations of the rates can be explained with random errors. In this context, the error propagation of the measurements was considered in real-time as a novel approach. The investigation of the soft sensor concept shows that in the case of a valid model, the degree of reduction balance with the calculated rates of the measured quantities and the C-balance with the reconciled rates best describe the bioreaction. There were achieved accuracies below 10% for the estimation for the biomass relative to reference measurements such as optical density and dry weight measurements demonstrated with a *Saccharomyces cerevisiae* cultivation. A workflow was developed which should guarantee a successful fermentation using this robust soft sensor concept.

2 Table of contents

3 List of Symbols

Symbols	Description	Unit
$\dot{F}_S$	feed of substrate	l/h
q_s	specific substrate uptake rate	g/(g·h)
c_s	substrate concentration	g/l
X_{stp}	current biomass setpoint	g
r_X	biomass rate	c-mol/h
r_S	substrate rate	c-mol/h
r_{O_2}	oxygen uptake rate	mol/h
r_{CO_2}	carbon dioxide evolution rate	mol/h
r_N	nitrogen consumption rate	mol/h
r_m	measured rates vector	-
E	Elemental Matrix	-
E_m	measurement Elemental Matrix	-
E_c	calculation Elemental Matrix	-
${E_c}^{\#}$	pseudo inverse calculation Elemental Matrix	-
R	Redundancy matrix	-
r_{mb}	best estimates of r_m	-
β	measured error vector	-
R_r	reduced redundancy matrix	-
ε	residue vector	-
$U \cdot S \cdot V$	Singular value decomposition Matrix	-
δ	measurement error vector	-
e_m	relative error vector of r_m	-
F	variance covariance matrix of r_m	-
φ	covariance matrix of the residuals	-
h	statistical index value h	-
$\hat{r}_m$	reconciled measurement vector	-
$\dot{V}_{in}$	input volume stream	mol/h
$c_{i,in}$	input volume stream concentration	mol/m³
$\dot{V}_{out}$	output volume stream	mol/h
$c_{i,out}$	output volume stream concentration	mol/m³
V_R	reactor volume	m³
r_i	rates i	mol/(m³·h)
$\dot{n}_{i,in}$	input mol stream i	mol/h
$\dot{n}_{i,out}$	output mol stream i	mol/h
n_{gas}	gas amount in reactor	mol
p_{norm}	pressure under normal conditions	101325 Pa
$\dot{V}_{j,in,norm}$	inlet volume stream j under normal conditions	m³

R	Universal gas constant	8,314 J/(mol·K)
T_{norm}	temperature under normal conditions	273,15 °K
V_{gas}	volume of gas in system	m³
V_{Rt}	reactor total volume	m³
V_R	current reactor volume	m³
$m_{R(t)}$	measured reactor weight	kg
m_{R0}	initial measured reactor weight	kg
ρ	density of broth	kg/m³
V_{R0}	initial reactor volume	m³
p_R	reactor pressure	Pa
T_R	reactor temperature	°K
n_{inert}	inert gas in reactor	mol
TR	transport rate	mol/h
$y_{inert,off}$	inert gas fraction in offgas	-
$y_{O_2,off}$	oxygen fraction in offgas	-
$y_{CO_2,off}$	carbon dioxide fraction in offgas	-
$y_{H_2O,off}$	water fraction in offgas	-
$y_{o_2,wet}$	oxygen fraction in wet offgas	-
$y_{O_2,air}$	oxygen fraction in air	-
$\dot{V}_{in,\,inert,norm}$	inert volume stream under normal condition	m³/h
$\dot{V}_{in,N_2,norm}$	N_2 inlet volume stream under normal condition	m³/h
$\dot{V}_{in,air,norm}$	air inlet volume stream under normal condition	m³/h
TR_{inert}	transport rate of inert gas	mol/h
$\dot{n}_{out}$	outlet stream	mol/h
$\dot{n}_{out,inert}$	outlet stream of inert gas	mol/h
$e_{abs,i}$	absolute error of rate i	mol/h
$e_{noise,i}$	absolute noise error of rate i	mol/h
Δf_{r_i}	propagation of uncertainty of rate i	mol/h
w_S	substrate weight signal	g
c_{Feed}	concentration of feed	kg/m³
c_{mol}	c-mol fraction of glucose	6
ρ_{Feed}	density of feed	kg/m³
M_{Feed}	molar mass of glucose	g/mol
Δe_{w_S}	absolute error of substrate feed weight signal	g
$\Delta e_{c_{Feed}}$	absolute error of feed concentration	kg/m³
w_{NH3}	ammonia weight signal	g
c_{NH3}	concentration of ammonia	kg/m³
ρ_{NH3}	density of ammonia feed	kg/m³
M_{NH3}	molar mass of ammonia	g/mol
$\Delta e_{w_{NH3}}$	absolute error of ammonia feed weight signal	g
$\Delta e_{c_{NH3}}$	absolute error of ammonia feed concentration	kg/m³

F_{AIRin}	air inlet stream variable	Nl/min
F_{O2in}	oxygen inlet stream variable	Nl/min
$\Delta e_{F_{AIRin}}$	error of air inlet stream variable	Nl/min
$\Delta e_{F_{O2in}}$	error of oxygen inlet stream variable	Nl/min
$\Delta e_{y_{O_2,off}}$	error of oxygen fraction in offgas	-
$\Delta e_{y_{CO_2,off}}$	error of carbon dioxide fraction in offgas	-
$y_{02_{wet},off}$	oxygen fraction in wet offgas	-
X	biomass	g
μ_{max}	maximum specific growth rate	1/h
S	substrate	g
K_S	saturation constant of substrate	g/l
$Y_{\frac{X}{S}}$	yield biomass/substrate	g/g
OTR	oxygen transfer rate	g/(l·h)
kla	volumetric mass transfer coefficient	1/h
c^*	saturation concentration	g/l
c_{O_2}	current oxygen concentration	g/l
$e_{c,rel(t)}$	current relative error of carbon balance	%
$e_{c,abs(t)}$	absolute error of carbon balance	mol
$e_{DoR,rel(t)}$	current relative error of degree of reduction balance	%
$e_{DoR,abs(t)}$	current absolute error of degree of reduction balance	mol
$e_{N,rel(t)}$	current relative error of nitrogen balance	%
$e_{N,abs(t)}$	current absolute error of nitrogen balance	mol
$e_{c_{rconc},rel(t)}$	current relative error of carbon balance after rates reconciliation	%
$e_{c_{rconc},abs(t)}$	absolute error of carbon balance after rates reconciliation	mol
$e_{DoR_{rconc},rel(t)}$	current relative error of degree of reduction balance after rates reconciliation	%
$e_{DoR_{rconc},abs(t)}$	current absolute error of degree of reduction balance after rates reconciliation	mol
$e_{N_{rconc},rel(t)}$	current relative error of nitrogen balance after rates reconciliation	%
$e_{N_{rconc},abs(t)}$	current absolute error of nitrogen balance after rates reconciliation	mol

4 Introduction

4.1 Motivation

All bioprocesses involve microorganisms that use nutrients to form a product of interest. Furthermore, a fundamental goal of bioprocesses is to produce a product corresponding with a required quality. This requirement involves continuous process control and monitoring. Differential typical process parameters such as temperature, pH, O_2 and CO_2 in the offgas, biomass etc. can be measured offline and online measured and provide fundamental information about the state of bioprocesses. Moreover, the data collected online and offline form the basis for a better monitoring and controlling process, as well as for model development based on cell specific characteristics. A defined quality and high yield of the product referring to the broth volume can be achieved by advantageous cultivation conditions and a defined physiological state of the culture. In this regard, Fed batch processes are the method of choice to control the cell metabolism by a substrate limiting feed profile and further parameters such as temperature, pH and the dissolved O_2 concentration. In order to ensure a consistent limited substrate supply for the growing biomass, the feed is regulated as a function of the biomass quantity. In this context, a key variable is the biomass concentration.

Differential online and offline methods for the biomass concentration determination exists [1]. In general, offline measurement methods are time-consuming and come with high operator-dependent measurement errors. The current biomass concentration of the bioprocess, which can be detected by online sensors, can be classified in hard-type and soft type sensors [2]. Several hard type sensors exists based on different methods (residual transmission, microscope, fluorescence, infra-red absorption, capacitance, impedance) which are able to do an in-line quantification of the biomass concentration [1] [3] [4] [5]. It is common that challenges occur in these measurements concerning: calibrations, dynamic range, interferences due to different process modes, physiological states of the culture, media composition in the process. Moreover, the dry weight measurements as reference for the calibrations can also vary depending on changing process characteristics [6] [7]. In this regard, hard type sensors do not often meet the required accuracy in determining the biomass concentration due to the changing process conditions and challenges that focus on the calibration.

Another approach to indirectly quantify the biomass concentration online is to use the accessible measurements such as the quantified feed, gas inlet and outlet streams with its oxygen and carbon dioxide concentrations to determine the biomass via mass conservation. This indirect determination can be summarized as soft type sensors. Several works were done which demonstrate the functionality of so called soft sensors for the biomass estimation.[8] [9] [10] [11].

However, in process development time is a limiting factor so a reasonable strategy is to use available measurements in combination with an overall bioreaction describing model based on mass conservation. This balance system can be solved for the variable of interest, in this case the biomass concentration. In this context, the soft sensor technology needs easily accessible measurements, models and estimation algorithms to work. Previous works investigated the calculability and balance-ability including diagnostics and error handling of such biochemical networks [12] [13] [14] [15] [16].

If a soft sensor concept for estimating the biomass is implemented, the operator must consider the physiological characteristics, technical limits and model assumptions of the system to have reliable biomass estimation results. A successful run can be expected when the process is performed in the intersection of all three circles shown in Fig. 1.

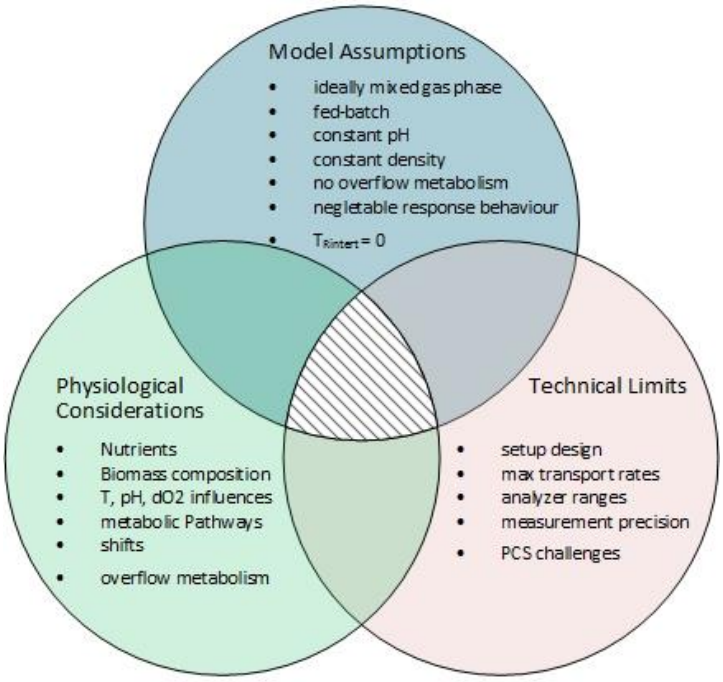

Fig. 1 Venn diagram for successful microbial fermenting considerations. In the hatched space, felicitous controlled processes meet the model assumptions, do not cross technical limits and deal with the physiology of the used strain.

The reasons for setting up soft sensor concepts today lies in the simple fact that direct measurements of the variables of interest are frequently too expensive, difficult to install or simply not reliable or accurate enough. A soft sensor provides continuous estimates of the variable instead of expensive, periodically performed measurements. Furthermore, the direct instrumentation for the direct measurement could make the process more difficult to control. It could be also possible that no instrumentation exists that can directly measure the process variable such as biomass concentration due to harsh conditions or simply lack of space in the bioreactor. In this case, a soft sensor concept is a suitable way to estimate the variable of interest. Moreover, with implementing a soft sensor, a model exists which is fed with measured data of the system. Additionally, the errors on each measurement can be considered for the calculation. This could also be used for a simple method to improve the reliability accuracy and fault detection of the model and data acquisition. Similarly, combining of information from different sensors allows us to combine the strength of each technology and reduce the weaknesses of the estimation. In this view, redundancy can be introduced in the measurement model system to provides precise estimates of the unmeasured variable.

Derived from these advantageous aspects, this master thesis shows a general introduction on implementing a soft senor concept with its important considerations for various microbial cultivation processes.

4.2 Problem statement and goals of the work

A brief overview about pervious cultivations and process behaviours lead to the following challenges.

Challenge I

Soft sensor concepts for the estimation of variables of interest are widespread in industrial applications. This technology is not popular in laboratory scale cultivation experiments.

Goal I: The first goal of this work is to find the problems as to why it is uncommon to use a soft sensor technology for estimating the biomass during cultivation runs. Furthermore, an identification of the "issues" in the technical, handling and software section should provide necessary information. In this context, a developed strategy should avoid the "issues". This should be a basis for a successful soft sensor development.

Challenge II

Specific feeding profiles are related to the biomass concentration in the reactor. The amount of biomass is not available in real-time due to the lack of reliable and accurate biomass quantification results from online direct determinations.

Goal II: A soft sensor concept should be developed to estimate the biomass amount with easily accessible measurements. The provided biomass determination should satisfy the accuracy of a specific substrate uptake rate control strategy. This concept should be developed in a multi-paradigm numerical computing environment. Finally, a proof of this novel soft sensor approach must be done.

Challenge III

A soft sensor concept should be reliable, robust and easily adaptable to other processes in a bioprocess development environment. In this context, no works investigate the behaviour of the applied soft sensor with the in real-time considered measurement errors.

Goal III: A detailed analysis of error propagation should be done for a powerful reconciliation procedure, a robust estimation of the biomass and a meaningful statistical test. Finally, a workflow should be established for a successful biomass estimation via this robust soft sensor concept.

5 Materials & Methods

5.1 Cultivation

A yeast strain, *Saccharomyces cerevisiae*, CBS 8340-Wild type, was used in a defined media according to Table 1 to Table 3. The shake flask culture for the batch phase inoculation were maintained in 200 ml H_2O solution. It contained 20 g/l glucose, 10 g/l yeast extract and 20 g/l peptone under pH 4.8, 200 rpm and 30 °C in laboratory incubator.

Table 1: Medium composition for the batch and fed-batch

Medium	quantity	unit	Annotation
Glucose	200	g/l	Fed Batch; 2 Liter
	20	g/l	Batch, 3 Liter
(NH4)2SO4	5	g/l	
KH2PO4	3	g/l	
MgSO4*7H2O	0,5	g/l	
Struktol J 650	0,1	g/l	
Trace Elements (750x)	1,33	g/l	
Vitamines (750x)	1,33	g/l	

Table 2: Trace elements for the batch and fed-batch medium.

Trace elements (750x) pH 4	quantity	unit	Annotation
EDTA	15	g/l	
ZnSO4*7H2O	4,5	g/l	
MnCl2*4H2O	1	g/l	
CoCl2*6H2O	0,3	g/l	
CuSO4*5H2O	0,3	g/l	
Na2MoO4*2H2O	0,4	g/l	
CaCl2*2H2O	4,5	g/l	
FeSO4*7H2O	3	g/l	
H3BO3	1	g/l	
KI	0,1	g/l	

Table 3: Vitamins for the batch and fed-batch medium.

Vitamins (750x) pH 6,5	quantity	unit	Annotation
Biotin	0,05	g/l	
Ca pentothenate	1	g/l	
Nicotinic acid	1	g/l	
m-inositol	25	g/l	
Thiamin hydrochloride (B1)	1	g/l	
Pyridoxine hydrochloride(B6)	1	g/l	
Para-amino benoic acid (PABA)	0,2	g/l	
FeSO4*7H2O	3	g/l	
H3BO3	1	g/l	
KI	0,1	g/l	

In the batch and fed batch phase, the set points of temperature (30°C), pH (4,8), dissolved oxygen concentration (35 %), initial aeration (1,5 l/(l·min)) and initial stirrer speed were chosen in compliance to the supplied information from previous experiments with this *Saccharomyces cerevisiae* yeast strain. As the growing biomass amount will be available online in this approach, a q_s (specific substrate uptake rate) control feed strategy is the method of choice. The fed batch run was realised with different constant q_s. The pH control was performed with a 6,25 w% NH_3 solution as base and 6,25 w% H_3PO_4 solution as acid.

5.2 Offline and Online Analytics

Several sampling routings were made to compare and validate the results with other biomass estimation methods as demonstrated here by OD (optical density) measurement and DW (dry weight) measurement. The biomass was determined with an OD –three-fold determination and a DW-five-fold determination. The sampling interval was chosen to be increased at higher q_s and reduced at lower q_s to follow a trend of biomass growth behaviour. The glucose and NH_3 concentrations during the experiment were analysed for each sample routine by an automated analyser (Cedex BioHT Biochemical analyzer, Roche, Mannheim, Germany). In order to have an online biomass determination tool, an online microscope (Ovizio Imaging Systems, Brussels, Belgium) was added to compare the estimated biomass via the microscope result.

5.3 Hardware

A glass bioreactor with a total volume of 7,5l and a working volume of 5l was used (Infors 3, Bottmingen, Switzerland). The controlled parameters, pH and temperature and cooling system were regulated with an internal controller unit on the bioreactor system. The base and acid were observed by using attached peristaltic pumps.

The following devices in Table 4 have been installed to calculate the input and output streams. Scales were used to define the input flow stream of glucose and the NH_3 solution as base. In addition, the reactor vessel was also weighted by a scale to determine the reactor volume. Two mass flow controllers were used to quantify the gas inlet flow streams of air and pure oxygen. Moreover, the oxygen and carbon dioxide concentration were measured by an offgas analysator.

Table 1 lists the measured process variables including the used devices and necessary general features such as range and accuracy/readability for considering the measurement errors.

5.4 Software

The process control system Lucullus PIMS (Securecell, Switzerland) was used for controlling the temperature, pH, dissolved oxygen and substrate feed controlling. MATLAB (MathWorks, USA) as a multi-paradigm numerical computing environment was used for enabling the soft-sensor concept. A link from Lucullus to a database phyton server via OPC interface and afterwards a Matlab connection to this database enables an online computing source for controlling and data analysis. Fig. 2 shows the schematic setup of this connection.

Table 4: Measured process variables including the used devices and necessary general features

Variables	labelling	Principle	Device	Range	Accuracy / Readability
m_R	103.3	Scale	Kern & Sohn ITS 35 K1IP	max. 35 kg	± 0.1g,
T_R	103.5	-	PT 100	80 °C	-
pH	103.7	-	Hamilton Easyferm Plus HB K8 425	0 to 14	57 to 59 mV / pH at 25 °C
O_2	101.3	-	Hamilton VISIFerm DO 425	0-100%	± 0.2 %
rpm	101.5	-	-	80 - 1400 rpm	-
$y_{O_2,off}$	102.4	ZrO_2	BlueInOne Ferm	0 – 50%	±0.02%
$y_{CO_2,off}$	102.5	IR	BlueInOne Ferm	0-25%	< ±1%
$\dot{V}_{Air,in,norm}$	101.1	MFC	Vogtlin Instruments GSC B5SA BB26	5 NL/min	± 1% F.S
$\dot{V}_{O_2,in,norm}$	101.2	MFC	Brooks 4800 Series	10 NL/min	+/- 3,0% of F.S
m_s	102.1	Scale	Sarturius Signum 1	max. 35 kg	± 0,1g, d=0,1g
m_n	103.1	Scale	Sarturius CPA 34001S	max. 34 kg	± 0,1g d=0,1g

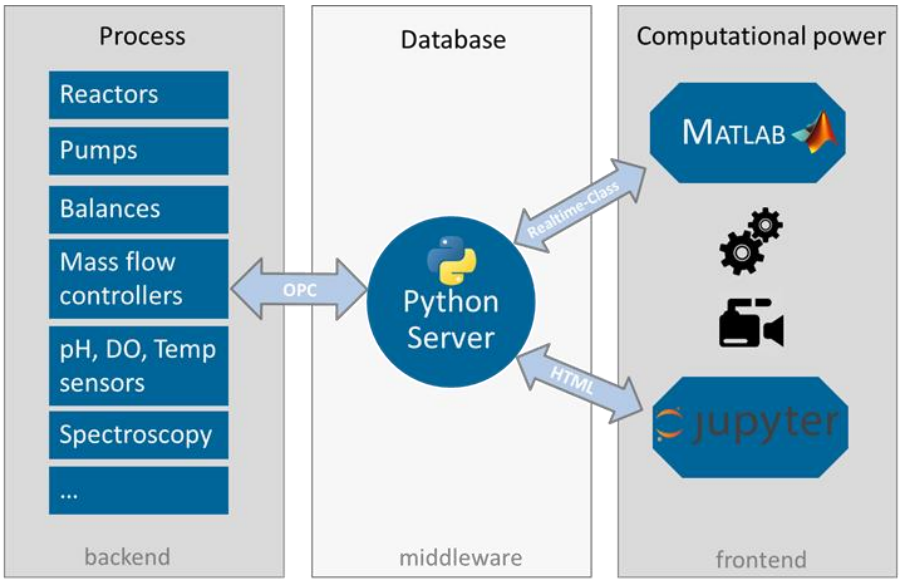

Fig. 2: Schematic structure of the online connection from local process control system via OPC-Python Server to Matlab.

The entire process flow diagram designed for the experiment is shown in Fig. 3.

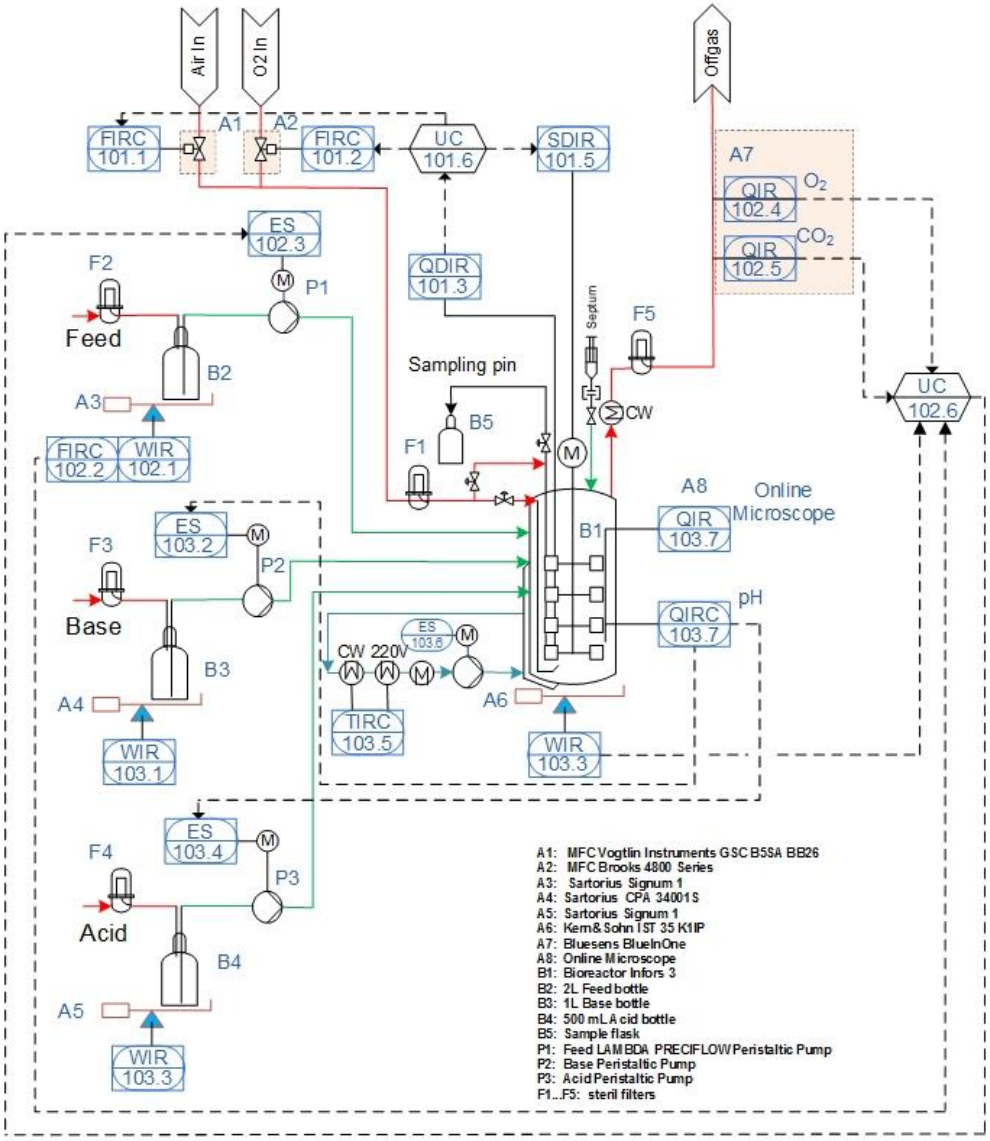

Fig. 3: Process-flow-Diagram with all basic equipment for measuring and controlling culture parameters. The dissolved O_2 is a closed loop controlled via Lucullus Software (UC 101.6). pH value is controlled via internal Infors software tool with peristaltic pumps. The q_s control feed strategy is controlled via OPC server with the Bioprocess Technology Tool (BPTT) in Matlab 2017b. The red lines represent gas streams and the green lines fluid streams.

5.5 Control strategy

In general, for successful cultivations, a well-working process control strategy is indispensable. Principally, microorganisms need defined process conditions to develop their optimal performance. If the regulation of the real values parameters such as pH, dissolved O_2, temperature, feeding profile is not satisfying the desired conditions, arbitrary pathway switching, limitations and finally cell death can be possible. Hence targeted cultivations, and high productivities are not achievable. Based on previous literary work, a cascade control subfunction for three steps O_2 control [17] [18] has been built, which is described in more detail, see subsection 5.5.1. The q_s control strategy is realized as closed loop subfunction which is explained in subsection 5.5.2. These subfunctions were implemented in the process information management system Lucullus. The closed loop controllers are schematic shown in Fig. 4.

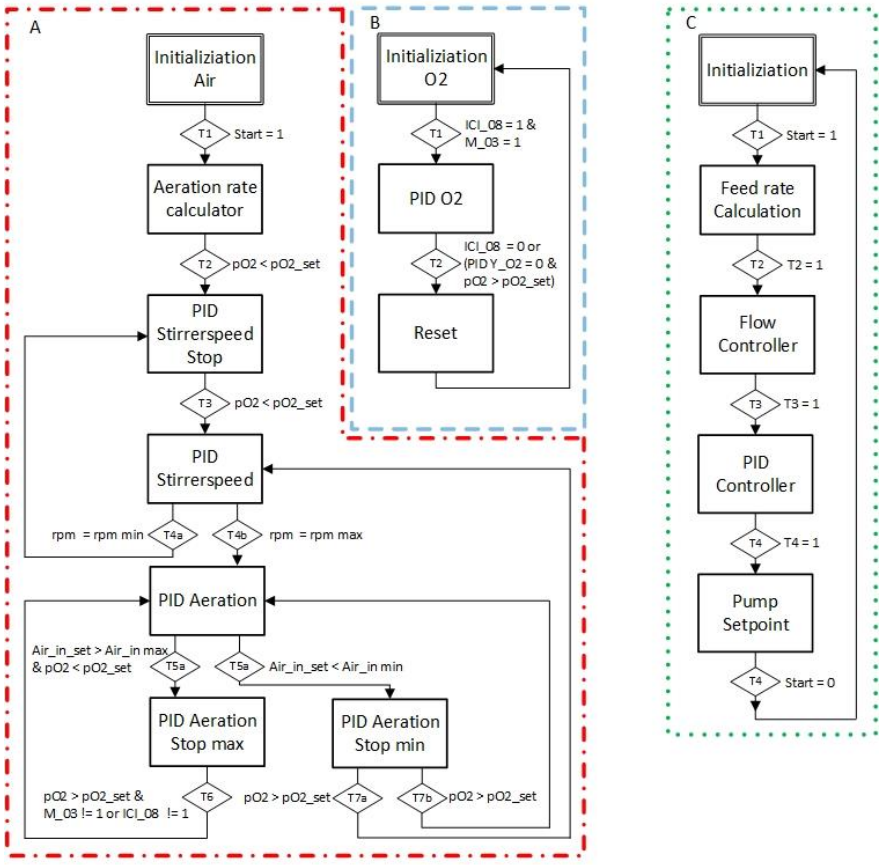

Fig. 4: Sequential control of pO_2 regulation and q_s setpoint control unit.
A: cascade control function unit with primarily stirrer speed regulation and as second step controlling the air gassing. **B:** Step function for constant volume flow with chancing air/O_2 ratio. **B** is enabled if **A** is not satisfying the required dissolved oxygen concentration. **C:** Step function for PID controlled feed pump setpoint dependent on q_s.

5.5.1 O_2 control

Due to the fact, that cultivations should not be restricted to oxygen, the concentration of dissolved oxygen should be regulated to a certain level to ensure the oxygen uptake. Therefore, two subfunctions were established for robust versatile cultivation applications. The subfunction **A** shown in Fig. 4 uses at first the agitator speed and as second step the aeration rate to regulate the O_2 concentration. Each actuating variable is a closed loop controlled via a PID controller subdevice in which the derivative term is zero. The limits for the stirrer speed, aeration rate and the initial gassing rate in vvm ($l/(l_{volume} \cdot min)$) are set in the initialization step. The subfunction is enabled if the O_2 concentration is under the critical chosen value. In this phase the stirrer speed is regulates the dissolved oxygen. If the maximum stirrer speed is reached and is not able to hold the dissolved O_2 on the setpoint, the PID controller for the aeration rate is enabled. The stirrer speed at the so called PID Aeration steps is still at the maximum. Steps with a *Stop* in the labelling are for stopping the PID internal calculating algorithm. The aim is to avoid overflows and in addition to make it resistant against handling errors.

If the controller of subfunction **A** cannot get the dissolved oxygen level, the extend subfunction **B** will be activated. This controller strategy is based on a constant gassing flow with changing air/pure oxygen ratio and maximum stirrer speed. The PID controller is changing the gas composition to hold the O_2 concentration on the setpoint. The limits of the O_2 mass flow controller and setpoint of the constant flow throw the reactor can be adjusted in the initial step.

This cascade strategy with subfunction **A** and **B** can be run in both directions. Adjustments of the values for the proportional and integral terms in the PID controller subdevices depends primarily on the actuating variable size units and the calculating intervals. As an approximate initial adjustment, the proportional term is 1% of the lower limits from each actuating variable entered in the initial step. The integral term is approximately a tenth of the proportional term.

5.5.2 Feed control

The controlled q_s constant strategy was realized with a PID and Flow controller app sub device in Lucullus which is shown in Fig. 4.

C. The flow controller app returns the real feed rate calculated by the changing feed weight signal. This calculated real feed rate is the controlled process variable for the PID controller subdevice. Step *Pump setpoint* calculates the setpoint for the feed pump. The current biomass setpoint X_{stp} was received by Python Server and OPC link from Matlab. q_s and the substrate concentration c_s of the feed are set in the initialization step. Equation Eqs. 5-1 shows the relation for the feed flow.

$$\dot{F}_S = \frac{q_s \cdot X_{stp}}{c_s} \qquad \textit{Eqs. 5-1}$$

5.6 Conditions for application

The created program presents a soft sensor for microorganism in fed-batch dynamic process conditions to indirectly determine the amount of biomass. The biomass estimation and reconciliation with this soft-sensor tool was possible with known elemental composition of the biomass in (C, H, N, O) and quantified inlet flows with defined media content. The shown application is a base for one substrate and a nitrogen source.

Furthermore, the balance equation system is constructed for the carbon-, degree of reduction- and nitrogen balance. In this contrast, it must be stated that the balances only describe the overall bioreaction and that there are a lot of other reactions beside this basic specification. Data input reconciliation cannot be performed without redundancy. That means the system of equations must be overdetermined. If needed, gross error detection is only feasible by using all three stated balances. In order to execute the calculations as precise as possible, the measurement frequency shall be high as possible. Furthermore, the linear regression window size for the calculated input rates should be chosen as low as possible to reduce the impact of average values.

The defined equation system for describing the overall bioreaction use various assumptions as follows:

- fluids are ideally mixed
- no pH changes during the cultivation
- constant density and temperature of the cultivation broth
- neglectable sensor response behaviour
- neglectable transport rate of inert gas
- no change of biomass composition
- product formation is not considered
- no accumulation of substrate
- no accumulation of nitrogen

5.7 Soft sensor concept

It is a good choice to use a soft sensor to estimate the unmeasured biomass concentration during a culture run. For this purpose, a model will be created according to the conservation law and its physical limitations to describe the bioreaction. Assuming oxidative metabolism, the bioreaction can be described according Eqs. 5-2 for the growth of *Saccharomyces cerevisiae* [19]. It should be noted that many different chemical reactions are running in a living cell. For this reason, the equation Eqs. 5-2 applies only to a general bioreaction formulation.

$$r_S \cdot CH_{pH}O_{pO} + r_{o_2} \cdot O_2 + r_N \cdot NH_3 \rightarrow r_x \cdot CH_{zH}O_{zN} + r_{Co_2} \cdot CO_2 \qquad \textit{Eqs. 5-2}$$

In order to estimate unmeasured variables at a given set of independent system constraints, a minimum number of measurements are required to calculate the estimates. In this case the system is determined. If there are more measurements than the minimum, redundancy exists in the measurements which can be exploited for data reconciliation. The overall bioreaction can be split and written as elemental balances for carbon and nitrogen, according to Eqs. 5-3. For donating and receiving electrons of each molecule, the degree of reduction (DoR) balance describes the electrons exchange. The elemental balances can be summarized to an elemental matrix and a rate vector analogous to the procedure in previous works [19] [20] [21] [22]. The elemental matrix and the rate vector are constructed as follows in Eqs. 5-4 and Eqs. 5-5.

$$Input - Output + conversation = accumulation \qquad \textit{Eqs. 5-3}$$

$$E = \left(\begin{array}{c|c|c|c|c|c} \textit{Bal.} & X & S & O_2 & CO_2 & NH_3 \\ \hline C & 1 & 1 & 0 & 1 & 0 \\ \hdashline DoR & 4{,}18 & 4 & -4 & 0 & 0 \\ \hdashline N & 0{,}176 & 0 & 0 & 0 & 1 \end{array}\right) \qquad \textit{Eqs. 5-4}$$

$$r_i = \left(\begin{array}{c} r_X \\ \hdashline r_S \\ \hdashline r_{O_2} \\ \hdashline r_{CO_2} \\ \hdashline r_N \end{array}\right) \qquad \textit{Eqs. 5-5}$$

The matrix is defined by means of known information of the biomass and substrate composition. For general purpose, the elemental matrix system (Eqs. 5-4) with its carbon containing sources is standardized on *c-mol*. Besides the rate r_X, the remaining rates (r_S, r_{O_2}, r_{CO_2}, r_N) are calculated by measured variables illustrated in subchapter 5.7.1. In this context, the rates are positive donated if production or negative donated in the case of consumption.

The matrix structure of the balance system enables math calculation operations. Several insightful literature[19] [20] [21] [23] exists for solving the system to estimate the unmeasured rates in over determined systems. Therefore, a full treatment of their foundations was not the subject of this work. Following operations and brief descriptions explain the strategy based on that.

Assuming the balance model satisfies the overall bioreaction without faulty data by convention, the system can be summarized according Eqs. 5-6.

$$\sum_{i=1}^{k} r_i \cdot E \stackrel{\text{def}}{\equiv} 0 \qquad \textit{Eqs. 5-6}$$

It is possible to partition the elemental matrix E and the rate vector r_i in measured and calculating components in Eqs. 5-7. The biomass production quantified by the production rate r_X is member of the unknown vector r_c which can be determined with Eqs. 5-8 by using the pseudo-inverse of of E_c.

$$r_m \cdot E_m + E_c \cdot r_c \stackrel{\text{def}}{\equiv} 0 \qquad \textit{Eqs. 5-7}$$

$$r_c = -{E_c}^{\#} \cdot E_m \cdot r_m \qquad \textit{Eqs. 5-8}$$

The redundancy matrix R defined by Eqs. 5-9 expresses the relations between all measured rates and inform about balance ability and redundancy. The rank of R (Degree of Redundancy) equals the independent balances and must be greater than one for gross error detection.

$$R = E_m - E_c \cdot {E_c}^{\#} \cdot E_m \qquad \textit{Eqs. 5-9}$$

As another requirement, the rank of R must be greater or equal to the number of tested balances minus the number of calculated components for the full calculability of non-measured rates. The remaining number of equations for reconciliation alignment, called degree of freedom S is the number of tested balances minus the rank of the redundancy matrix R.

The goal of rates reconciliation is to obtain the best estimates rm_b of measured variables rm via least square minimization procedure in Eqs. 5-10 with the unknown measured error vector β.

$$r_{mb} = r_m + \beta \qquad \textit{Eqs. 5-10}$$

It is obvious that the Eqs. 5-6 is not exactly met due to measurement noise. Therefore, the residue vector ε is calculated according Eqs. 5-11.

$$\varepsilon = R_r \cdot r_m \qquad \textit{Eqs. 5-11}$$

For further calculations, a reduced redundancy matrix R_r is needed because the inverse of the variance-covariance matrix of the residues φ does not exist or lead to incorrect results. Therefore, singular value decomposition (SVD) was executed on R to find the non-singular values of R according Eqs. 5-12.

The reduced redundancy matrix is computed from the right-hand singular vector V. Using j columns $(1 \leq j \leq S, \quad \mathrm{S} \ldots \mathrm{degree\ of\ freedom})$of vector V, the reduced redundancy matrix R_r is generated (Eqs. 5-13).

$$R = U \cdot S \cdot V \qquad \textit{Eqs. 5-12}$$

$$R_r = \left(V_{ij}\right)^T \qquad 1 \leq i \leq S \qquad \textit{Eqs. 5-13}$$

The measurement error vector δ contains information on the error in measured rates created by Eqs. 5-14. Further, a variance covariance matrix F of the measured rates is formed in Eqs. 5-15. This matrix is only diagonal, assuming that the variance-covariance matrix is uncorrelated.

$$\delta = e_m \cdot r_m \qquad \textit{Eqs. 5-14}$$

$$F = E(\delta \cdot \delta^T) \qquad \textit{Eqs. 5-15}$$

The covariance matrix of the residuals φ is calculated in Eqs. 5-16.

$$\varphi = R_r \cdot F \cdot {R_r}^T \qquad \textit{Eqs. 5-16}$$

With this information, the consistency index h according Eqs. 5-17 can be determined. The index h is a suitable value to test the significance of errors caused by errors either the measurement set or in the system description. In this case the residual vector differs significantly from zero.

$$h = \varepsilon \cdot \varphi^{-1} \cdot \varepsilon^T \qquad \textit{Eqs. 5-17}$$

It has been shown that the h value follows a X^2 distribution with the degrees of freedom equals the rank of R [24]. Moreover, the residuals are weighted according to their accuracy. If the h value is higher than the level of significance at the appropriate degree of freedom an error is significant.

Back to the data reconciliation to find a solution for the unknown measured error vector β in Eqs. 5-10. This problem can be solved according to [25]. The reconciled measurement vector $\hat{r}_m$ can be calculated by Eqs. 5-18.

$$\hat{r}_m = \left(I - F \cdot {R_r}^T \cdot \varphi^{-1} \cdot R_r \right) \cdot r_m \qquad \textit{Eqs. 5-18}$$

The reconciled measurement vector Eqs. 5-18 can be used in Eqs. 5-8 to calculate an improved biomass estimate.

5.7.1 Rates Calculators

The measured velocity vector is fundamental to a soft sensor design in addition to the element matrix. Furthermore, the errors on the rates by error propagation of measurement inaccuracy are taken into account as a novelty in this paper. The following calculated rates are set according to Fig. 5. The expressions in the following equations are described in chapter 3.

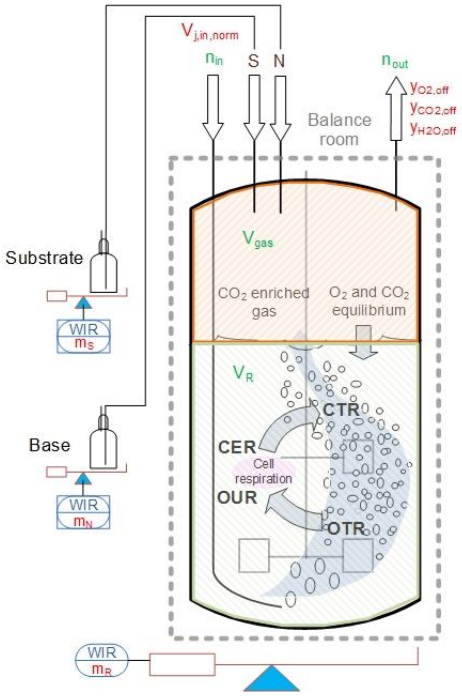

Fig. 5: Balance space for rates calculation. Directly measured process variables have font colour red, indirectly calculated major variables are green highlighted.

The general material balance for component i is formulated in Eqs. 5-19 is balanced as input minus output plus reaction expression and corresponds to the accumulation.

$$\dot{V}_{in}c_{i,in} - \dot{V}_{out}c_{i,out} + V_R r_i = V_R \frac{\partial c_i}{\partial t} + c_i \frac{\partial V_R}{\partial t} \qquad \textit{Eqs. 5-19}$$

The Eqs. 5-19 applies to all substances of the system such as substrate, nitrogen, oxygen, carbon dioxide and biomass [26] [27]. Applying of Eqs. 5-19 for calculating r_i with $i = S$ for substrate rate and $i = N$ for the nitrogen rate leads to Eqs. 5-20. Furthermore, one can assume that the accumulation term and the output mol current are zero according to Eqs. 5-21 The rates can be calculated by a difference quotient (Eqs. 5-22) from the measured signals according to Eqs. 5-23.

$$\frac{dn_i}{dt} = \dot{n}_{i,in} - \dot{n}_{i,out} + r_i \qquad \textit{Eqs. 5-20}$$

$$\frac{dn_i}{dt} = 0,\ \dot{n}_{i,out} = 0 \qquad \textit{Eqs. 5-21}$$

$$r_i = \dot{n}_{i,in} = \frac{\Delta n_i}{\Delta t} = \frac{n_{i+1} - n_i}{\Delta t} \qquad \textit{Eqs. 5-22}$$

$$\dot{m}_{i,in} = \frac{\Delta m_i}{\Delta t} = \frac{m_{i+1} - m_i}{\Delta t} \qquad \textit{Eqs. 5-23}$$

The outlet molar stream $\dot{n}_{out}$ must be determined before it is possible the calculate the gaseous rates r_{O_2} and r_{CO_2}. The outlet molar stream $\dot{n}_{out}$ is calculated from Eqs. 5-24 to Eqs. 5-37.

$$\frac{dn_{gas}}{dt} = \dot{n}_{in} - \dot{n}_{out} + TR \qquad \textit{Eqs. 5-24}$$

$$\dot{n}_{in} = \frac{p_{norm} \cdot \sum_{j=1}^{n} \dot{V}_{j,in,norm}}{R \cdot T_{norm}} \qquad \textit{Eqs. 5-25}$$

$$V_{gas} = V_{Rt} - V_R \qquad \textit{Eqs. 5-26}$$

The current reactor volume V_R is calculated from the measured balance signal $m_{R(t)}$ and the assumed constant broth density ρ.

$$V_R = V_{R0} + \frac{m_{R(t)} - m_{R0}}{\rho} \qquad \textit{Eqs. 5-27}$$

$$n_{gas} = \frac{p_R \cdot V_{gas}}{R \cdot T_R} \Rightarrow \frac{dn_{gas}}{dt} = \frac{\Delta n_{gas}}{\Delta t} \qquad \textit{Eqs. 5-28}$$

$$\frac{dn_{inert}}{dt} = \dot{n}_{in,inert} - \dot{n}_{out,\,inert} + TR_{inert} \qquad \textit{Eqs. 5-29}$$

$$\frac{\Delta n_{inert}}{\Delta t} = y_{inert,off} \cdot \frac{\Delta n_{gas}}{\Delta t} \qquad \textit{Eqs. 5-30}$$

$$y_{inert,\,off} = 1 - y_{O_2,off} - y_{CO_2,off} - y_{H_2O,off} \qquad \textit{Eqs. 5-31}$$

The water content is calculated in Eqs. 5-32 with the measured $y_{o_2,wet}$ in the wet outlet stream, provided there is no bioreaction in the reactor.

$$y_{H_2O,off} = 1 - \frac{y_{o_2,wet}}{y_{O_2,air}} \qquad \textit{Eqs. 5-32}$$

$$\dot{V}_{in,\,inert,norm} = \dot{V}_{in,N_2,norm} + \left(1 - y_{O_2,air} - y_{CO_2,air}\right) \cdot \dot{V}_{in,air,norm} \qquad \textit{Eqs. 5-33}$$

$$\dot{n}_{in,\,inert} = \frac{p_{norm} \cdot \dot{V}_{in,\,inert,norm}}{R \cdot T_{norm}} \qquad \textit{Eqs. 5-34}$$

The transport rate TR_{inert} of the inert gas can be supposed to be zero.

$$TR_{inert} = 0\ mol/h \qquad \textit{Eqs. 5-35}$$

$$\dot{n}_{out,inert} = \dot{n}_{in,inert} + TR_{inert} - \frac{\Delta n_{inert}}{\Delta t} \qquad \textit{Eqs. 5-36}$$

Finally, the output stream can be calculated with Eqs. 5-37.

$$\dot{n}_{out} = \frac{\dot{n}_{out,inert}}{y_{inert,off}} \qquad \textit{Eqs. 5-37}$$

The application of Eqs. 5-38 to Eqs. 5-44, leads to the calculation of r_i with $i = O_2$ for the oxygen the uptake rate and $i = CO_2$ for the carbon evolution rate. The accumulation term can be formed according Eqs. 5-39 by means of a difference quotient.

$$\frac{dn_i}{dt} = \dot{n}_{i,in} - \dot{n}_{i,out} + TR_i \qquad \textit{Eqs. 5-38}$$

$$\frac{dn_i}{dt} \overset{\text{def}}{=} \frac{\Delta n_i}{\Delta t} = y_{i,off} \cdot \frac{\Delta n_{Gas}}{\Delta t} \qquad \textit{Eqs. 5-39}$$

$$\dot{n}_{i,in} = y_{i,in} \cdot \dot{n}_{in} \qquad \textit{Eqs. 5-40}$$

$$y_{i,in} = \frac{\dot{V}_{i,in,\,norm} + y_{i,Air} \cdot \dot{V}_{air,in,norm}}{\sum_{j=1}^{n} \dot{V}_{j,in,norm}} \qquad \textit{Eqs. 5-41}$$

$$\dot{n}_{i,in} = y_{i,in} \cdot \dot{n}_{in} \qquad \textit{Eqs. 5-42}$$

$$\dot{n}_{i,out} = y_{i,off} \cdot \dot{n}_{out} \qquad \textit{Eqs. 5-43}$$

$$\Rightarrow r_i = TR_i = \dot{n}_{i,in} - \dot{n}_{i,out} - \frac{\Delta n_i}{\Delta t} \qquad \textit{Eqs. 5-44}$$

A glance on the equations Eqs. 5-23 Eqs. 5-36 and Eqs. 5-44 shows the dependence of the rates on the selected time interval Δt. This influence on the method of finite difference approximation has been studied[28]. As a new approach, the traditional method of finite difference approximation was replaced by a linear regression function. The Matlab function *regress* is performed on a certain amount of rate datapoints, which are adjusted by a moving window with the selected windowsize. [b,bint] = regress(y,X) returns a vector b where the first line represents the gradient corresponding to the current rate.

5.7.1.1 Propagation of uncertainty and noise error

The total absolute error e_{abs} of the rates according [29] corresponds to the sum of the noise error $e_{noise,i}$ and the propagated uncertainty error Δf_{r_i} , which is determined by the measured uncertainty according Eqs. 5-45.

$$e_{abs,i} = e_{noise,i} + \Delta f_{r_i} \qquad \textit{Eqs. 5-45}$$

[b,bint] = regress(y,X) also returns a *p*-by-2 matrix *bint* of 95% confidence intervals for the coefficient estimates. The first column of *bint* contains lower confidence bounds for each of the *p* coefficient estimates. The noise error $e_{noise,i}$ of the rates is computed with this information and always the first datapoint of the moving window was taken for the calculation.

The absolute propagated uncertainty error Δf_{r_S} of the substrate rate r_S as function of the weight signal and feed concentration can be calculated according Eqs. 5-46 to Eqs. 5-48.

$$\frac{\partial f_{rS}}{\partial w_S} = \frac{1}{dt} \cdot \frac{c_{Feed} \cdot c_{mol}}{\rho_{Feed} \cdot M_{Feed}} \qquad \textit{Eqs. 5-46}$$

$$\frac{\partial f_{rS}}{\partial c_{Feed}} = \frac{dw}{dt} \cdot \frac{c_{mol}}{\rho_{Feed} \cdot M_{Feed}} \qquad \textit{Eqs. 5-47}$$

$$\Delta f_{r_S} = \sqrt{\left(\frac{\partial f_{rS}}{\partial w_S}\right)^2 \cdot \left(\Delta e_{w_S}\right)^2 + \left(\frac{\partial f_{rS}}{\partial c_{Feed}}\right)^2 \cdot \left(\Delta e_{c_{Feed}}\right)^2} \qquad \textit{Eqs. 5-48}$$

The absolute propagated uncertainty errorΔf_{r_N} of the nitrogen rate r_N as function of the weight signal and ammonia concentration can be calculated according Eqs. 5-49 to Eqs. 5-51.

$$\frac{\partial f_{rN}}{\partial w_{NH3}} = \frac{1}{dt} \cdot \frac{c_{NH3} \cdot c_{mol} \cdot 14}{17 \cdot \rho_{NH3} \cdot M_{NH3}} \qquad \textit{Eqs. 5-49}$$

$$\frac{\partial f_{rN}}{\partial c_{NH3}} = \frac{dw_{NH3}}{dt} \cdot \frac{14}{17 \cdot M_{NH3}} \qquad \textit{Eqs. 5-50}$$

$$\Delta f_{r_N} = \sqrt{\left(\frac{\partial f_{rN}}{\partial w_{NH3}}\right)^2 \cdot \left(\Delta e_{w_{NH3}}\right)^2 + \left(\frac{\partial f_{rN}}{\partial c_{NH3}}\right)^2 \cdot \left(\Delta e_{c_{NH3}}\right)^2} \qquad \textit{Eqs. 5-51}$$

The absolute propagated uncertainty error $\Delta f_{r_{O2}}$ of the oxygen uptake rate r_{O_2} as function of the gas inlet streams AIRin and O2in, measured $y_{O_{2,}\mathrm{off}}$ and $y_{CO_2,off}$ concentrations in the offgas are calculated according Eqs. 5-52 to Eqs. 5-56.

$$\frac{\partial f_{O_2}}{\partial F_{AIRin}} = -0{,}2096 + \frac{0{,}7804 \cdot 0{,}2096 \cdot y_{O_{2,}off}}{y_{0_2,wet} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)} \qquad \textit{Eqs. 5-52}$$

$$\frac{\partial f}{\partial F_{O2in}} = -1 \qquad \textit{Eqs. 5-53}$$

$$\frac{\partial f_{O_2}}{\partial y_{O_{2,}off}} = \frac{F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot \left(y_{0_2,wet} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right) + 0{,}2096 \cdot F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot y_{O_{2,}off}}{\left(y_{0_2,wet} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right)^2} \qquad \textit{Eqs. 5-54}$$

$$\frac{\partial f_{O_2}}{\partial y_{CO_{2,}off}} = \frac{0{,}2096 \cdot F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot y_{O_{2,}off}}{\left(y_{0_2,wet} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right)^2} \qquad \textit{Eqs. 5-55}$$

$$\Delta f_{r_{O2}} = \sqrt{\left(\frac{\partial f_{O_2}}{\partial F_{AIRin}}\right)^2 \cdot \left(\Delta e_{F_{AIRin}}\right)^2 + \left(\frac{\partial f_{O_2}}{\partial F_{O2in}}\right)^2 \cdot \left(\Delta e_{F_{O2in}}\right)^2 + \left(\frac{\partial f_{O_2}}{\partial y_{O_{2,}off}}\right)^2 \cdot \left(\Delta e_{y_{O_{2,}off}}\right)^2 + \left(\frac{\partial f_{O_2}}{\partial y_{CO_{2,}off}}\right)^2 \cdot \left(\Delta e_{y_{CO_{2,}off}}\right)^2} \qquad \textit{Eqs. 5-56}$$

The absolute propagated uncertainty error $\Delta f_{r_{CO2}}$ of the carbon evolution rate r_{CO_2} as function of the gas inlet stream AIRin, measured $y_{O_{2,}\mathrm{off}}$ and $y_{CO_2,off}$ concentration in the offgas are calculated according Eqs. 5-57 to Eqs. 5-60.

$$\frac{\partial f_{CO_2}}{\partial F_{AIRin}} = -0{,}000407 + \frac{0{,}7804 \cdot 0{,}2096 \cdot y_{CO_{2,}off}}{y_{02_{wet}} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)} \qquad \textit{Eqs. 5-57}$$

$$\frac{\partial f_{CO_2}}{\partial y_{CO2_{out}}} = \frac{F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot \left(y_{02_{wet},off} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right) + 0{,}2096 \cdot F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot y_{CO_{2,}off}}{\left(y_{02_{wet}} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right)^2} \qquad \textit{Eqs. 5-58}$$

$$\frac{\partial f_{CO_2}}{\partial y_{O_{2,}off}} = \frac{0{,}2096 \cdot F_{AIRin} \cdot 0{,}7804 \cdot 0{,}2096 \cdot y_{CO_{2,}off}}{\left(y_{02_{wet},off} - 0{,}2096 \cdot \left(y_{O_{2,}off} + y_{CO_{2,}off}\right)\right)^2} \qquad \textit{Eqs. 5-59}$$

$$\Delta f_{r_{CO2}} = \sqrt{\left(\frac{\partial f_{CO_2}}{\partial F_{AIRin}}\right)^2 \cdot \left(\Delta e_{F_{AIRin}}\right)^2 + \left(\frac{\partial f_{CO2}}{\partial y_{CO_{2,}off}}\right)^2 \cdot \left(\Delta e_{y_{CO_{2,}off}}\right)^2 + \left(\frac{\partial f_{CO2}}{\partial y_{O_{2,}off}}\right)^2 \cdot \left(\Delta e_{y_{O_{2,}off}}\right)^2} \qquad \textit{Eqs. 5-60}$$

5.8 Quantification of Setup

The experiment must be performed within the technical limits of the setup. Therefore, a kinetic model for the simulation of the critical system variables process-time, biomass and reactor volume as function of the specific substrate rate qs with different volumetric mass transfer coefficients is created. Whether or not the experiment with the online biomass estimation setup is ready it is necessary to provide a basic decision tool.

5.8.1 Fed Batch Model

A fed batch model was established according Eqs. 5-61 - Eqs. 5-63 to quantify the change in biomass, substrate concentration and volume of the system.

$$\frac{dX}{dt} = \mu_{max} \cdot \frac{S}{K_S + S} \cdot X \qquad \textit{Eqs. 5-61}$$

$$\frac{dS}{dt} = q_s \cdot X - \frac{\left(\mu_{max} \cdot \frac{S}{K_S + S} \cdot X\right)}{Y_{\frac{X}{S}}} \qquad \textit{Eqs. 5-62}$$

$$\frac{dV}{dt} = \frac{q_s \cdot X}{c_f} \qquad \textit{Eqs. 5-63}$$

5.8.2 Simulation study

The previously in Eqs. 5-4 and Eqs. 5-5 described Fed Batch model is used to quantify the substrate and nitrogen consumption for the biomass production. In addition, the respective volumetric mass coefficient k_La was calculated with the knowledge of r_{O2}, which can be considered to correspond to the OTR (oxygen transfer rate) according Eqs. 5-64. For each q_s phase, the k_La was compared with the k_La max of 121 h^{-1} given from a previous k_La determination experiment.

$$kla = \frac{OTR}{\left(c^* - c_{O_2}\right)} \qquad \textit{Eqs. 5-64}$$

6 Results and Discussion

The chapter six presents the results of the performed *Saccharomyces cerevisiae* cultivation starting with the evaluation of historical data, including a user story to study general application requirements and regulatory requirements. Furthermore, the results of the setup quantification and the simulation study provide an overview of the boundaries and behaviour of the designed cultivation run. The following subsections 6.4.1 to 6.4.3 show and discuss the results of the estimated biomass with calculated and reconciled rates for different model types.

6.1 Historical Data Evaluation

A user story was added to identify the issues and challenges to implement the online biomass estimation soft sensor concept. A metabolic balance analyser tool for estimating the biomass already exists in the process control system Lucullus, but was not applied in the cultivations. It was found out that an internal function has an error and returns wrong results. Furthermore, there is no possibility to fix this problem without updating the library packages, so this instrument is not available for biomass estimations.
Focusing on the challenges for process control strategies, physiological assumptions and technical limits, it was observed that following process parameters were exceeded:

- The oxygen uptake rate, OUR is higher than the oxygen transport rate OUR. That means that the volumetric mass transfer coefficient is too small in the selected setup.
- Usage of pure oxygen instead of air leads to be out of the offgas analyser ranges.
- Secondary metabolites due to insufficient feeding strategy.
- Inadequate control of the dissolved oxygen caused by the back pressure of the offgas analysis.

The acquired knowledge of the user story emphasises the general goal to control the cultivation under defined conditions, such as a suitable feeding profile. This feeding profile depended on the amount of biomass which can be computed with this soft sensor concept. Table 5 lists the significant cultivation parameters with its initial values and desired accuracy. In this context, the existing pH and temperature control is satisfying the demands.

Table 5: significant cultivation parameters

State variable	Priciple	Setpoint	Accuray
dO_2	direct measured	30%	+-5%
pH	direct measured	7	6,8-7.2
T	direct measured	37°C	+-1 °C
q_S	indirect measured	-	±0,03 g/g/h

6.2 Quantification of Setup

By using the Fed batch model and the initial parameter inputs shown in Table 6, the critical system parameters such as critical process time, maximum generated biomass amount and critical broth volume can be calculated. It depends on the specific uptake rate q_s. and the respective system-dependent maximum volumetric mass transport of oxygen shown in Fig. 6.

Table 6 Parameters of Fed batch model

Parameters	Description	Initial Values	Unit
X	initial biomass	23	g
V	initial volume	3	l
S	substrate	0	g
μ_{max}	specific growth rate of X	0,2	h^{-1}
Y_{xs}	yield biomass/substrate	0,35	g/g
c_f	glucose concentration	200	g/l
H	Henry constant oxygen	$1{,}3*10^{-3}$	mol/l/atm
y_{O2}	O_2 fraction of aeration	0,21	-
d_{O2}	dissolved O_2 concentration	35	%

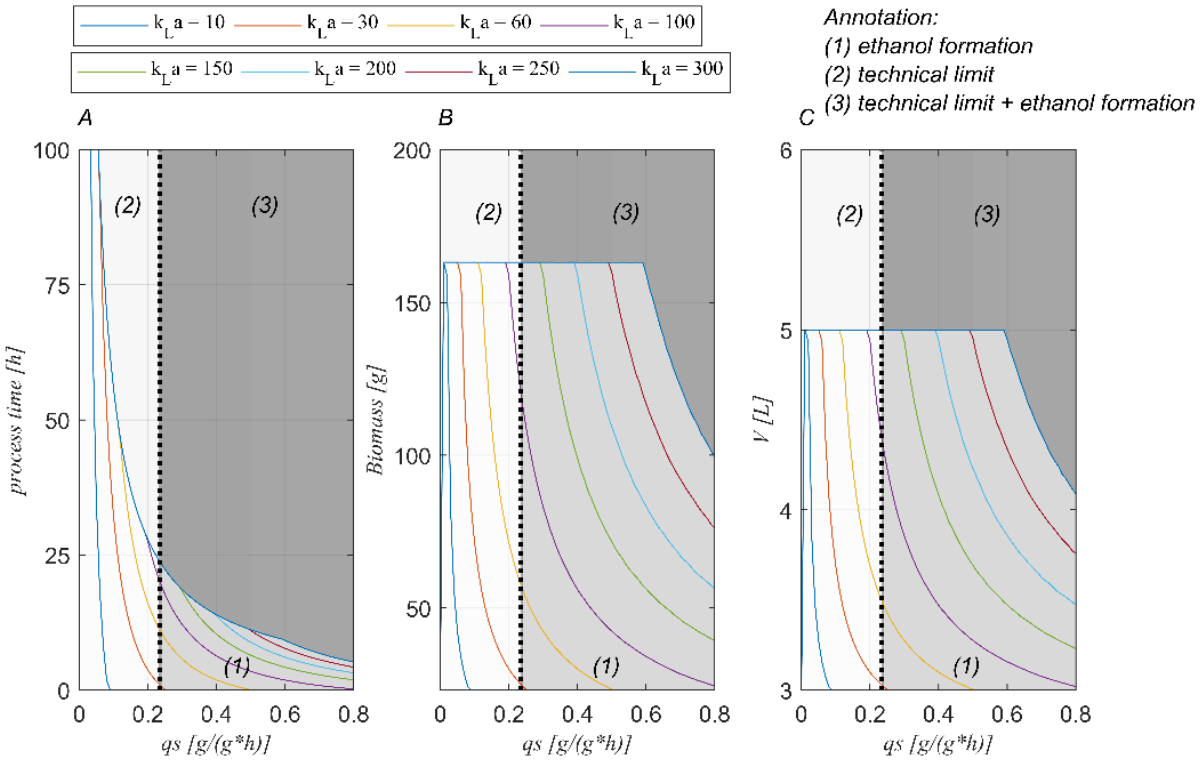

*Fig. 6: **A:** The critical process time, **B:** critical biomass and **C:** critical volume for different volumetric mass transfer coefficients k_La dependent of q_s.*

As described in 5.6 there are limitations for the utilization of the created equation system model. Assumed physiological behaviours and technical limitations should not be exceeded for a successful application of the soft sensor. In this context, it should be noted that the cultivation run should not reach areas (1) - (3). Area (1) shows the area with the ethanol formation. It should be to be assumed, started with a specific rate of q_s = 0,233 g/(g·h) according [30]. The technical limit (2) is caused by oxygen transfer limitations described by the volumetric mass transfer coefficient k_La. In area (3), ethanol is formed and the technical design space exceeded.

The obtained maximum volumetric mass transfer coefficient kla_{design} 121 h^{-1} from a previous k_La determination, including a safety of 0,8* $k_La_{design} \approx 100\ h^{-1}$ is the basis of designing the experiment. The experiment is designed with a specific uptake rate of q_s =0,1 and 0,2 g/(g·h), which satisfies the assumptions and is below the technical limits. Moreover, a q_s =0,3 g/(g·h) it is chosen to interpret the behaviour if the thresholds were exceeded.

It can be observed that a cultivation with a constant specific rate of q_s = 0.1 g/(g·h) is only limited by the maximum working volume of 5l and is able to run approximately 70 h. Furthermore, a cultivation with a constant specific rate q_s = 0,2 g/(g·h) is limited in the oxygen transfer, which can be interpreted by the violet line (k_La =100 h^{-1}) as a threshold. In this case the maximum grown biomass is approximately 140 g and the runtime which is limited up to 28 hours.

The last part of the experiment at a specific uptake rate q_s = 0.3 g/(g·h) is initially located in the region that represents ethanol formation and which is also limited by the oxygen transfer. Under these conditions, the cultivation is complete after 12 hours and a biomass of 75 g. The setup was constructed according to Fig. 9.

Fig. 9: Setup of Fermenter 1

6.3 Simulation study

The setup quantification information shows the maximum amount of biomass that can be achieved using the selected substrate uptake rates q_s and gives an indication of the maximum possible run time. The distribution of the different chosen q_s as 0,1 0,2 0,3 in g/(g·h) leads to the overall simulated experiment shown in Fig. 10. The estimated biomass, consumed substrate glucose and nitrogen is shown in subplot A. Subplot B shows the simulated k_La and the increasing broth volume as a function of time associated with the threshold values of these parameters.

In this context the performance of such cultivations also includes shaping the length of time periods for each q_s in a manner to optimally quantify the biomass increase alternatively with the measurement of optical density and dry weight. Consequently, the sampling interval of the low specific q_s rate q_s = 0.1 g/(g·h) and an additional rate, q_s = 0,05 g/ g/(g·h), is lower and these experimental periods are also run overnight.

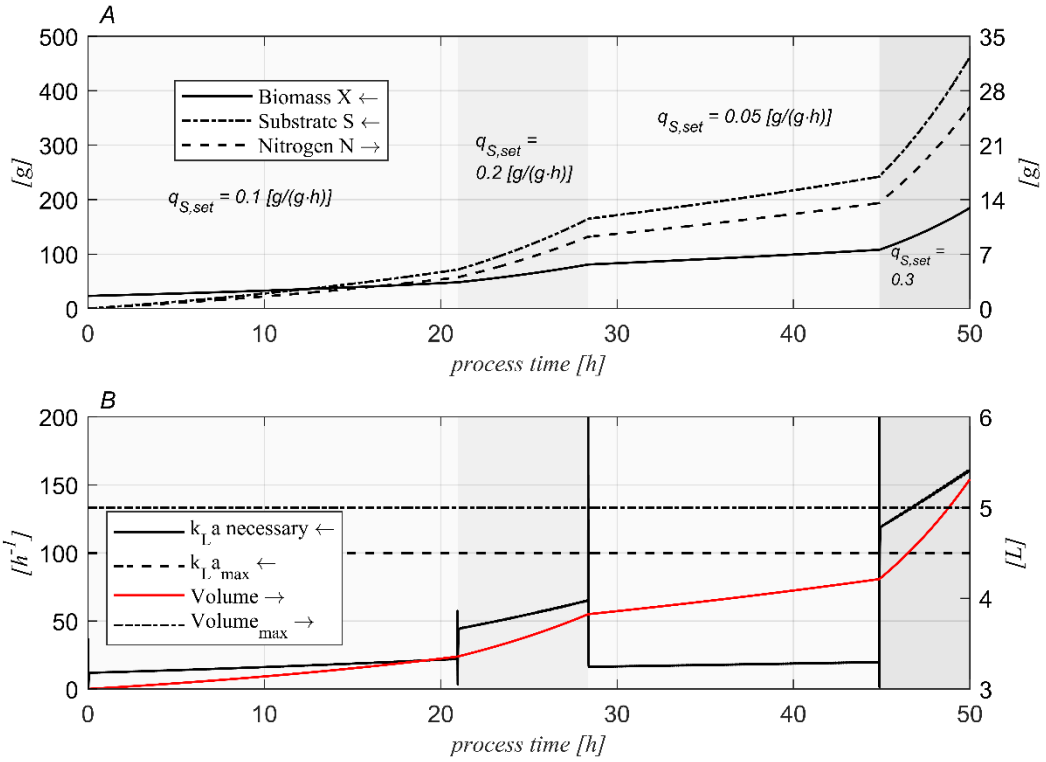

Fig. 10: A: Simulation of consuming substrate(glucose), nitrogen and produced biomass in g. B: Trends of k_La and broth volume increase various q_s periods as a function of time. The threshold for k_La max is 100 h^{-1} and for the maximum working volume 5 l.

The simulation of the required volumatric mass coefficient of oxygen and the increase in the broth volume shows that the thresholds are exceeded in the period q_s = 0,3 g/(g*h). This period should show the behaviour of the cultivation outside the technical limits and the additional ethanol formation.

6.4 Process data results

Following subchapters shows the results of two different model types starting with the equation system with 3 balances (C, DoR, N) in subchapter K3S1 and second model with 2 balances (C, DoR) in subchapter K2S1. An additional run shown in subchapter K2S1_2 was performed with a feed concentration of 400 g/l instead of 200 g/l to investigate the performance of higher q_s rates.

The evaluation of the results for the various equation models runs K3S1, K2S1, K2S1_2 followed the workflow starting with checking the consistency of the data using the statistical X^2 test. Subsequently, the results of each input rate were displayed in contrast to the adjusted rate. In addition, the time - dependent relative errors of the calculated rates are shown. As a next investigating step, the calculated biomass from each mass balance including the estimated biomass with calculated and reconciled rates were illustrated. Moreover, the discrepancies of the estimated biomass with calculated and reconciled rates were shown and investigated. A comparison of the estimated biomass with calculated and reconciled rates in contrast to the dry weight (DW) and optical density (OD) measurements completed the examination of the used soft sensor concept. In this context, it should be noted that the online microscope was not available as an online tool for determining the biomass due to the increased cell density.

6.4.1 K3S1

Several statistical tests are available in literature to investigate the quality of the results [31], [32] and [33]. In this thesis, the consistency check for the K3S1 model was performed using the statistical global X^2 test, which was also proposed in [34] [35]. The result is shown in Fig. 11. Subplot A represents the calculated consistency index value h with a bar interval of 1 hour against the levels of significance α = (0,1 0,05 0,025 0,01) with the degree of freedom of 2 equal to the rank of the redundancy matrix R. It can be observed that the consistency index value h is exceeding the thresholds, respectively all levels of significance. This means a significant error has been detected. The classification of the errors can be done according [21] as noted:

I. At any rate, on one of the primary measurements exists a significant (gross) error.
II. The equations system is incorrect.
    a. a significant component is not considered in the equation system or
    b. a component has a different composition from the specified one.
III. The variances of the rates are too small resulting in a too sensitive X^2 test.

Identifying the gross error measurements when the original complete data set does not pass the hypothesis test can be determined by deleting a measurement and calculating the consistency test again[24]. If the model does not describe the bioreaction, it is possible to remove an equation from the model and recalculate the consistency test to identify the error containing equation.

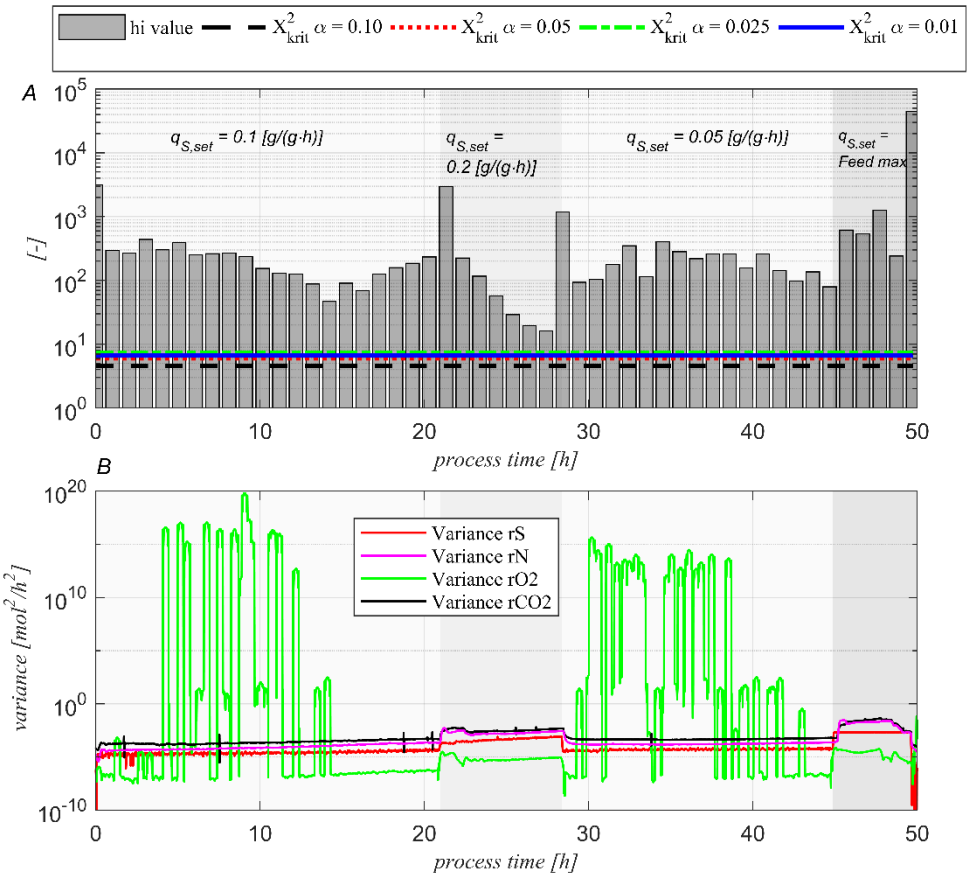

Fig. 11: K3S1 ***A:*** *Statistical Chi-quadrat test in 60 min intervals for model validation. The straight lines represent the levels of significance for the statistical test.* ***B:*** *Time-dependent variances of the K3S1 model input rates.*

Subplot B shows the variances of each rate as a function of time and set q_s rates. In this context, the time course result of the r_N variance is significantly manipulated by Savitzky-Golay filtering to prepare the outcome in a vivid depiction. The variance values higher than 1 $(mol/h)^2$ were caused by the discontinuous working nitrogen pump resulting in a low to zero nitrogen rate leading to a relative error that goes towards the infinite and culminate to a very high variance. In contrast to that, the other variances of the rates were very low all the time. Moreover, the r_N variance was lowest when the nitrogen feed was on.

Furthermore, it can be observed that in the phase qs_{set} = 0.2 g/(g·h), the h value decreased exponentially. Decreasing in the direction of threshold values and the consistency value increased again with a smaller qs_{set} = 0.05 g/(g·h) and got about 2 magnitudes higher. This suggests that there can be an error in the model, or the variances of the rates are too small resulting in a too sensitive X^2 test assuming there is no gross error on the calculated rates. Investigating the rates, including examining each balance gives more information about this behaviour and was studied subsequently.

Nevertheless, the results for the K3S1 were shown without passing the global statistical X^2 test to demonstrate and observe the behaviour of estimating the biomass via measured and reconciled rates.

The calculated substrate rate of glucose (red line), the confidence intervals with a level of significance of (α= 0.05) and the reconciled rate in c-mol are shown in Fig. 12. It could be observed that the calculated and the reconciled rate is similar in the low q_s phases (q_s = 0,1; 0,05 g/(g·h)). The q_s = 0,2 g/(g·h) phase represented a distinction of the reconciled rate and the calculated rate at the

beginning of the phase. Moreover, they are completely different in the phase of q_s = Feed max. In this context, it must be stated that the planed q_s = 0,3 g/(g·h) could not be performed the whole time in this phase caused by the upper pump rate limitation. A supplementary experiment of higher q_s rates was realized in subchapter 6.4.3.

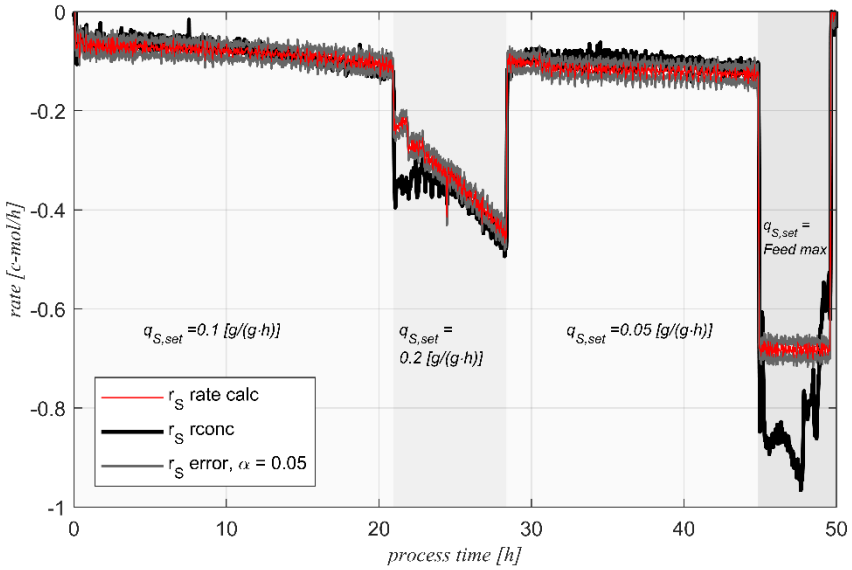

Fig. 12: K3S1 substrate consumption rate r_S. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_S. The reconciled rate r_S is shown as black trend.

Fig. 13 shows the relativ error in the substrate rate as the sum of the propagation and noise error. The window size of the linear regression was set to 10 which corresponds to the accounted data points. These 10 data points represent a time-interval of 5 minutes, accorrding to a signal receiving interval of 30 seconds from the Python Server. It could be observed that the relative propagation error and noise error are significantly higher in the low q_s phases caused by the poor signal to noise ratio (SNR) which was also observed in [28]. Moreover, the noise error has less impact of the time-dependent error.

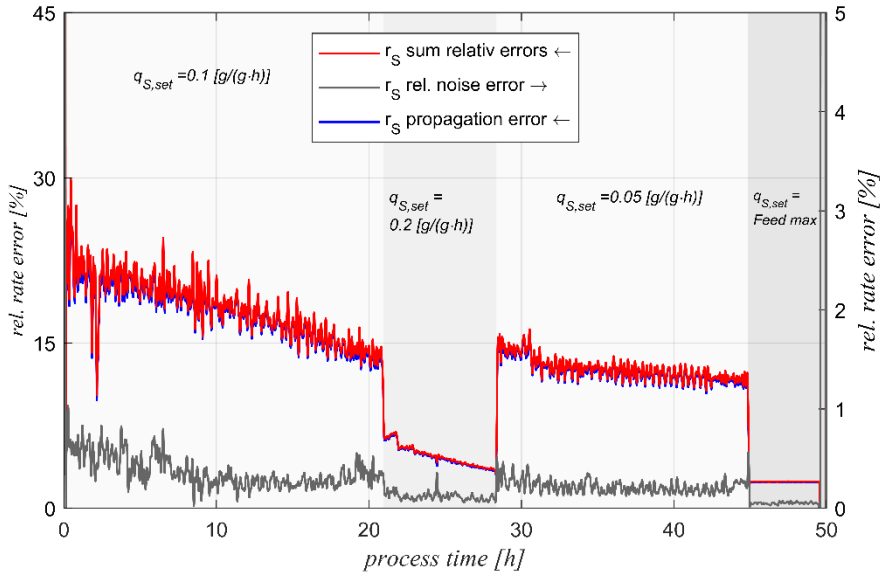

Fig. 13: K3S1 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of substrate consumption rate r_S.

The results of the calculated nitrogen input rate r_N, including the 95% confidence intervals and the reconciled rate are shown in in Fig. 14. It is interesting that the reconciled rate was completely equal to the calculated rate, especially at the beginning, in phase q_s =0,2 g/(g·h) and in the phase q_s = Feed max in contrast to the results of the substrate rate before.

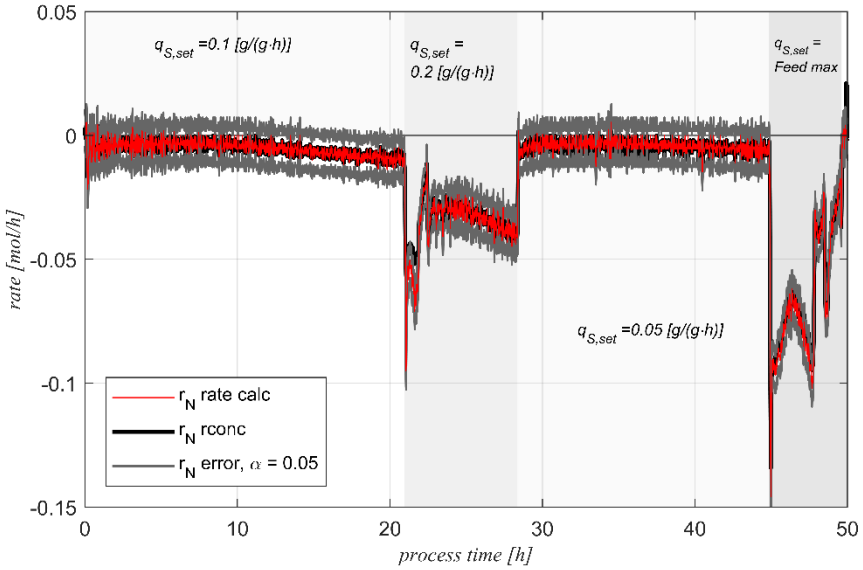

Fig. 14: K3S1 nitrogen consumption rate r_N. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_N. The reconciled rate r_N is shown as black trend.

The results of the relative error by means of the scale calculated r_N rate (Fig. 15) shows throughout different trends in contrast to the relative error of the r_S error shown in Fig. 13. The propagation error has a significantly higher impact of the total relative error in contrast to the error results of r_S. Furthermore, there were outliers caused by extremely low to zero rate calculations, especially in the phases with low q_s rates.

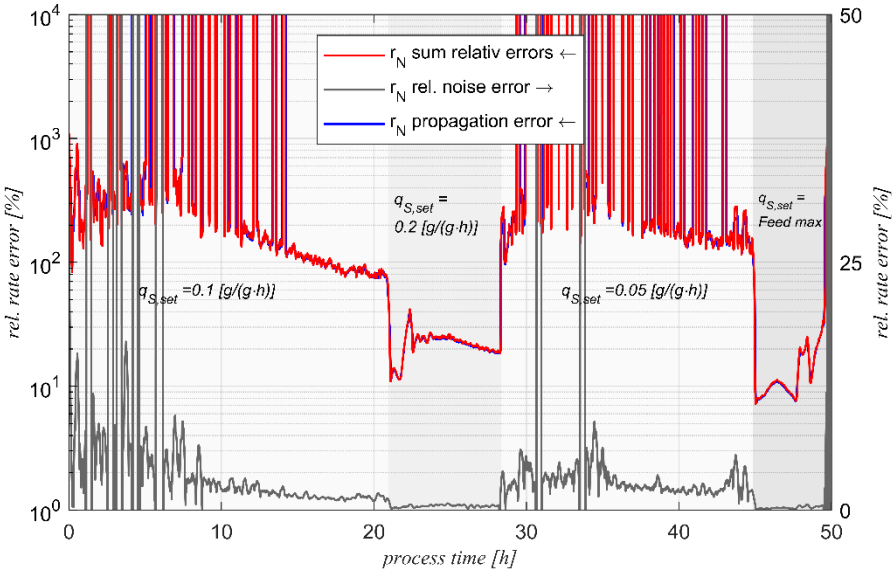

Fig. 15: K3S1 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the nitrogen consumption rate r_N.

The calculated gas rates, including the 95% confidence intervals are shown as following, beginning with the results of the oxygen uptake rate r_{O2} in mol/h as a function of time. It must be stated that the O_2 mass flow controller is not considered in K3S1 which otherwise leads to higher confidence intervals. It could be observed that the calculated and reconciled rate were similar in low q_s phases and there were mismatches at the beginning of the phase q_s = 0,2 g/(g·h) and in the phase q_s = Feed max phase. The reconciled r_{O2} rate is underrated in contrast to the calculated rate in these parts.

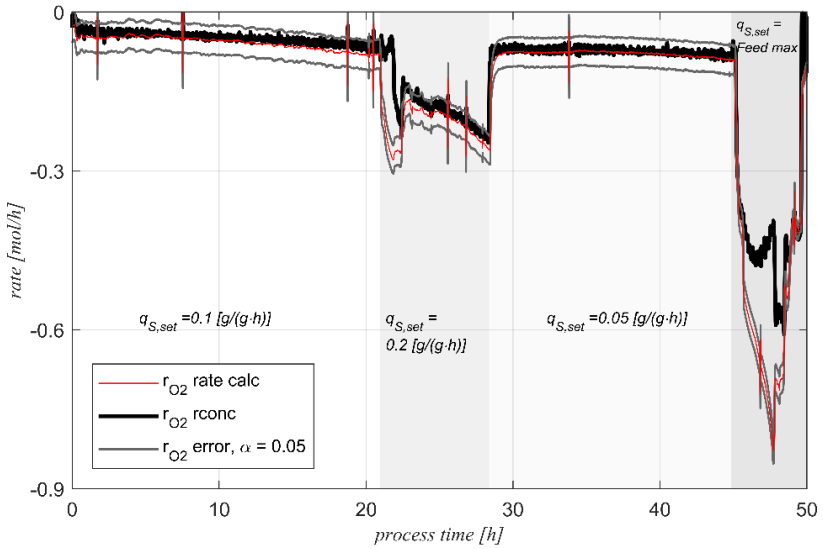

Fig. 18: K3S1 oxygen uptake rate r_{O2}. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_{O2}. The reconciled rate r_{O2} is shown as black trend.

The outcome of the relative r_{O2} rate error is represented in Fig. 19. It could be detected that the relative error on r_{O2} is mainly dependent on the error propagation results. The noise error was at least approximately three magnitudes smaller than the propagated error. The outliers in the noise error and propagated error results were caused by the sampling procedure during the run. It could be identified that the calculated relative error on r_{O2} was higher in contrast to r_S and r_N calculated with scale signals.

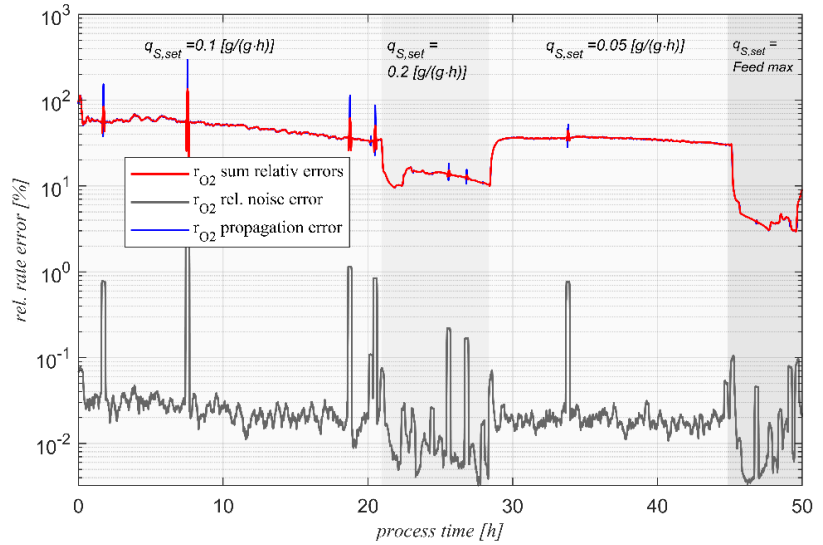

Fig. 19: K3S1 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the oxygen uptake rate r_{O2}.

The behaviour of the carbon evolution rate r_{CO2} in mol/h is revealed in Fig. 20. As similar to the r_{O2} rate the calculated and reconciled rate were just the same in the low q_s phases. Moreover, the reconciled rate was also underrated at the beginning of phase q_s = 0,2 g/(g·h) and the phase q_s = Feed max.

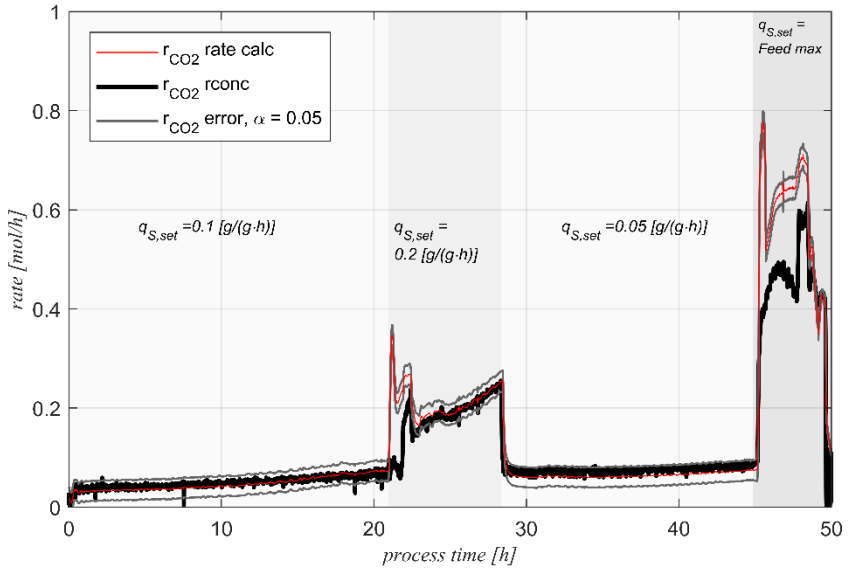

Fig. 20: K3S1 carbon dioxide evolution rate r_{CO2}. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_{CO2}. The reconciled rate r_{CO2} is shown as black trend.

The trend of the relative error on r_{CO2} during the trial is shown in Fig. 21. The error characteristic was similar to the r_{O2} error. The outliers in the noise signal was caused by the sample procedure. The time-dependent relative error was mainly driven by error propagation. As seen the noise error was at least 4 magnitudes smaller than the propagated error.

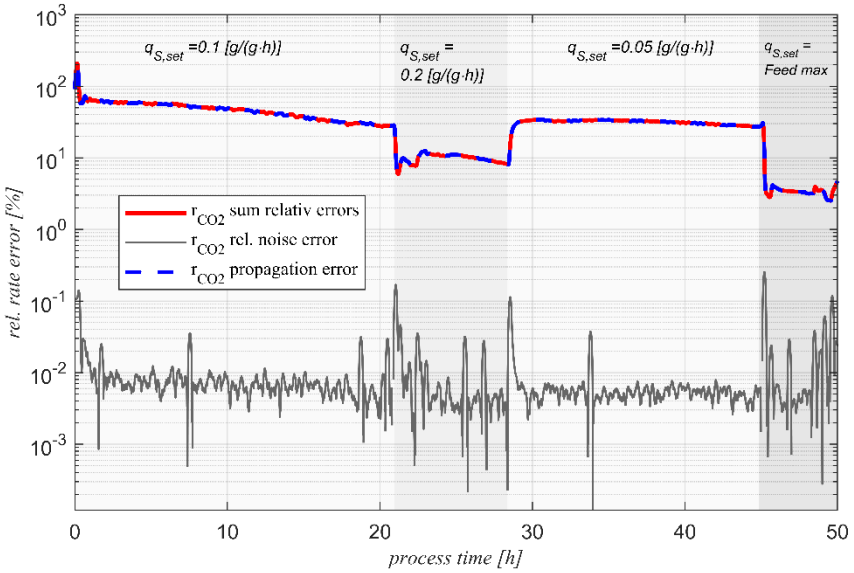

Fig. 21: K3S1 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the carbon dioxide uptake rate r_{CO2}.

The results of the dissolved O_2 concentration control strategy are shown in Fig. 22. The setpoint of 35% and the control of it is given in subplot A for all phases. The actuating variables agitator speed and the air flow are revealed in subplot B. Most of the time the varying agitator speed was sufficient to control the O_2 concentration setpoint. The change of the specific uptake rate q_s led to the singular overshoot at the point of 22 hours. Moreover, a good control behaviour is observed when q_s is decreased to 0,05 g/(g·h).

A different manner was detected in the q_s rate of Feed max. Here the O_2 concentration was not only reached by manipulating the stirrer speed. In this phase the air flow regulation entered the control loop and strong overshoots could be discovered. This behaviour could be primarily caused by the high q_s change from q_s = 0,05 g/(g·h) to the maximum feed pump rate and also by switching the metabolic mode depending of the glucose levels[36] [37].

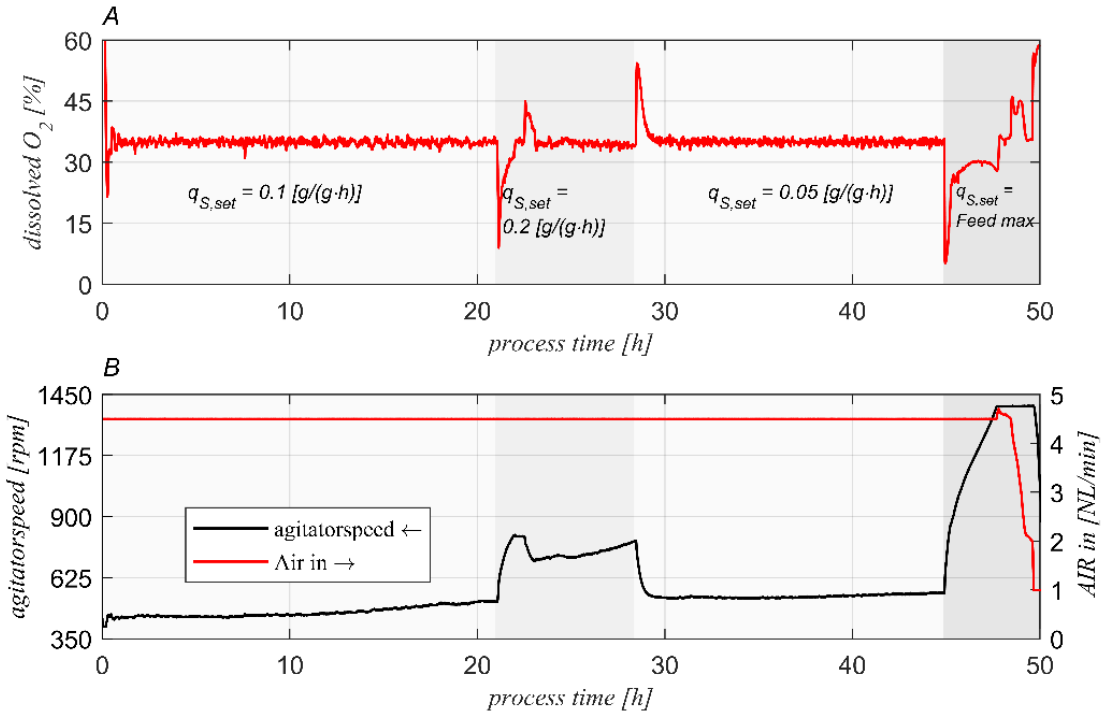

Fig. 22: ***A:*** *closed loop controlled dissolved O_2 on 35%.* ***B:*** *Subplot shows the actuating variables agitator speed and aeration rate of the cascade control system.*

To examine the results more closely, each mass balance is shown in Fig. 23. Subplot A shows the calculated biomass, which was calculated independently from each mass balance, including the estimated biomass result, which is displayed as a red line.

In the q_s =0,1 g/(g·h) phase, the calculated biomass amount of each balance was reasonably similar and distinguish with the beginning of the q_s = 0,2 g/(g·h) phase. The DoR balance predicted the biomass amount result best in contrast to the C-balance and N-balance determined biomass quantity. An interesting aspect accounted the fact that the estimated biomass amount is nearly similar to the determined biomass, according the DoR balance. So it could be justified by a low error of this balance.

The relative and cumulative absolute error with the estimated biomass rate was examined on the different mass balances shown in subplot B-D. The relative error was related to the overall conversion of each balance according Eqs. 6-1 to Eqs. 6-3.

$$e_{c,rel(t)} = 100 \cdot \frac{e_{c,abs(t)}}{1 \cdot r_{X(t)} + 1 \cdot | r_{S(t)} | + 1 \cdot r_{CO2(t)}} \qquad \textit{Eqs. 6-1}$$

$$e_{DoR,rel(t)} = 100 \cdot \frac{e_{DoR,abs(t)}}{4{,}159 \cdot r_{X(t)} + 4 \cdot | r_{S(t)} | + 4 \cdot | r_{O2(t)} |} \qquad \textit{Eqs. 6-2}$$

$$e_{N,rel(t)} = 100 \cdot \frac{e_{N,abs(t)}}{0{,}176 \cdot r_{X(t)} + 1 \cdot | r_{N(t)} |} \qquad \textit{Eqs. 6-3}$$

Subplot B shows the absolute and relative error of the C-balance as a function of time. After solving the equation system according to Eqs. 5-8, it could be observed that the biomass was underestimated with regards to the C-balance and led to a cumulated error by 0,4 mol of biomass.

Investigating the cumulative and relative error of the DoR-balance as a function of time, it was found out that the relative error in terms of the C and N-balance was lowest. This low error is consistent to the estimated biomass, which was near to the calculated biomass from the DoR-balance revealed in subplot A. It could be seen in subplot C the cumulative error was increased to approximately 0,13 mol. This can be also interpreted that the biomass is overestimated.

The relative and cumulative error of the N-balance for the chosen time interval is shown in subplot C. A look on the relative error a much higher error on this balance in contrast to the C and DoR balance was identified. In addition, the cumulative error was almost minus 0,3 moles, which can be interpreted that 0,3 moles of nitrogen were not available for the estimated biomass, since it can be assumed that all the nitrogen input is used to produce biomass. In conclusion, these results of the N balance made this balance to be a candidate of a systematic balance error, which could cause the overall not passing consistency test and was further discussed after the same procedure with reconciled rates.

Nevertheless, the estimated biomass calculated according Eqs. 5-8 for the K3S1 model (C, DoR, N-balance) with glucose as substrate (S1) showed that the DoR-balance is describing the mass transfer of the overall bioreaction most suitable.

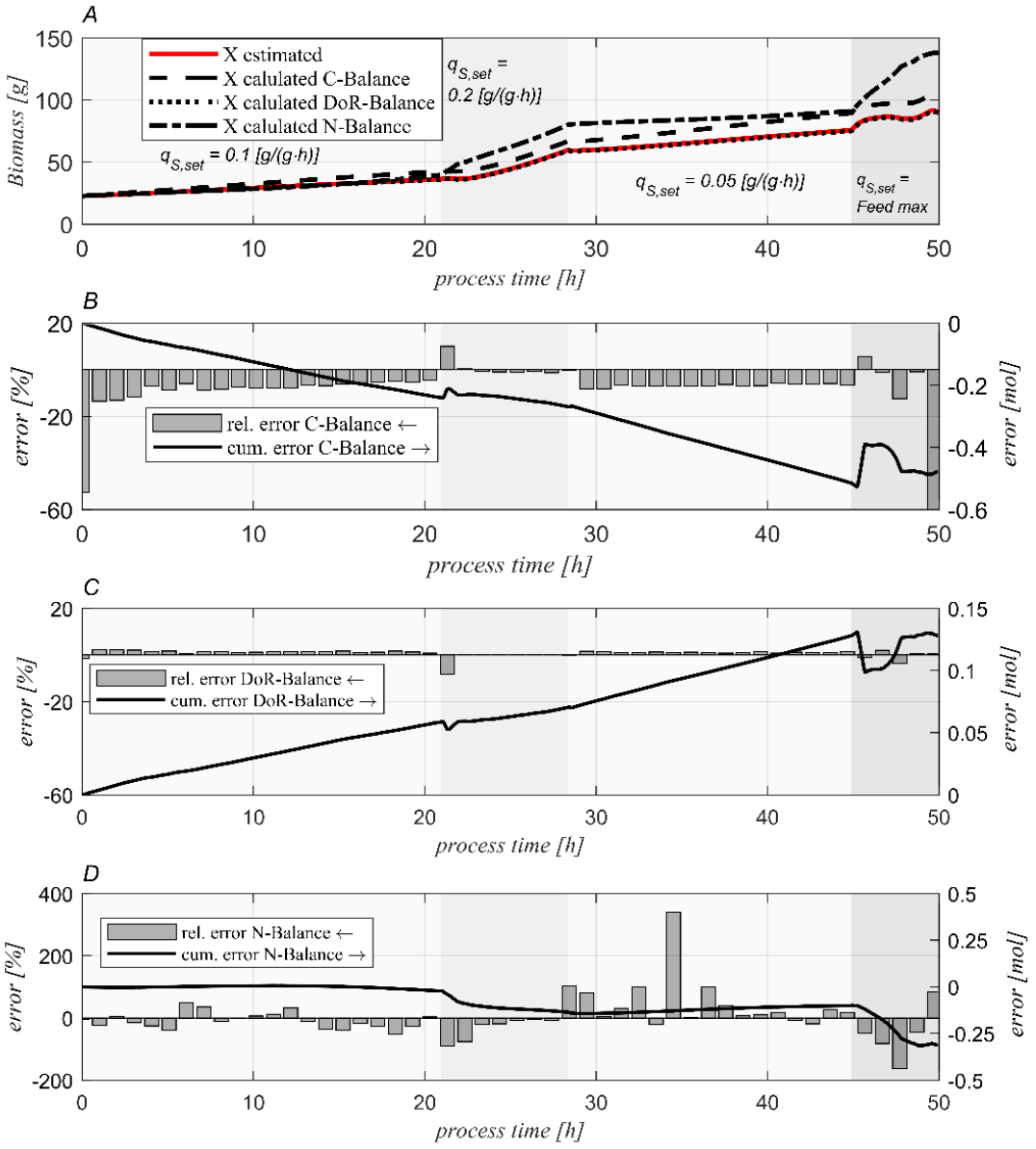

Fig. 23: K3S1 ***A:*** *Calculated biomass from C, DoR and N balance with additional estimated biomass result as a function of the time.* ***B:*** *C-balance cumulative error in mol. The relative error relates to the total carbon across the balance space.* ***C:*** *DoR-balance cumulative error in mol. The relative error relates to the total electron's actions across the balance space.* ***D:*** *N-balance cumulative error in mol. The relative error of N-balance relates to the total nitrogen across the balance space.*

The results shown in Fig. 23. revealed discrepancies in the balances, especially in the N and C balance in general. As a next step, it is possible to alter the measured rates having regard to their accuracies quantified by their standard deviations of the rates according Eqs. 5-18.

The biomass estimation results with the reconciled rates are shown in the subplots A-D in Fig. 24. The estimated biomass with the reconciled rates represents the blue line in subplot A. In contrast to the results in Fig. 23, the estimated biomass was well predicted from the N-balance. Subplot B shows each balance residue. The residue of the nitrogen balance was here also significantly higher than the residues of the C- and DoR-balance and was contributing the conclusion that this nitrogen balance does not characterize the bioreaction in terms of nitrogen with this discontinuous r_N feeding profile.

The relative errors were performed according Eqs. 6-4 to Eqs. 6-6, to investigate the shown cumulative errors in subplot B in another context. The cumulative errors in the balances were related to the cumulative sum of the absolute input rates plus absolute output rates and estimated biomass generation as a function of time for each balance.

$$e_{c_{rconc},rel(t)} = \frac{\int_0^t e_{c_{rconc},abs(t)}}{\int_0^t r_{x(t)} + | r_{S(t)} | + r_{CO2(t)}} \qquad \textit{Eqs. 6-4}$$

$$e_{DoR_{rconc},rel(t)} = \frac{\int_0^t e_{DoR_{rconc},abs(t)}}{\int_0^t 4{,}1590 \cdot r_{x(t)} + 4 \cdot | r_{S(t)} | + 4 \cdot | r_{O2(t)} |} \qquad \textit{Eqs. 6-5}$$

$$e_{N_{rconc},rel(t)} = \frac{\int_0^t e_{N_{rconc},abs(t)}}{\int_0^t 0{,}176 \cdot r_{x(t)} + 1 \cdot | r_{N(t)} |} \qquad \textit{Eqs. 6-6}$$

Subplot C shows the outcomes of this interpretation. In this view, it could be observed that the nitrogen balance fits most suitable related to the cumulative sum of the nitrogen input and nitrogen amount in the generated estimated biomass.

The results show that the N-balance after the reconciliation of the calculated rates is more significant for the biomass estimation. This fact can be explained by the statistical basis of data reconciliation. It arises from the properties that are assumed for the random errors on the rates. It is assumed that the random errors follow a normal distribution with zero mean and a known variance covariance matrix F generated according Eqs. 5-15. By convention, the matrix F is diagonal respectively the rates are independent from each other, the variances on each rate can be interpreted as a weighting factor. That means, a higher value of the variance implies that the calculated current rate is less accurate, so the more accurate rates have larger weights. Moreover, the variances in a balance can be simply summed up due to this independence and give a suitable information of the weighting of each balance according its weighting for the estimation.

Returning to Fig. 11, the variance of r_N was lowest as opposed to the other rates when the NH_3 feed pump operates. In the N balance, only the r_N variance occurs and the overall sum of the variances in this balance is lower than all other balances, so it is evident that the estimated biomass is predicted to the N-balance after the reconciliation procedure.

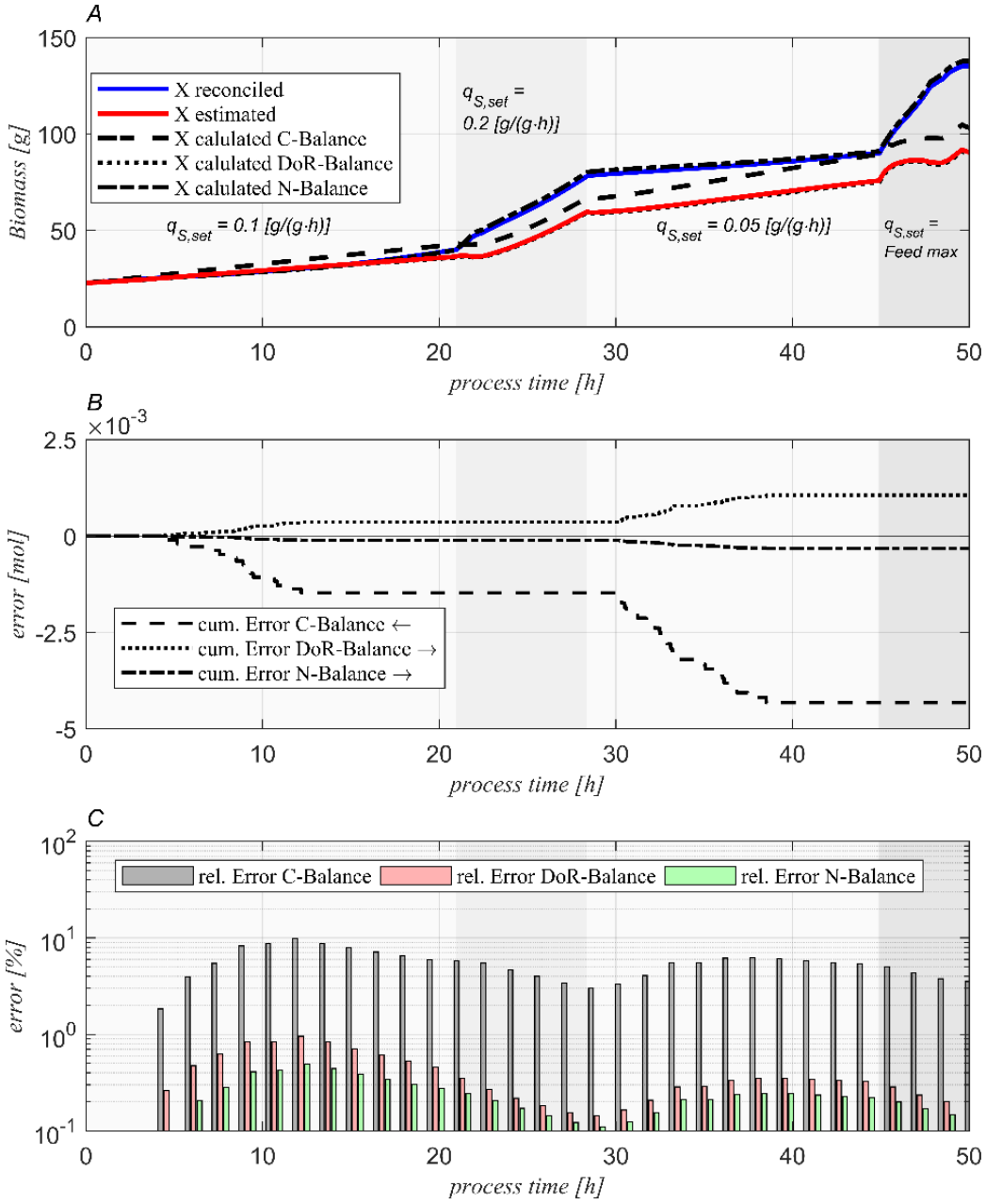

Fig. 24: K3S1 ***A:*** *Calculated biomass from C, DoR and N balance with additional estimated and reconciled biomass results as a function of time.* ***B:*** *Cumulative error of C, DoR and N balance after reconciliation.* ***C:*** *Relative balance errors related to total C, electron actions (DoR) and N usage across the balance space.*

As described in subchapter *Materials and Methods*, alternative biomass determinations with optical density measurements and dry weight measurements were done to compare the results of these common procedures with the estimated biomass on calculated and reconciled rates. Fig. 27 shows in subplot A the estimated biomasses with the calculated rates (green) and the reconciled rates (blue), including alternative offline biomass determinations with the confidence intervals of 95%.

Subplot B reveals the absolute error of the estimated biomasses in comparison to the optical density measurements at the respective time points. Subplot C shows the absolute error of the estimated biomasses in comparison to the dry weight results at the respective time points. The absolute mean errors of the biomasses with respect to the OD respectively DW measurements are also shown in subplot B and C.

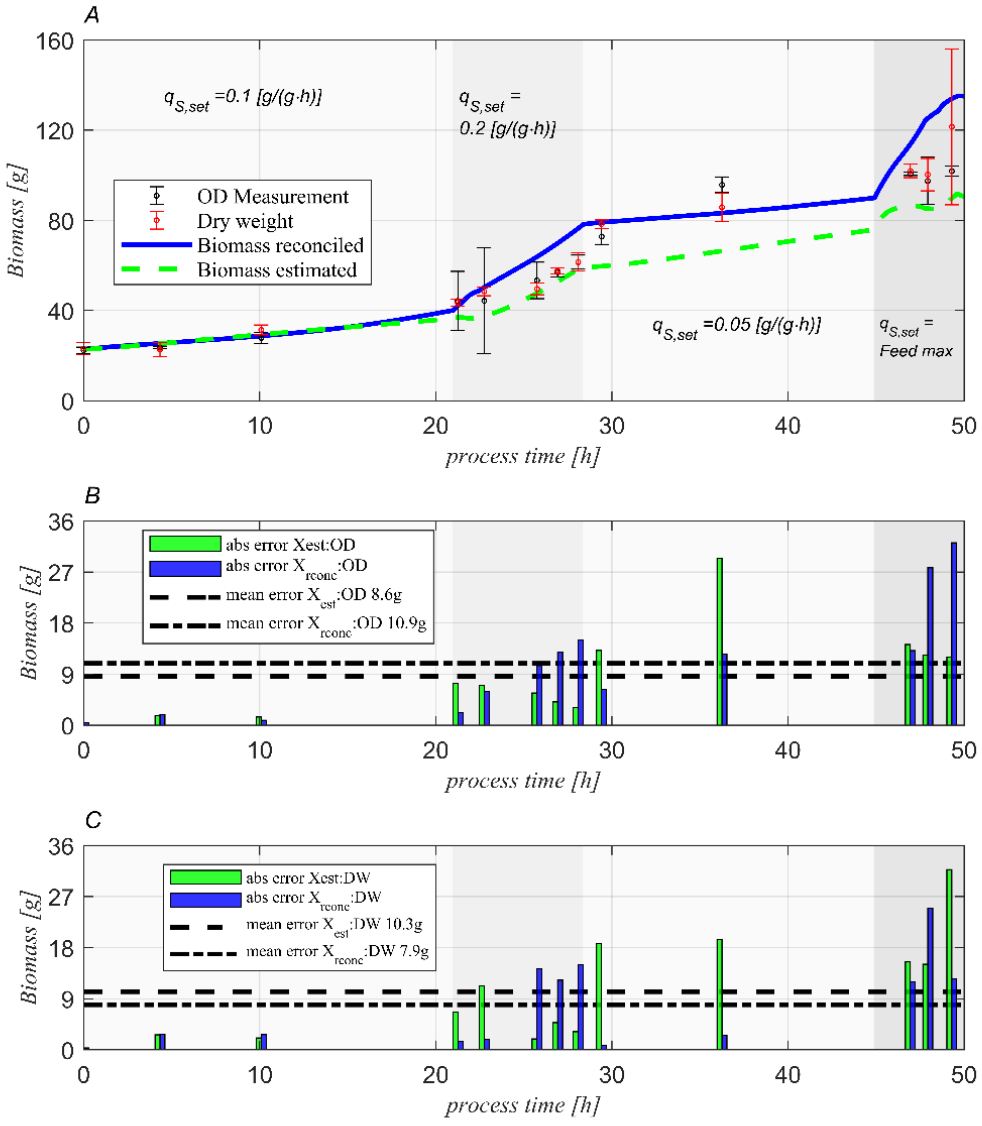

Fig. 27: K3S1 Independent biomass content comparison and validation, 13 samples were under consideration. ***A:*** *Estimated and reconciled biomass trend as functions of time in different q_s phases, including optical density (OD) and dry weight (DW) measurements.* ***B:*** *Absolute errors of estimated and reconciled biomass related to optical density (OD) measurement method. In addition, the mean error over all different q_s process phases.* ***C:*** *Absolute errors of estimated and reconciled biomass related to dry weight (DW) measurement method. In addition, mean error over all different q_s process phases.*

A view on the subplot A shows that the estimated biomass with the reconciled rates was overrated and the estimated biomass by the calculated rates was underrated most of the time. This phenomenon takes place, especially at higher q_s rates (q_s = 0,2 g/(g·h)).

It could be observed that the mean error of the estimated biomasses (*Xrconc, Xest*) with and without reconciling data in terms of the optical density measurements yielded in 8,6 or rather 10,9 g. This fact shows that, the data reconciliation does not generally improve the performance of the estimation. That is, if the equation model does not pass the consistency test, it is possible that the biomass estimate is not optimal. The mean error on the estimated biomasses related to the dry weight biomass determination yielded in 10,3 respectively 7,9 g. Reconciling the rates in this case improves the biomass estimation compared to the OD measurements in contrast to the DW measurements. Back to subplot A, the best estimation is between the dotted green line and the blue curve. Moreover, the results of the offline methods for biomass determination are also afflicted with errors concerning washing procedure and error propagation on DW and calibration, dynamic range and attachment of cells or antifoam on OD measurements[38].

It could be stated that best accordance was given by the estimated biomass with reconciled rates in comparison to the dry weight measurement, which led to an error of approximately 7%, related to the generated biomass.

Back to the initial situation, it was observed that the equation model did not meet the global consistency criteria explained before. Using small q_s rates (q_s = 0,1 / 0,05 g/(g·h)), small variances led to a too sensitive consistency test. This also confirmed by the observation of an exponential decrease of the consistency value h in the q_s = 0,2 g/(g·h) phase.

Nevertheless, the model does not fulfil the quantification of the global consistency criteria. This suggests that some constraints of the equation model are incorrect. It has been observed that the relative error in the N-balance was significantly higher in the intervals seen in subplot D in Fig. 23 in comparison to the C- and DoR-balance relative error after the estimation without reconciled rates. The glucose and NH_3 amounts were measured from the samples and are shown Fig. 30 in subplot A.

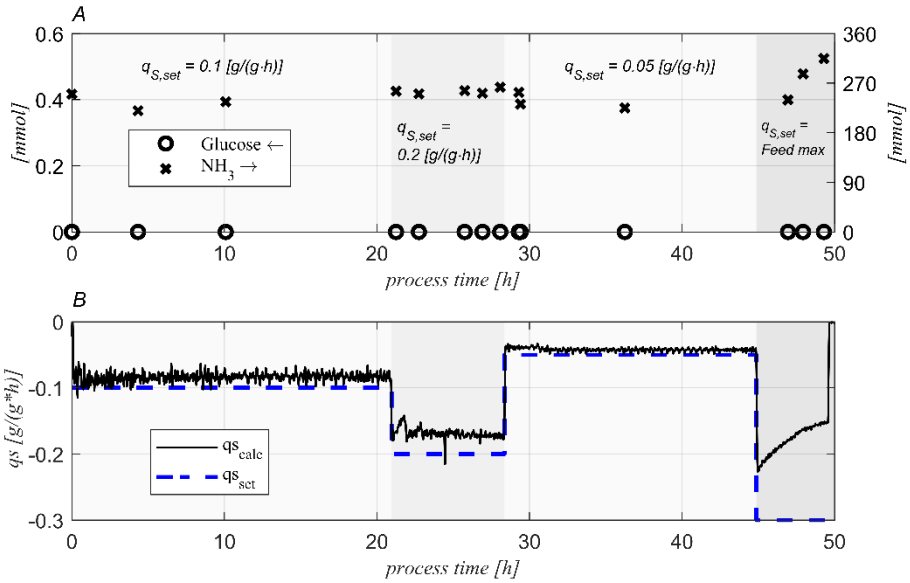

Fig. 30: K3S1 A: Glucose and and NH_3 amount in mmol during the fed batch run in different q_s phases. Instead of Glucose, the ammonium content during the process does not meet the assumption to be zero. B: Pre-defined q_s setpoints and calculated specific rate q_s as a function time.

The assumption that all glucose is metabolised is fulfilled. There is no glucose accumulation at the overall run. Another assumption which concerns that all nitrogen input is used for the biomass generation by holding the pH constant cannot be accepted after the changing NH_3 concentration demonstrated in subplot A. The change is significantly high at the end of the run in the phase q_s = Feed max.

Subplot B in Fig. 30 shows the real calculated specific uptake rate q_s in comparison to the q_s setpoint. The real calculated q_s was similar to the q_s setpoint at low q_s setpoint rates. In the part of the q_s = 0,2 g/(g·h) phase, the real calculated q_s differed from the q_s setpoint. This could be interpreted that the integral term of the PI -Controller must be higher for further runs. The technical limit of the pump in the q_s max phase was reached, so the q_s of 0,3 g/(g·h) was not performed.

In conclusion, the estimation of the biomass with calculated and reconciled rates. It can be summarized that the significance of the balances on the biomass results differed after the reconciliation procedure. If the N-balance would satisfy the assumptions, it would be the most significant balance after the reconciliation procedure in this experimental case. The valuation of each balance in relation to the OD and DW measurements are assessed by their weighting on the estimated biomass results and classified in calculated and reconciled rates input according Table 7.

Table 7: K3S1 Performance of each balance with calculated and reconciled rates input

Balances	Rank with calculated rates	Rank with reconciled rates
N-Balance	less satisfying	best
Degree of Reduction Balance	best	satisfactory
C-Balance	satisfactory	less satisfying

6.4.2 K2S1

A suitable approach is to investigate the run without the used nitrogen balance. The model decreases to two equations (C- and DoR-balance) and one substrate (Glucose). K2S1 is also an over-determined system, where only one (biomass) rate is unknown. Therefore, the reconciliation of the rates is also possible. As a first consequence, the consistency test should be improved in comparison to the K3S1 model. The analysis of the K2S1 model is equally structured such as the previous investigation.

The statistical global X^2-test for the K2S1 equation system is shown in Fig. 31. Firstly, it could be observed that the consistency index h is one magnitude smaller in comparison to the consistency test of the K3S1 model in Fig. 11. Furthermore, the model passed the test at a specific uptake rate of q_s = 0,2 g/(g·h) as an opposite to the K3S1 model. This means, that the K2S1 model is trustful in this phase and describes the overall bioreaction. It can be said that we have a representative biomass estimation. However, the calculated h value exceeded the thresholds in the low q_s phases. This could be a result of switching the mode of the metabolism detected by a lower respiratory quotient RQ of approximately 0,8 instead of 1 in these phases. Furthermore, the maintenance also had a large proportion on the metabolism in these low q_s phases. Otherwise, this high consistency value h could be also from the very low variances resulting in a very sensitive test. It is obvious, that the most accurate rate is r_S followed by r_{CO2} and r_{O2} in contrast to Fig. 11 subplot B.

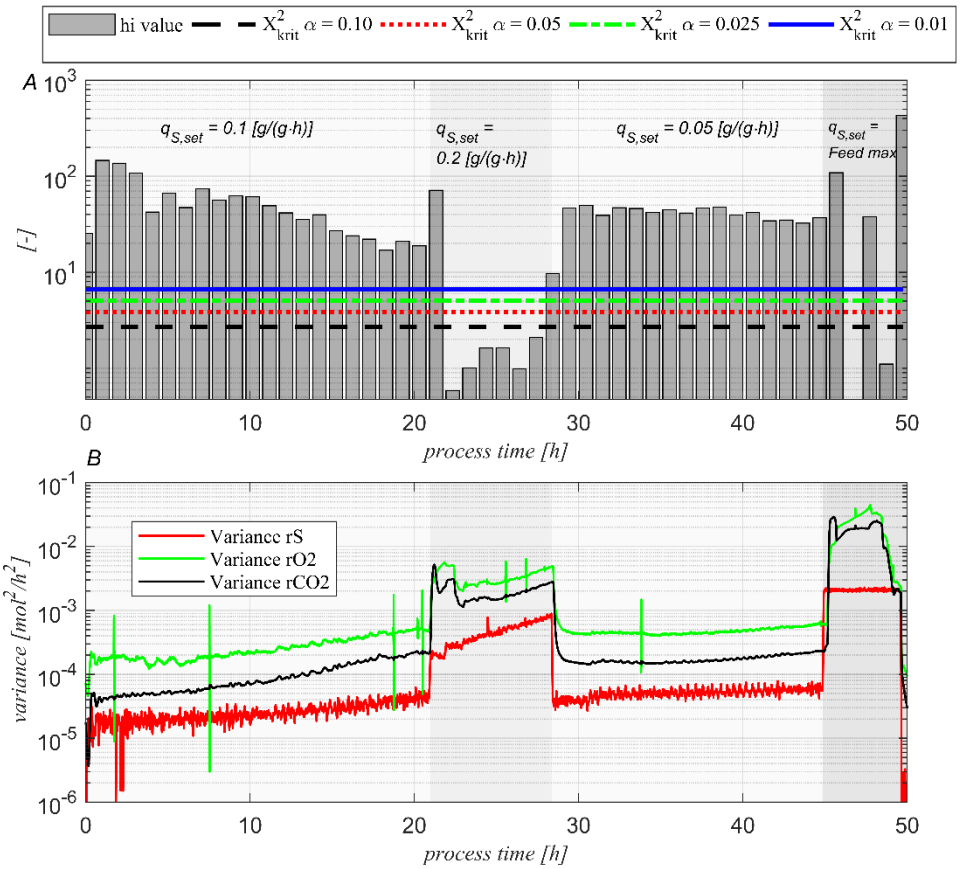

Fig. 31 K2S1 ***A:*** *Statistical Chi-quadrat test in 60 min intervals for model validation. The straight lines represent the levels of significance for the statistical test.* ***B:*** *Time-dependent variances of the K2S1 model input rates.*

The calculated substrate rate of glucose (red line), the confidence intervals with a level of significance (α= 0.05) and the reconciled rate in c-mol are shown in Fig. 32. It could be observed that the calculated and the reconciled rate was exactly equal in all q_s phases in contrast to the similar plot of the K3S1 model. The reconciliation procedure did not significantly modify the calculated rate.

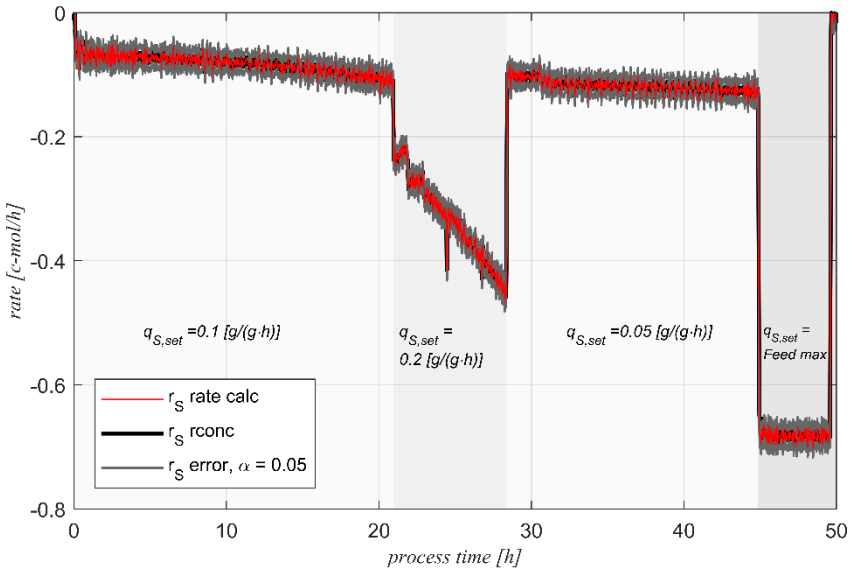

Fig. 32: K2S1 substrate consumption rate r_S. The red line including grey confidence intervals (α= 0.05) represents the online calculated r_S. The reconciled rate r_S is shown as black trend.

The calculated oxygen uptake rate r_{O2} (red line), the confidence intervals with a level of significance α = 0.05 (grey line) and the reconciled rate (black line) in mol are shown in Fig. 33. It could be observed that the reconciled rate was slightly lower than the calculated rate in the phases q_s = 0 g/(g·h) and q_s = 0,05 g/(g·h). There were mismatches at the beginning of phase q_s = 0,2 g/(g·h) and phase feed max. The reconciled r_{O2} rate in these parts were underrated in contrast to the calculated rate. It could be observed that the mismatches were significantly smaller after the reconciliation procedure in comparison to the r_{O2} analysis in K3S1 shown in Fig. 18.

The calculated carbon dioxide evolution rate r_{CO2} (red line), the confidence intervals with a level of significance α = 0.05 (grey line) and the reconciled rate (black line) in mol are shown in Fig. 34. It was found out that the calculated rate was similar to the reconciled rate over wide q_s phases. Slight distinctions were also observed at the beginning of the q_s = 0,2 g/(g·h) phase and at the beginning of the maximal feed phase. These distinctions were not so pronounced in contrast to the reconciled rate of the K3S1 model shown in Fig. 20.The errors of the rates r_S, r_{O2} and r_{CO2} were the same as in the K3S1 model shown in subchapter 6.4.1.

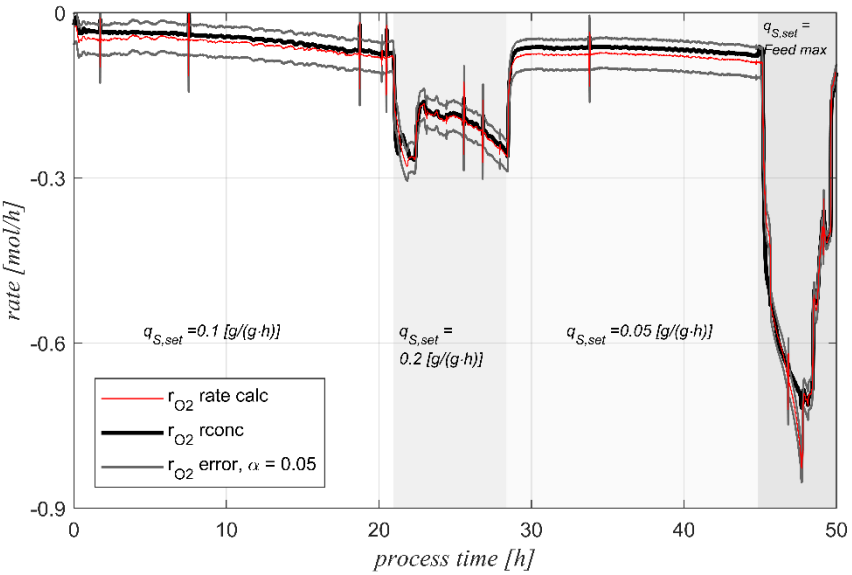

Fig. 33: K2S1 oxygen uptake rate r_{O2}. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_{O2}. The reconciled rate r_{O2} is shown as black trend.

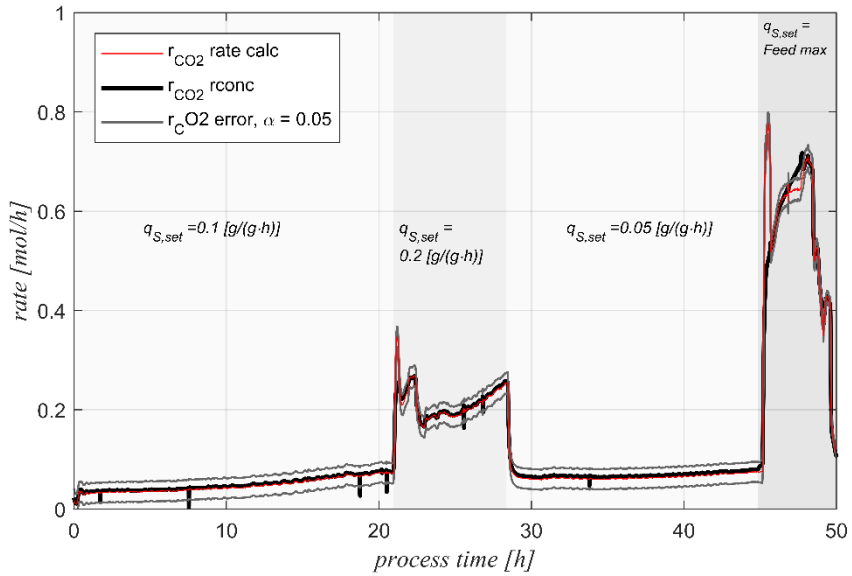

Fig. 34: K2S1 carbon dioxide evolution rate r_{CO2}. The red line including grey confidence intervals (α= 0.05) represents the online calculated r_{CO2}. The reconciled rate r_{CO2} is shown as black trend.

The biomass estimation results with the K2S1 model of each balance (C and DoR) demonstrated in Fig. 35 were consistent in comparison to the K3S1 model. The estimated biomass with the calculated rates was also located in the determined biomass of the DoR balance shown in Fig. 35 in subplot A.

The results of the reconciliation procedure of the K2S1 model differs from the previous K3S1 model. Here, the estimated biomass was located in the determined biomass of the C-balance in contrast to model K3S1. A look on subplot B shows the cumulative absolute balance errors of the C and DoR balance. The cumulative absolute errors are 12 magnitudes lower in comparison to the absolute balance errors of the K3S1 model after the reconciliation procedure.

The critical examination according Eqs. 6-4 to Eqs. 6-6 revealed in subplot C led also to different results in contrast to the K3S1 evaluation. Here, the C-balance is more significant for the biomass estimation than the DoR balance. The sum of the rate variances in the C-balance is lower than in the DoR balance, that means that the C-balance is more accurate and has more influence on the reconciliation. As a general statement, it can be said that the K2S1 equation model system describes the overall bioreaction well with higher q_s rates such as 0,2 g/(g·h). This circumstance led to a small discrepancy of the estimated biomass with calculated and reconciled rates.

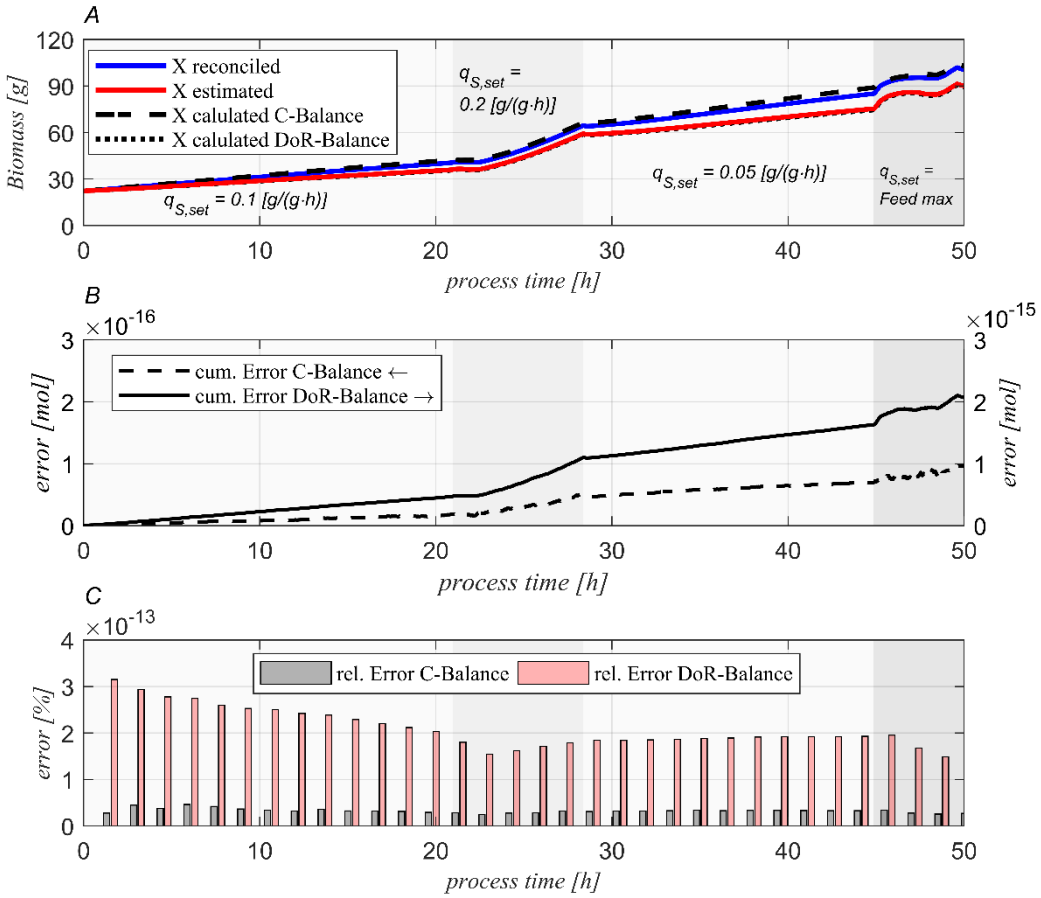

Fig. 35: K2S1 ***A:*** *Calculated biomass from C, DoR and N balance with additional estimated and reconciled biomass results as a function of time.* ***B:*** *Cumulative error of C and DoR balance after reconciliation.* ***C:*** *Relative balance errors related to total C or rather electron actions (DoR) across the balance space.*

The comparison of the estimated biomass with the OD and DW measurements according Fig. 36 shows an improvement in contrast to the K3S1 model. Subplot A shows the absolute estimated biomass with calculated and reconciled rates, including the OD and DW results. The discrepancies between the two different estimated biomasses are low in the first two q_s phases and in the last q_s phase.

Subplot B reveals the absolute error of the estimated biomasses in comparison to the optical density measurements at the respective time points. Subplot C shows the absolute error of the estimated biomasses in comparison to the dry weight results at the respective time points. The absolute mean errors of the estimated biomasses with respect to the OD respectively DW measurement are also shown in subplot B and C.

It could be observed that the mean error of the estimated biomasses (Xrconc, Xest) with and without reconciled rates in terms of the optical density measurements yielded in 8,88 or rather 4,3g. This fact shows in this case that, the rates reconciliation improves the performance of the estimation. This happens, if the equation model (K2S1) passes the consistency test or if the failed consistency test is only caused by too small variances. The mean error on the estimated biomasses compared to the dry weight biomass determination yielded in 10,63 g respectively 6 g. Reconciling the rates also improved the biomass estimation compared to the OD measurements. It could be stated that the best accordance is given by the estimated biomass with reconciled data in comparison to the OD measurement. In this context, a biomass estimation with a relative error of 5 % based on the biomass produced could be obtained.

The valuation of the balances in relation to the OD and DW measurements was assessed by their weighting on the estimated biomass results. Furthermore, balances are classified in calculated and reconciled rates input according Table 8.

Table 8: K2S1 Performance of each balance with calculated and reconciled rates input

Balances	Rank with calculated rates	Rank with reconciled rates
Degree of Reduction Balance	best	satisfactory
C-Balance	satisfactory	best

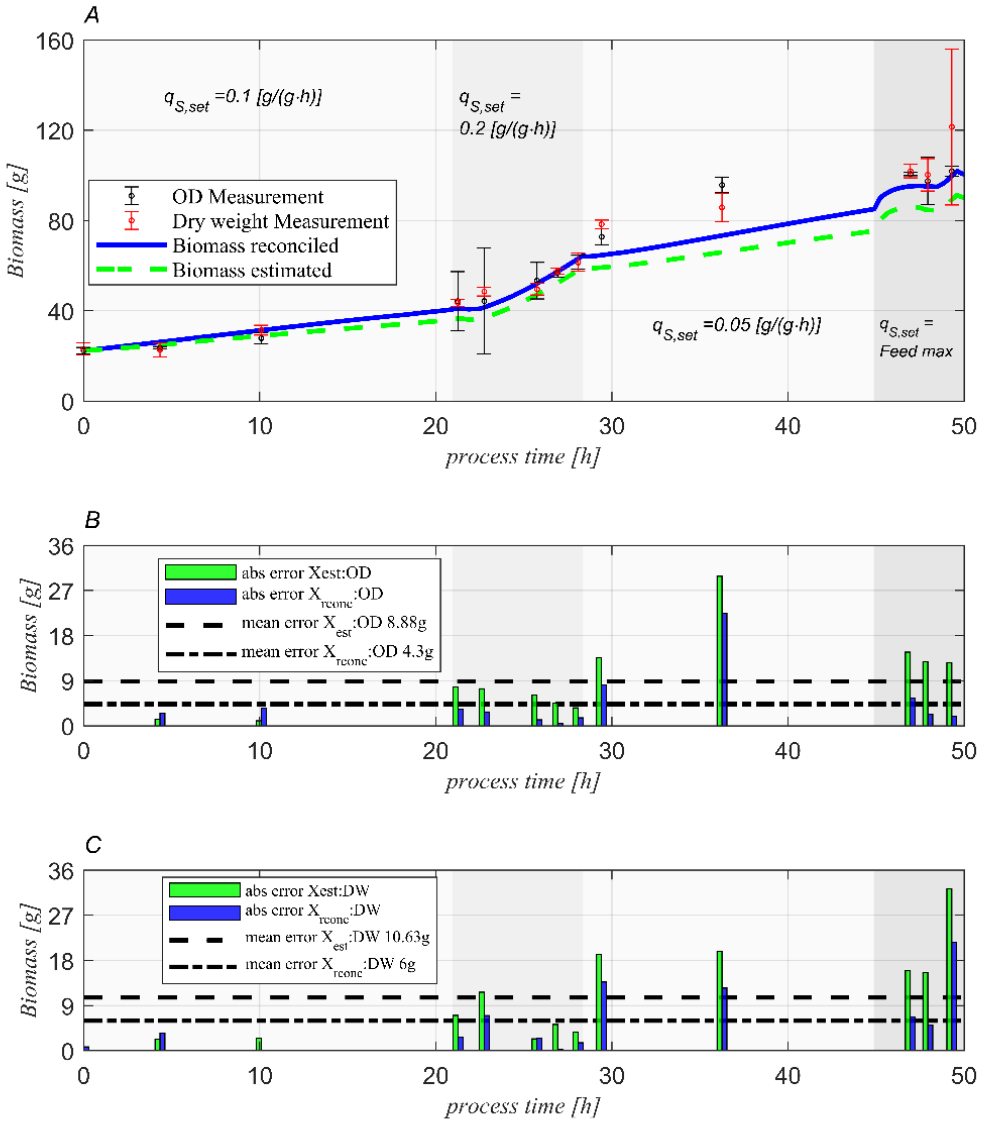

Fig. 36: K3S1 Independent biomass content comparison and validation. 13 samples were under consideration. ***A:*** *Estimated and reconciled biomass trend as a functions time in different q_s phases, including optical density (OD) and dry weight (DW) measurements.* ***B:*** *Absolute errors of estimated and reconciled biomass related to the optical density (OD) measurement method. OD mean error over all different q_s process phases.* ***C:*** *Absolute errors of estimated and reconciled biomass related to the dry weight (DW) measurement method. DW mean error over all different q_s process phases.*

6.4.3 K2S1_2

The following analysis is similarly structured such as subchapter 6.4.1. The assumptions are equal to the run before. This additional run with the K2S1 model was performed with a feed concentration of 400 g/l instead of 200 g/l to investigate the behaviour of higher q_s rates. In this task, the run was executed at a specific substrate uptake rate of q_s = 0,3 g/(g·h) and q_s = 0,4 g/(g·h). Fig. 37 shows the calculated specific uptake rate, including their setpoints as a function of time. A good feeding control performance is at the first third of the phase q_s = 0,3 g/(g·h). The integral term in the PI controller has not had the required control power which was caused by the exponential biomass growth. Especially, this effect is shown in the phase q_s = 0,4 g/(g·h). Further optimizations can be done in this field in prospective works. Furthermore, the whole glucose was metabolised as a result of the samples investigation.

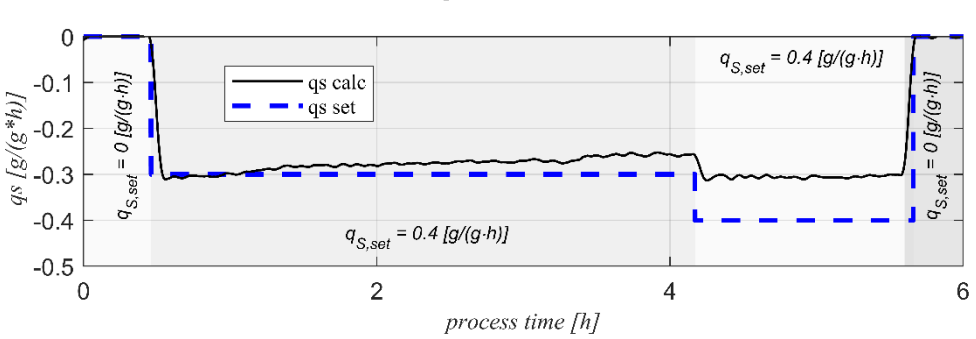

Fig. 37: Specific calculated glucose uptake rate with q_s setpoints.

The application of the global consistency test on the K2S1 model with higher rate values leads to different results in contrast to Fig. 31. Here, the behaviour shown in Fig. 38 is caused by the higher setpoints of the specific uptake rates and a higher starting biomass amount of 93 g. Subplot A shows the calculated consistency index h also as a bar plot with a bar width of 10 minutes. The threshold values represent the levels of significance with α = (0,1 0,05 0,025 0,01). Contrary to the findings of the previous K2S1 investigation, here the K2S1 model fits in combination with the executed q_s feeding setpoints. That means, that the biomass estimation is successful. It is also obvious that the consistency index h increases if there is no feeding phase or a possible ethanol metabolism.

The variances of the rates are shown in subplot B. The variances of all rates are significantly higher in contradiction to the variances of the previous run. This means, the risk of a too sensitive global consistency test caused by very low variances is minimal.

A higher r_{O2} variance related to the other variances in contrast to the r_{O2} variance results in Fig. 31 subplot B was also identified. This behaviour is caused by the considered pure O_2 mass flow controller and its accuracy allowance.

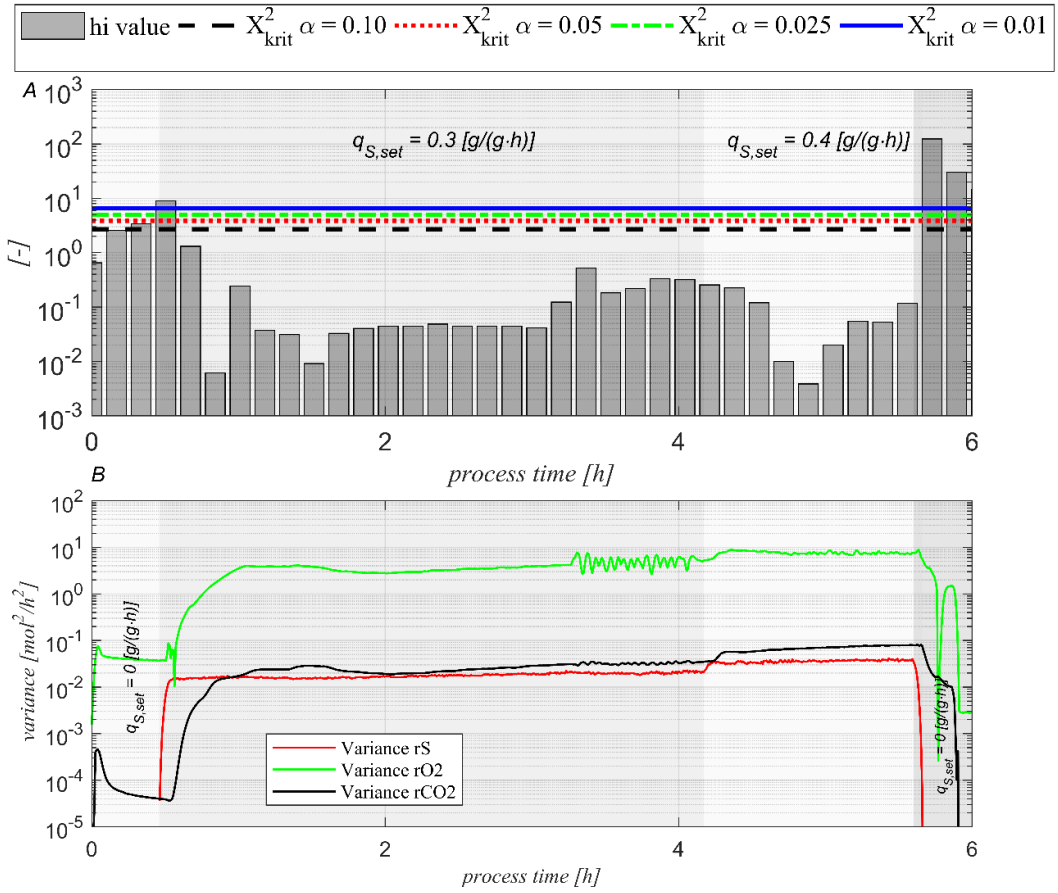

Fig. 38 K2S1_2 ***A:*** *Statistical Chi-quadrat test in 60 min intervals for model validation. The straight lines represent the levels of significance for the statistical test.* ***B:*** *Time-dependent variances of the K2S1_2 model input rates.*

The calculated substrate rate of glucose (red line), the confidence intervals with a level of significance (α= 0.05) and the reconciled rate in c-mol is shown in Fig. 39 on the next side. It could be observed that the calculated and the reconciled rate is exactly equal in all q_s phases. Moreover, the reconciliation procedure does not significantly modify the calculated rate.

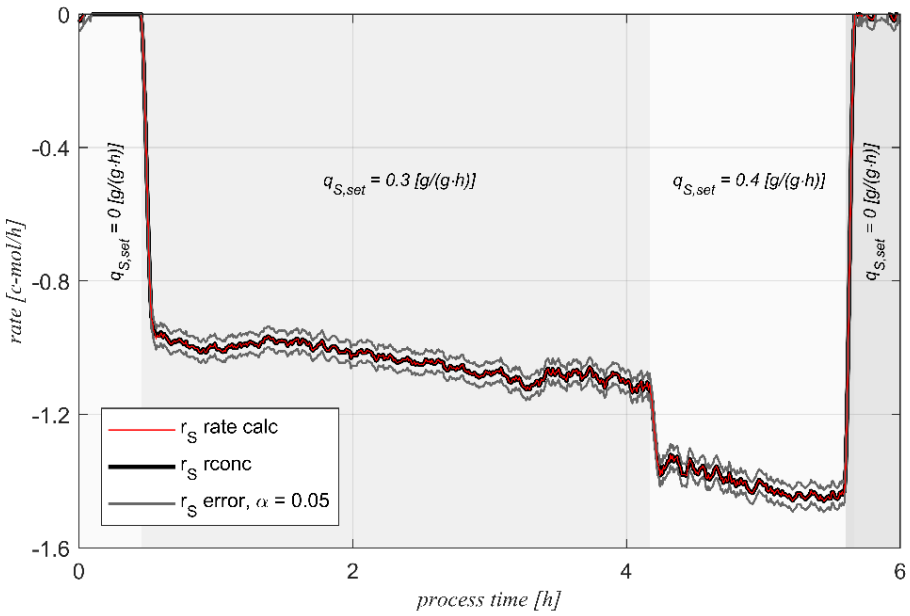

Fig. 39: K2S1_2 substrate consumption rate r_S. The red line, including the grey confidence intervals (α = 0.05) represents the online calculated r_S. The reconciled rate r_S is shown as black trend.

Fig. 40 shows the relative error of the substrate rate as the sum of the propagation and noise error. The relative error is approximately 5% during the feeding phase. It is obvious that the relative error goes to infinite if the rate goes to zero. This behaviour can be seen before the feed switch on and after the feed switch off timepoints.

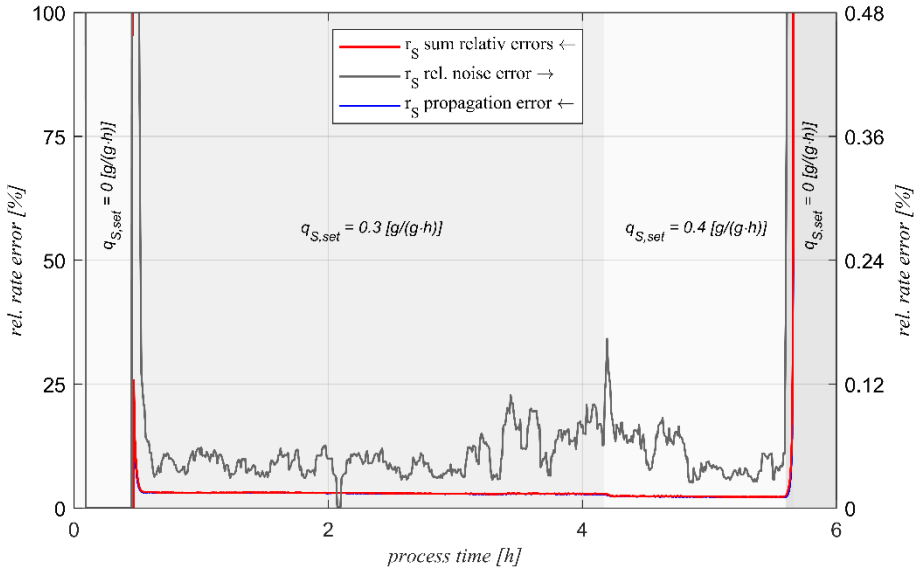

Fig. 40: K2S1_2 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the substrate consumption rate r_S.

The calculated oxygen uptake rate (red line), the confidence intervals with a level of significance ($\alpha = 0.05$) and the reconciled rate in mol is shown in Fig. 41. The O_2 mass flow controller is here considered, in contrast to the K3S1 and K2S1 evaluation. This circumstance leads to bigger confidence intervals in contrast to the results in Fig. 33. It could be observed that the calculated and reconciled rate is almost similar in the two thirds of the experiment. The discrepancy increases when the O_2 mass flow controller entered in the dissolved O_2 control system. Which shows the noise at the end of the $q_s = 0{,}3$ g/(g·h) phase.

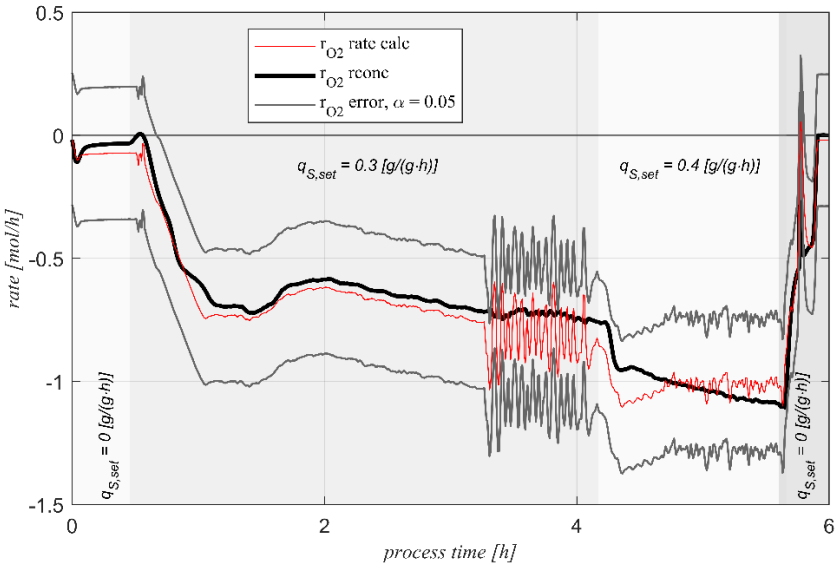

Fig. 41: K2S1_2 oxygen uptake rate r_{O2}. The red line, including the grey confidence intervals (α= 0.05) represents the online calculated r_{O2}. The reconciled rate r_{O2} is shown as black graph.

Fig. 42 shows the relative error on the substrate rate as the sum of the propagation and noise error. The relative error is higher in constrast to the K2S1 and K3S1 evalution because of the considered additional O_2 massflow controller.

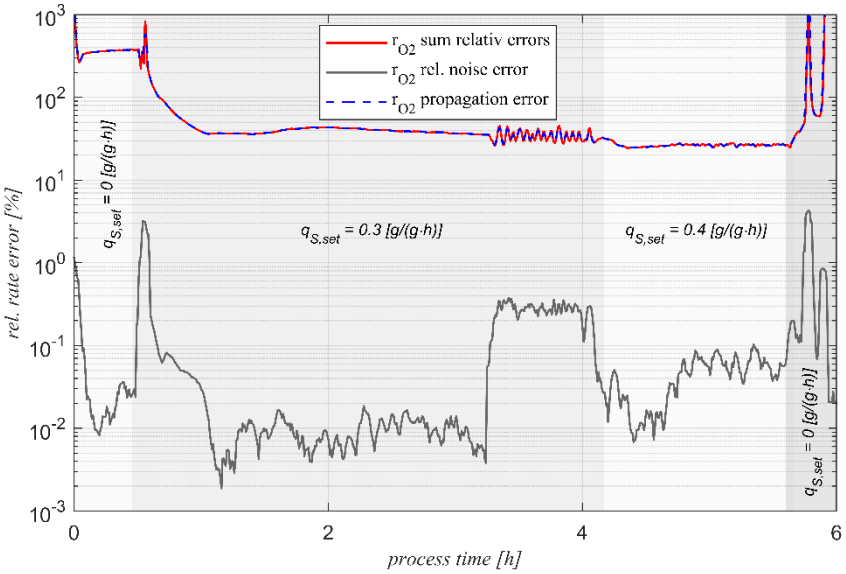

Fig. 42: K2S1_2 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the oxygen uptake rate r_{O2}.

The calculated carbon evolution rate (red line), the confidence intervals with a level of significance (α = 0.05) and the reconciled rate in c-mol is shown in Fig. 43. It could be observed that the calculated and the reconciled rate is exactly equal in all q_s phases. In this context, the reconciliation procedure did not significantly modify the calculated rate.

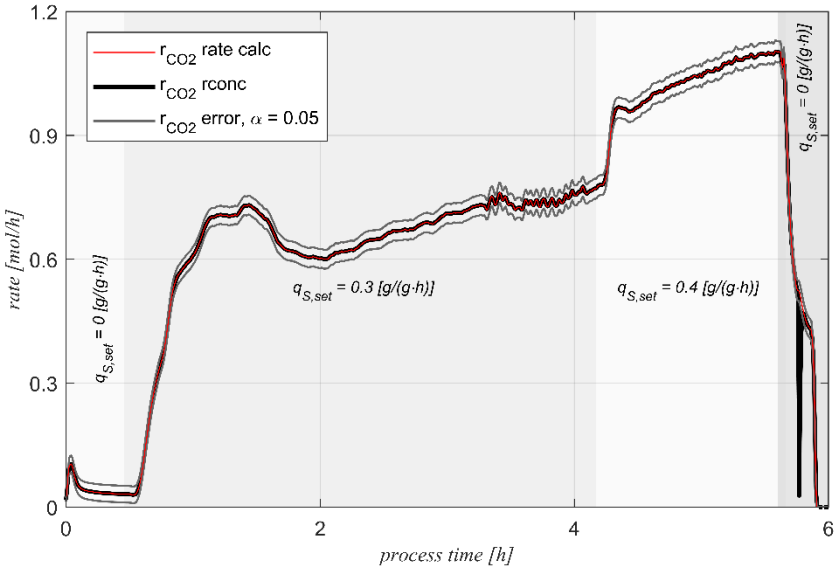

Fig. 43: K2S1_2 carbon dioxide evolution rate r_{CO2}. The red line, including grey confidence intervals (α= 0.05) represents the online calculated r_{CO2}. The reconciled rate r_{CO2} is shown as black graph.

Fig. 44 shows the relative error of the carbon evolution rate as the sum of the propagation and noise. The relative error is significantly lower in the feeding phases as in contrast to the previous K3S1 run. Here, the mean relative error remains around 5%.

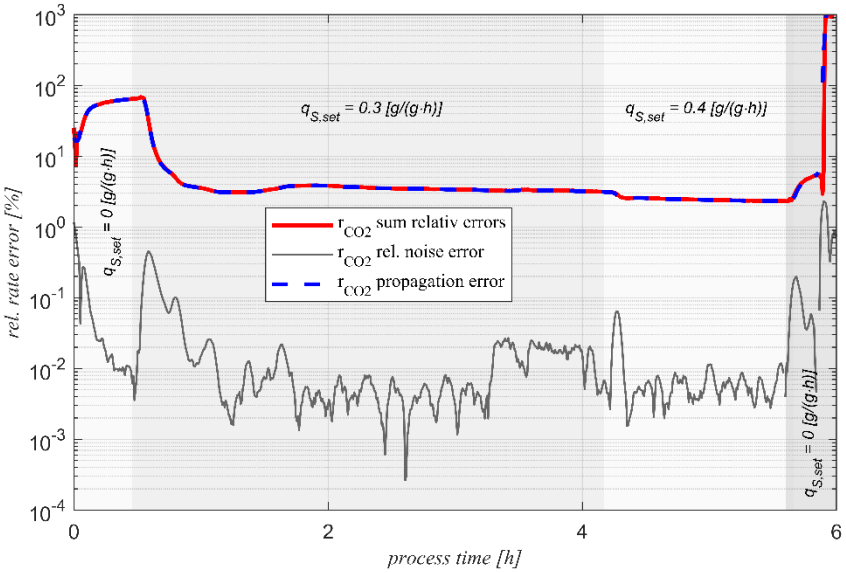

Fig. 44: K2S1_2 time-dependent relative error (red) separated in noise (grey) and propagation error (blue) of the carbon dioxide evolution rate r_{CO2}.

The jagged course in all rates plots at the end of the q_s =0,3 g/(g·h) phase was caused by the high minimum flow rate of the O_2 mass flow controller. The discrete turn on and turn off of this device in combination with the minimum flow rate of 0,2 l/min led to a reduced O_2 control stability. This reduced stability is strong when the air mass flow controller could not really hold the O_2 concentration on the setpoint. Consequently, this behaviour led to a higher calculated consistency value at a process time of 4 hours, shown in Fig. 38 subplot A. This behaviour can be avoided by a more convenient mass flow controller.

Fig. 47 shows the O_2 concentration control with the agitator speed and oxygen input stream as the actuating parameters. The O_2 concentration broke in at the start of feeding (q_s = 0,3 g/(g·h)). The control system required approximately one hour to reach the setpoint of 35 %. That is not uncommon with approximately 30 g/l biomass and an initial specific substrate uptake rate of 0,3 g/(g·h). The cascade control can be explained by the actuating variables in subplot B. When the agitator speed reaches the maximum, then the air mass flow controller enters the control system. If the mass flow controller reaches the maximum, then the additional O_2 mass flow controller enters the control system. The activation of the O_2 controller can be identified at a time point of 3,5 hours.

As a result of this plot, it can be said that the implemented O_2 control provides the required performance also with higher q_s rates. Furthermore, the additional pure oxygen control function extends the existing design space. In this context and for this specific case, a solid oxygen supply must be guaranteed. In conclusion, the oxygen supply could not be provided after approximately 5 hours, which caused the termination of the experiment.

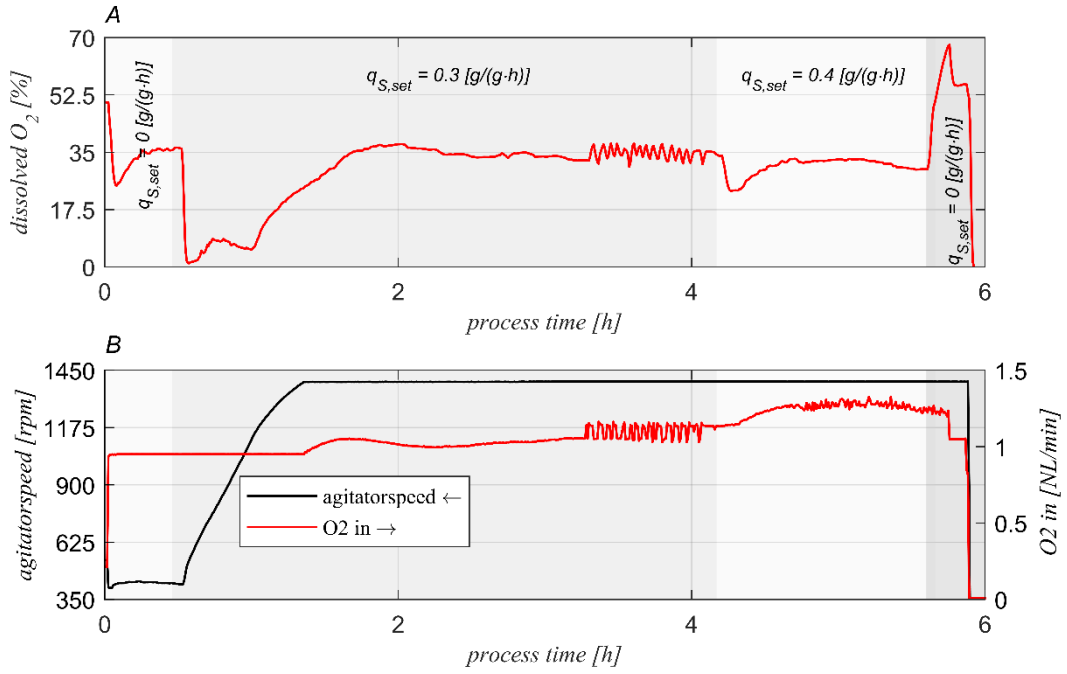

Fig. 47: K2S1_2 ***A:*** *closed loop controlled dissolved O_2 on 35%.* ***B:*** *Subplot shows the actuating variables agitator speed and aeration rate of the cascade control system.*

The results of the biomass estimation with the K2S1 model are revealed in Fig. 48. Subplot A shows the calculated biomass from each balance, including the estimated biomass with the measurements calculated rates.The estimated biomass coincided with the result of the DoR balance. The following subplots B and C represents the relative and cumulative errors of each balance after the estimation. The DoR-balance also describes the overall bioreaction more precise than in contrast to the C-balance. Moreover, the cumulative error of the C-balance is approximatly 40 times higher than the cumulative error of the DoR balance. In an another interpretation, it can be said that the biomass production is overrated in the C-balance and slightly underated in the DoR-balance.

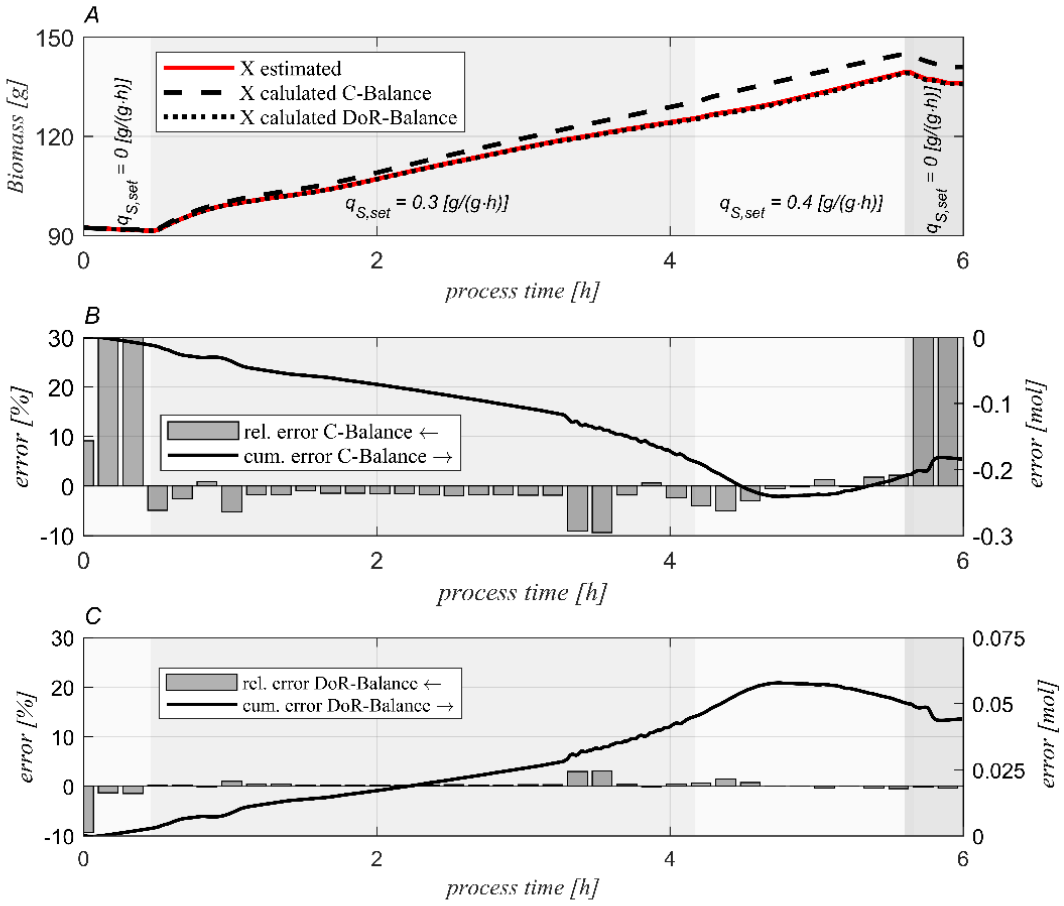

Fig. 48: K2S1_2 ***A:*** *Calculated biomass from C and DoR balance with additional estimated biomass result as a function of time.* ***B:*** *C-balance cumulative error in mol. The relative error relates to the total carbon across the balance space.* ***C:*** *DoR-balance cumulative error in mol. The relative error relates to the total electron's actions across the balance space.*

The results of the estimated biomass with the reconciled rates are shown in Fig. 49. The estimated biomass is also predicted on the C-balance. This behaviour can be explainted with the variances of the different rates. The variances shown in Fig. 38 reveals that variances of the r_S rate and the r_{CO2} rate are smaller than the variance of r_{O2}. The sum of the variances in the C balance is significantly lower than the sum of the variances in the DoR balance. This fact is caused by the high variance of the r_{O2} rate. Futhermore, this suggest that the C-balance is more significant for the reconciliation prodecure than the DoR balance.

Subplot B shows the cumulated error of each balance after the reconciliation procedure. The results are near to zero. In conclusion, the equation model also fits with higher q_s rates. Subplot C shows the relative error of the balance explained in more detail in subchapter K3S1. It can be summarized that the relative error related to the sum of the input, output and accumulation amount is higher on the DoR balance with the reconciled rates.

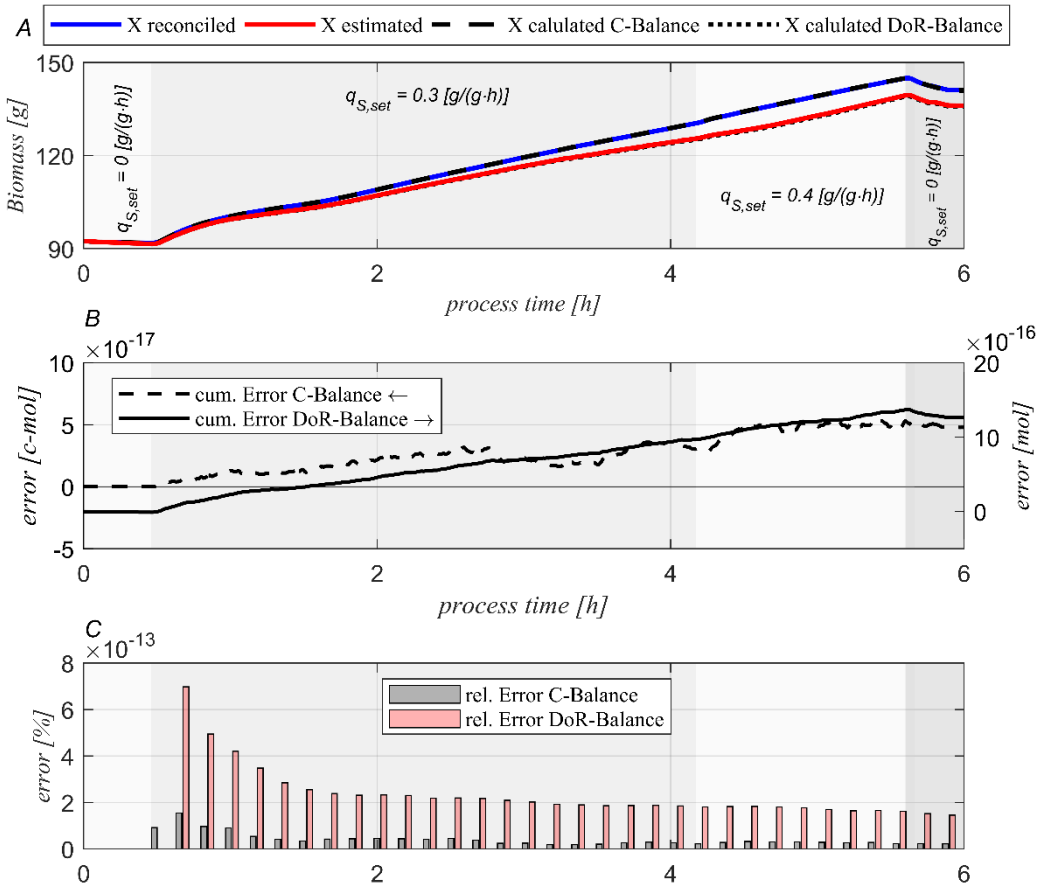

Fig. 49 K2S1_2 ***A:*** *Calculated biomass from C and DoR balance with additional estimated and reconciled biomass results as a function of time.* ***B:*** *Cumulated error of the C and DoR balance after reconciliation.* ***C:*** *Relative balance errors related to total C and electron actions (DoR) across the balance space.*

The comparsion of the estimated biomass with calculated and reconciled rates in contrast to the OD and DW measurements is shown in Fig. 50. Only 3 samples were done in this run. The first sample was to initialize the calculation with the biomass content of 93 g. It has been observed that the OD and DW measurements differ singnificantly in the q_s = 0,3 g/(g·h) phase.

In subplot B, it can be observed that the mean error of the estimated biomasses with calculated and with reconciled rates in comparison to the optical density measurements yielded in same value of 5,4 g. This fact demonstrates that the reconciliation did not improve the performance in this run.

In subplot C, the mean error of the estimated biomasses is represented with calculated and with reconciled rates in comparison to the dry weight measurements. The mean errors related to the DW measurements yielded in 8,8 g respectively 9,2g. This also suggests that the reconciliation procedure rather impair the estimation related to the DW measurements. The better estimation results in relation to the OD measurements are caused by the initial biomass value for the calculation in contrast to the DW comparison.The results of the mean errors related to the alternative measurements can be also related to the generated biomass. The results in this view yielded in a relative error of approximatly 10 % compared to the OD and approximatly 18% compared to the DW measurements. It can be generally stated that the offline biomass measurements are obviously less accurate than the estimated biomass.

In conclusion, the additional run with the K2S1 model and higher q_s rates is fitting the overall bioreaction checked by the consistency test. The formation of ethanol with higher q_s rates of 0,233 g/(g·h) did not lead to an not passing consistency test. Furhtermore, the higher rates, including the higher variances avoid a too sensitive consistency test in contrast to the results in subchapter 6.4.2. It can be stated the general performance of this run was satisfying the demands for the estimation of the biomass via this robust soft sensor concept. Also the balances can be weighted according Table 8.

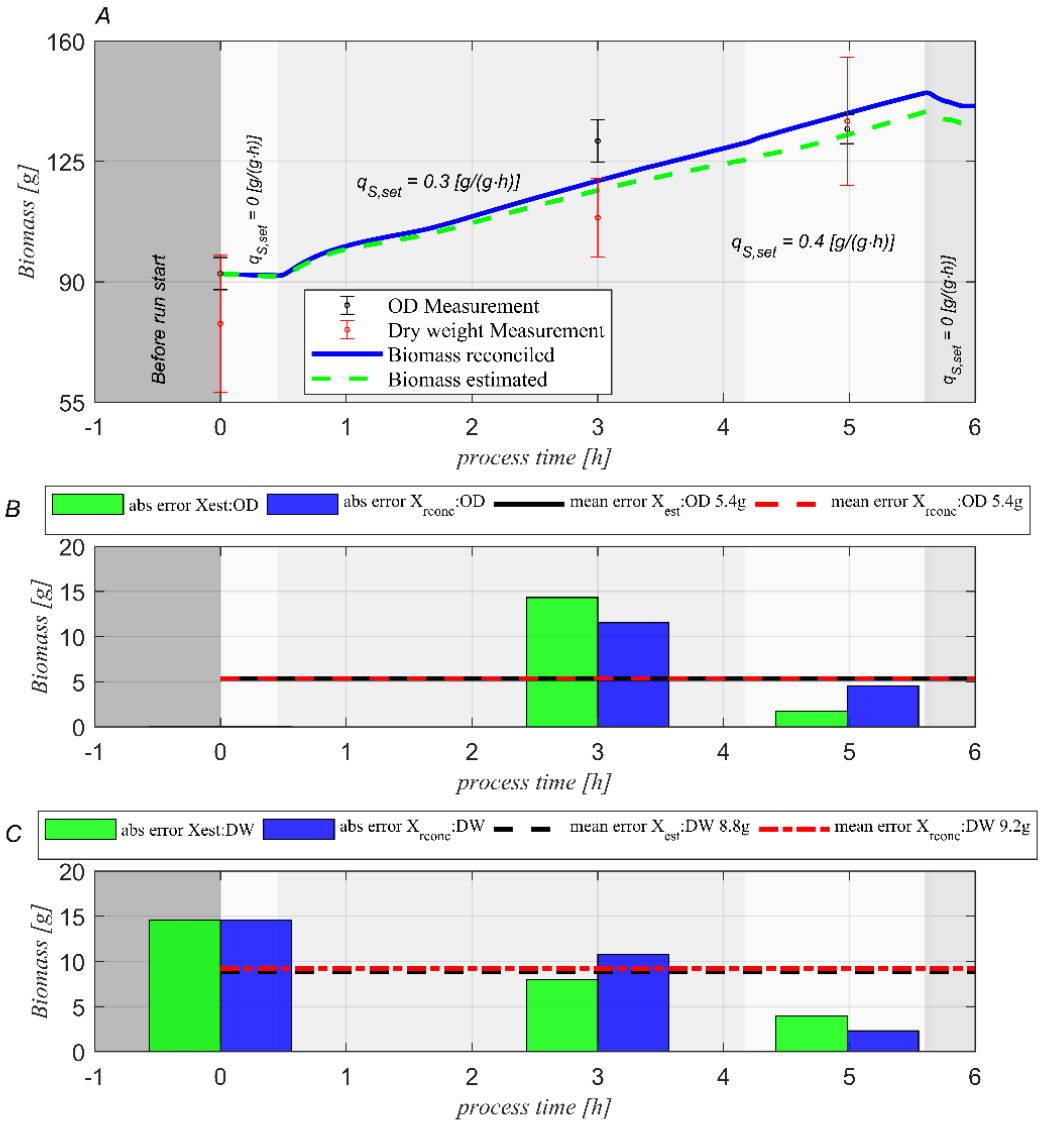

Fig. 50: K2S1_2 Independent biomass content comparison and validation, 3 samples were under consideration. ***A:*** *Estimated and reconciled biomass trends as a function of time in different q_s phases, including optical density (OD) and dry weight (DW) measurements.* ***B:*** *Absolute errors of the estimated and reconciled biomass related to the optical density (OD) measurement method.* ***C:*** *Absolute errors of the estimated and reconciled biomass related to dry weight (DW) measurement method.*

6.5 Workflow for a successful performance of microbial fermentations

Based on [23], a general strategy to run a successful cultivation with a robust biomass soft sensor can be summarized in a workflow shown in Fig. 51.

The reasons for the advantages of such this workflow is at first demonstrated by the first step of defining assumptions for the bioreaction network. The operator is encouraged to think about possible assumptions then to develop a simple bioreaction network that can describe the overall bioreaction with simple accessible measurements or derived quantities.

In the next step, the stochiometric matrix for various species and different reactions (C, DoR, N) is formulated. This reaction network is the basis for a basic experimental simulation. The practical benefit of a general fed batch simulation study, such as in this thesis, is that the setup can be constructed for the experiment without any knowledge from previous similar cultivations. Any problems, especially regarding to limitations especially of oxygen transfer, maximum volume and feeding profile can be prevented.

Subsequently after the setup of the reactor system, a “water” run should be conducted to test the control systems and the connection link to the process control system also via python server to MATLAB and in reverse direction. It needs to be stated at this point that a fermenting run should be performed under stable conditions without malfunctioning dissolved O_2, pH, temperature, and feed strategy controls that disturb the whole soft sensor system.

The initial biomass concentration after the batch phase must be known as initial biomass concentration for the biomass estimation calculator. In contrast to previous works [22] [39] [19] [28],this calculator takes the changing variances of the inputs during the run into account, which makes the biomass estimation with reconciled rates robust.

After the calculation of the reduced redundancy matrix and the covariance matrix of the residuals, it can be computed the constency value h. This value is also highly dependent on the variance from the rates as function of the time. If a gross error is detected, the gross error can be identified by sub sequential elimination of the measured rates. If there is no identified error on the measured rates, then the error must be in the model. The normal procedure is to modify the assumptions or bioreaction network. The input rate can be adjusted afterwards in the case of a gross error on a measured rate and the calculation procedure repeated. If the consistency test does not detect a gross error, then the measured rates can be readily reconciled. An improved biomass result can be computed with the reconciled rates.

Checking the control system, especially in the first 5 steps from the assumption involves a lot of work. These steps are time consuming, especially setting the system up in the ‘’water’’ run and is a disadvantage of this workflow. Moreover, the concept requires a certain amount of expert knowledge, which could represent a barrier to the use of the workflow.

Nonetheless, the showed soft sensor concept for the biomass estimation represents a robust soft sensor concept and due to the modular setup in MATLAB a slight adaption and extensible platform for other cultivations. The general concept can be also correlated with other online sensors to use MATLAB for advanced mathematical requirements.

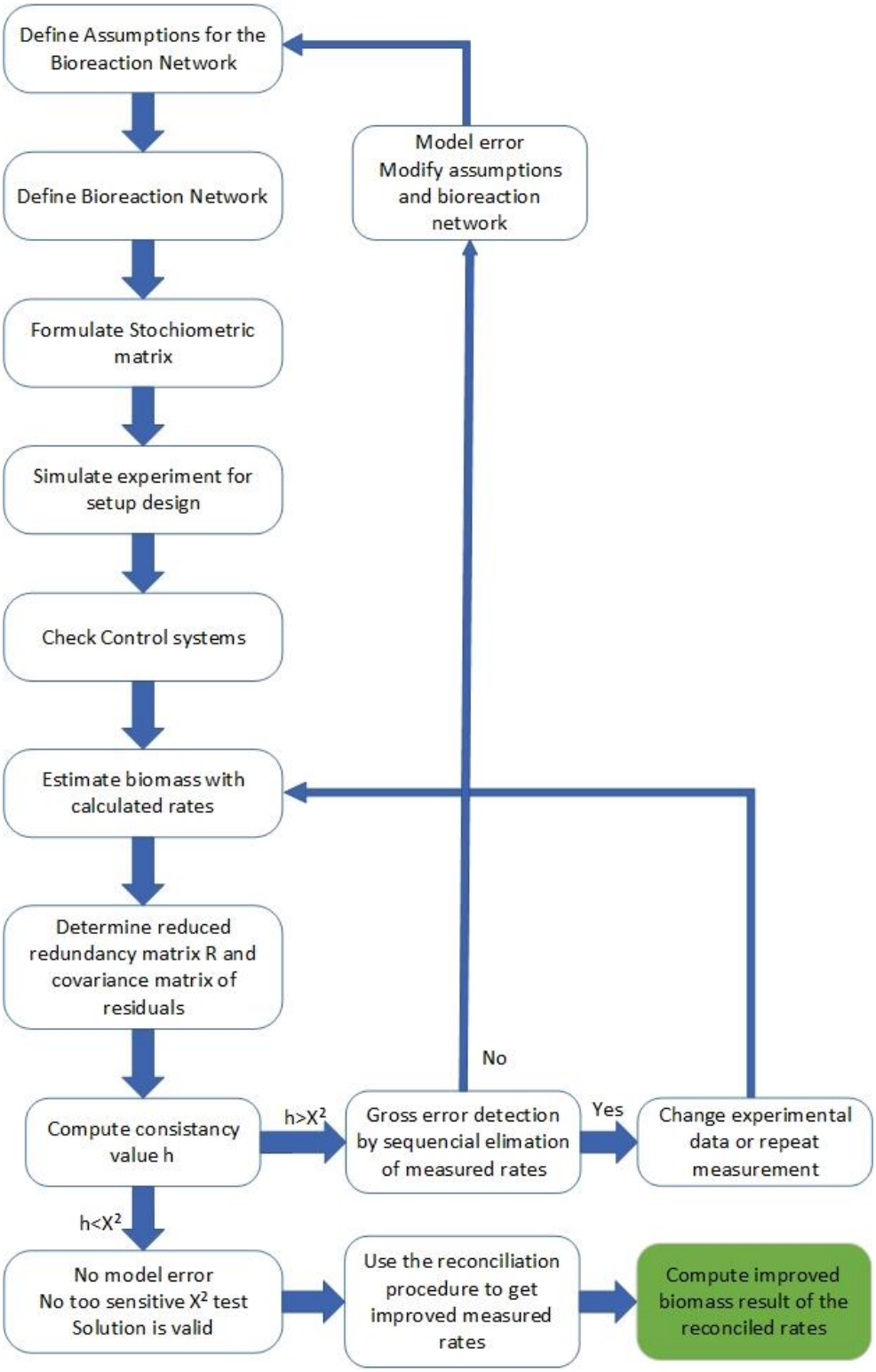

Fig. 51 Workflow for successful performance of microbial fermentations including estimating the biomass via soft-sensor. The green box represents the goal of the soft sensor concept.

7 Summary

The investigation of pervious cultivation runs showed that robust control strategies in combination with a successful soft sensor for the biomass estimation were not applied. Therefore, it is recommended to make a simulation of the experiment to avoid to be out of the design space. A general aspect of all sub - evaluations from K3S1 to the extended K2S1_2 run shows the importance of the statistical consistency test. The test verifies the model and the measured input rates. If the consistency test is not performed, wrong estimations can occur. This fact is demonstrated in the analysis with the K3S1 model, were the faulty nitrogen balance seems to be the most significant balance for the overall bioreaction after the rates reconciliation procedure. This is caused by the time-dependent consideration of the measurement errors.

The key novelty in this soft sensor concept is the real-time consideration of the measurement errors from the devices, including their propagation and measurement noise of the rates. This information was used for the consistency test and the reconciliation procedure during the run to estimate the best biomass result related to the current process conditions. This circumstance makes the concept robust against changing process conditions. This is the key improvement in contrast to a traditional soft sensor in previous works [19] [22] [28] [39] were the error of the rates was assumed by a constant value. Furthermore, [29] only investigated the behaviour of a biomass estimating soft sensor with a considered error propagation by Monte Carlo simulation with in silico generated data. In contrast to [29] this thesis investigates the behaviour in a real-time application. Therefore, following statements can be done for this robust soft sensor concept.

The results of the K3S1 subchapter show that small rates are linked with high relative errors which lead to very small variances. In further consequence, this can lead to a too sensitive consistency test whereby the informative value is reduced. A model error was detected and led to a reduced model without the nitrogen balance. The system passed the consistency test with the reduced K2S1 model with a specific feed rate of q_s = 0,2 g/(g·h). An additional run was performed to investigate the soft sensor concept with higher rates. It has been shown that the concept becomes stronger due to the decreased relative errors of the higher rates. This improved accuracy of the calculated rates can be impaired by poorly selected measurement devices. In this context, a suitable equipment is a mandatory for a successful soft sensor application. Moreover, it has been shown that the most significant balance for the biomass estimation result is the DoR-balance with calculated rates and the C - balance with the reconciled rates. This statement can be done with this setup, a *Saccharomyces cerevisiae* cultivation and a passed consistency test.

The estimated biomass results were compared with optical density and dry weight measurements at certain time points. It can be shown that these reference measurements are more inaccurate at higher biomass concentrations. The best results for the biomass estimation were achieved with the K2S1 model in comparison to the dry weight measurements. The mean relative errors of the estimated biomass are less than 10 % compared to the reference measurements. The investigation of the control strategies for a pre - defined run shows different results. The dissolved O_2 concentration control with the extended pure oxygen function provides the requirements for a smooth run. The control for the specific feeding profile has the necessary accuracy in low feed flows but declined with higher specific uptake rates. This was caused by the not considered changing requirements for an optimal setting of the PID controller.

A workflow for a successful performance of microbial cultivations based on these results is developed. It shows the requirements for a successful soft sensor application. It can be concluded that this novel soft sensor concept provides a precise biomass estimation in real – time.

8 Outlook

The biomass amount as a catalytic converter to develop a product of interest is one of the key variables in any microbial cultivation processes. In this thesis a *Saccharomyces cerevisiae* cultivation was only performed as an example to investigate the behaviour with this novel error consideration approach. The prospective aim of this work should be the implementation of this concept in different cultivations with similar initial process conditions as a first step. The slight implementation is provided due to the modular setup of the created calculator functions. In contrast to that, robust data transfer should be improved in terms of easy use and stability. A centralized approach for monitoring and control in combination with a decentral data management could reduce the susceptibility to errors.

Another improvement can be achieved with the workflow of [29]. This paper provides a workflow for an optimized pre-selection of the measurement devices for improved estimation results. From this point of view, it is interesting to compare the predicted results to a real-time application. Chapter K3S1 and K2S1 showed non-passing consistency tests behaviours. A suitable approach could be to enlarge the constant parameter *windowsize* for the rates calculation. This reduces the relative errors of the rates and enlarge the variances. In this context, a too sensitive consistency test can be avoided. A next approach could be done to control this parameter *windowsize* in relation to pre-defined maximum relative errors.

As shown in this thesis, 3 different balances were initially used to characterize the overall bioreaction for the biomass estimation. The nitrogen balance did not quantify this overall bioreaction system due to wrong assumptions. Further works can be done to make the nitrogen or other balances accessible to describe a wide range of bioreaction systems. In this context, useful literature was found [20] [25] [40] [41] [42]. Various specific uptake rates can cause temporary metabolites which are not considered with a statically C-balance. The advantage of the nitrogen balance consideration lies in the fact that possible carbon containing temporary metabolites without nitrogen such as ethanol do not disturb the nitrogen balance. An implementation of the N-balance needs a precise online determination of the nitrogen concentration in the broth. NIR and MIR spectroscopy can be used for online ammonia measurements, although this is an expensive method [43]. In this context, the behaviour of the setup can be investigated when the nitrogen is the limited nutrient [44].

A steady-state model was used in this approach. Innovative improvements can be done in the model development to characterize the bioreaction. One idea could be to create a more sophisticated adaptive soft sensor where the performance is improved by a change or adaption of the model in real-time. Furthermore, a dynamic structured balance model can be implemented in combination with kinetics to describe rate expressions as a function of the state variables. Literature was found concerning to [45] [46] [47]. It is also possible to use different models in a parallel mode.

Finally, this soft sensor concept can be used for the calibration of other direct online measurement methods or as an additional reference measurement for an improved biomass amount determination.

9 Conclusion

Challenge I

Soft sensor concepts for the estimation of variables of interest are widespread in industrial applications. This technology is rarely in laboratory scale cultivation experiments.

Goal I: The first goal of this work is to find the problems as to why it is uncommon to use a soft sensor technology for estimating the biomass during cultivation runs. Furthermore, an identification of the "issues" in the technical, handling and software section should provide necessary information. In this context, a developed strategy should avoid the "issues". This should be a basis for a successful soft sensor development.

Conclusion I: The use of soft sensor technologies in laboratory environments requires some expertise and can be an inhibitor for an application. Research has shown that processes are often performed outside a design space. The problem is now avoided by a developed fed batch simulation tool. This tool is designed for a pre-estimation of the design space of an existing setup and process.

Challenge II

Specific feeding profiles are related to the biomass concentration in the reactor. The amount of biomass is not available in real-time due to the lack of reliable and accurate biomass quantification results from online direct determinations.

Goal II: A soft sensor concept should be developed to estimate the biomass amount with easily accessible measurements. The provided biomass determination should satisfy the accuracy of a specific substrate uptake rate control strategy. This concept should be developed in a multi-paradigm numerical computing environment. Finally, a proof of this novel soft sensor approach must be done.

Conclusion II: Control concepts for the dissolved oxygen concentration and feeding profile have been developed to ensure a stable fermentation. These shareable tools are implemented in Lucullus PIMS to increase the process robustness. Based on the real-time communication to MATLAB, a novel soft sensor for a robust biomass estimation was implemented. This novel approach is based on the real - time consideration of the measurement device errors. The error propagation consideration only then provides a robust rate reconciliation procedure. The implementation is based on modular functions, which allow fast adaptions to other processes. The soft sensor was validated according to the pre-defined acceptance criteria. Biomass estimation errors around 5 % were reached.

Challenge III

A soft sensor concept should be reliable, robust and easily adaptable to other processes in a bioprocess development environment. In this context, no works investigate the behaviour of the applied soft sensor with the in real-time considered measurement errors.

Goal III: A detailed analysis of the error propagation should be done for a powerful reconciliation procedure, a robust estimation of the biomass and a meaningful statistical test. Finally, a workflow should be established for a successful biomass estimation via this robust soft sensor concept.

Conclusion III: The key novelty of this thesis is the real-time consideration of the changing errors on the rates. The errors on the rates were taken into account by an error propagation and noise error consideration of the measurements for a powerful reconciliation procedure. Furthermore, the rates were calculated with a regression analysis. This led to a very sensitive statistical test at low rates. The used N-balance could be identified as non-representative in this approach. A robust estimation of the biomass could be achieved with the K2S1 model (C and DoR balance). It was found out that the C-balance is the most significant balance for the biomass estimation after the reconciliation procedure. In general, a stable process is essential for a proper biomass estimation. The developed workflow for the biomass estimation therefore forms a basis for improved process developments.

10 Appendix

The program parts in addition with the generated source code of the soft sensor concept and further technical documents are enclosed in digital form.

11 Bibliography

[1] B. Sonnleitner, G. Locher, and A. Fiechter, “Biomass determination,” *J. Biotechnol.*, vol. 25, no. 1, pp. 5–22, Aug. 1992.

[2] P. Wechselberger, P. Sagmeister, and C. Herwig, “Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes,” *Bioprocess Biosyst. Eng.*, vol. 36, no. 9, pp. 1205–1218, 2013.

[3] C. Bittner, G. Wehnert, and T. Scheper, “In situ microscopy for on-line determination of biomass,” *Biotechnol. Bioeng.*, vol. 60, no. 1, pp. 24–35, 1998.

[4] K. Kiviharju, K. Salonen, U. Moilanen, and T. Eerikäinen, “Biomass measurement online: the performance of in situ measurements and software sensors,” *J. Ind. Microbiol. Biotechnol.*, vol. 35, no. 7, pp. 657–665, Jul. 2008.

[5] A. E. Cervera, N. Petersen, A. E. Lantz, A. Larsen, and K. V. Gernaey, “Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation,” *Biotechnol. Prog.*, vol. 25, no. 6, pp. 1561–1581, 2009.

[6] K. Kiviharju, K. Salonen, U. Moilanen, E. Meskanen, M. Leisola, and T. Eerikäinen, “On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes,” *J. Ind. Microbiol. Biotechnol.*, vol. 34, no. 8, pp. 561–566, Aug. 2007.

[7] I. Knabben, L. Regestein, J. Schauf, S. Steinbusch, and J. Büchs, “Linear Correlation between Online Capacitance and Offline Biomass Measurement up to High Cell Densities in Escherichia coli Fermentations in a Pilot-Scale Pressurized Bioreactor,” *J. Microbiol. Biotechnol.*, vol. 21, pp. 204–11, Feb. 2011.

[8] P. Wechselberger, P. Sagmeister, and C. Herwig, “Model-based analysis on the extractability of information from data in dynamic fed-batch experiments,” *Biotechnol. Prog.*, vol. 29, no. 1, pp. 285–296, Jan. 2013.

[9] P. Wechselberger and C. Herwig, “Model-based analysis on the relationship of signal quality to real-time extraction of information in bioprocesses,” *Biotechnol. Prog.*, vol. 28, no. 1, pp. 265–275, Jan. 2012.

[10] V. Steinwandter, T. Zahel, P. Sagmeister, and C. Herwig, “Propagation of measurement accuracy to biomass soft-sensor estimation and control quality,” *Anal. Bioanal. Chem.*, vol. 409, Jul. 2016.

[11] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, T. Langemann, and C. Herwig, “Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development,” *Chem. Eng. Sci.*, vol. 96, pp. 190–198, Jun. 2013.

[12] R. T. van der Heijden, J. J. Heijnen, C. Hellinga, B. Romein, and K. C. Luyben, “Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates,” *Biotechnol. Bioeng.*, vol. 43, no. 1, pp. 3–10, Jan. 1994.

[13] R. T. J. M. van der Heijden, B. Romein, J. J. Heijnen, C. Hellinga, and K. C. A. M. Luyben, “Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors,” *Biotechnol. Bioeng.*, vol. 43, no. 1, pp. 11–20, Jan. 1994.

[14] N. S. Wang and G. Stephanopoulos, “Application of macroscopic balances to the identification of gross measurement errors,” *Biotechnol. Bioeng.*, vol. 25, no. 9, pp. 2177–2208, Sep. 1983.

[15] F. Madron, V. Veverka, and V. Vaněček, “Statistical analysis of material balance of a chemical reactor,” *AIChE J.*, vol. 23, no. 4, pp. 482–486, Jul. 1977.

[16] C. T. Goudar, R. K. Biener, J. M. Piret, and K. B. Konstantinov, “Metabolic Flux Estimation in Mammalian Cell Cultures,” in *Animal Cell Biotechnology: Methods and Protocols*, R. Pörtner, Ed. Totowa, NJ: Humana Press, 2014, pp. 193–209.

[17] J. S. Alford, “Bioprocess control: Advances and challenges,” *Comput. Chem. Eng.*, vol. 30, no. 10, pp. 1464–1475, Sep. 2006.

[18] D. J. Korz, U. Rinas, K. Hellmuth, E. A. Sanders, and W.-D. Deckwer, “Simple fed-batch technique for high cell density cultivation of Escherichia coli,” *J. Biotechnol.*, vol. 39, no. 1, pp. 59–65, Feb. 1995.

[19] P. Wechselberger, P. Sagmeister, and C. Herwig, “Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes,” *Bioprocess Biosyst. Eng.*, vol. 36, no. 9, pp. 1205–1218, Sep. 2013.
[20] R. T. van der Heijden, J. J. Heijnen, C. Hellinga, B. Romein, and K. C. Luyben, “Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates,” *Biotechnol. Bioeng.*, vol. 43, no. 1, pp. 3–10, Jan. 1994.
[21] R. T. J. M. van der Heijden, B. Romein, J. J. Heijnen, C. Hellinga, and K. C. A. M. Luyben, “Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors,” *Biotechnol. Bioeng.*, vol. 43, no. 1, pp. 11–20, Jan. 1994.
[22] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, T. Langemann, and C. Herwig, “Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development,” *Chem. Eng. Sci.*, vol. 96, pp. 190–198, Jun. 2013.
[23] C. T. Goudar, R. K. Biener, J. M. Piret, and K. B. Konstantinov, “Metabolic Flux Estimation in Mammalian Cell Cultures,” in *Animal Cell Biotechnology: Methods and Protocols*, R. Pörtner, Ed. Totowa, NJ: Humana Press, 2014, pp. 193–209.
[24] N. S. Wang and G. Stephanopoulos, “Application of macroscopic balances to the identification of gross measurement errors,” *Biotechnol. Bioeng.*, vol. 25, no. 9, pp. 2177–2208, Sep. 1983.
[25] F. Madron, V. Veverka, and V. Vaněček, “Statistical analysis of material balance of a chemical reactor,” *AIChE J.*, vol. 23, no. 4, pp. 482–486, Jul. 1977.
[26] H. Chmiel, Ed., *Bioprozesstechnik*. Heidelberg: Spektrum Akademischer Verlag, 2011.
[27] C. Ratledge and B. Kristiansen, *Basic Biotechnology*. Cambridge: Cambridge University Press, 2006.
[28] P. Wechselberger, P. Sagmeister, and C. Herwig, “Model-based analysis on the extractability of information from data in dynamic fed-batch experiments,” *Biotechnol. Prog.*, vol. 29, no. 1, pp. 285–296, Jan. 2013.
[29] V. Steinwandter, T. Zahel, P. Sagmeister, and C. Herwig, “Propagation of measurement accuracy to biomass soft-sensor estimation and control quality,” *Anal. Bioanal. Chem.*, vol. 409, Jul. 2016.
[30] B. Sonnleitner and O. Käppeli, “Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis,” *Biotechnol. Bioeng.*, vol. 28, no. 6, pp. 927–937, Jun. 1986.
[31] H. Tong and D. Bluck, “An Industrial Application of Principal Component Test to Fault Detection and Identification,” *IFAC Proc. Vol.*, vol. 31, no. 10, pp. 201–206, Jun. 1998.
[32] A. C. Tamhane, “A note on the use of residuals for detecting an outlier in linear regression,” *Biometrika*, vol. 69, no. 2, pp. 488–489, Aug. 1982.
[33] S. Narasimhan and R. S. H. Mah, “Generalized likelihood ratio method for gross error identification,” *AIChE J.*, vol. 33, no. 9, pp. 1514–1521, Sep. 1987.
[34] J. A. Romagnoli and M. C. Sánchez, Eds., “7 Treatment of gross errors,” in *Process Systems Engineering*, vol. 2, Academic Press, 1999, pp. 109–135.
[35] F. Madron, “A new approach to the identification of gross errors in chemical engineering measurements,” *Chem. Eng. Sci.*, vol. 40, no. 10, pp. 1855–1860, Jan. 1985.
[36] K. Otterstedt *et al.*, “Switching the mode of metabolism in the yeast Saccharomyces cerevisiae,” *EMBO Rep.*, vol. 5, no. 5, pp. 532–537, May 2004.
[37] F. Rodrigues, P. Ludovico, and C. Leão, “Sugar Metabolism in Yeasts: an Overview of Aerobic and Anaerobic Glucose Catabolism,” in *The Yeast Handbook*, 2006, pp. 101–121.
[38] B. Sonnleitner, G. Locher, and A. Fiechter, “Biomass determination,” *J. Biotechnol.*, vol. 25, no. 1, pp. 5–22, Aug. 1992.
[39] P. Wechselberger and C. Herwig, “Model-based analysis on the relationship of signal quality to real-time extraction of information in bioprocesses,” *Biotechnol. Prog.*, vol. 28, no. 1, pp. 265–275, Jan. 2012.
[40] T. Chattaway, A. L. Demain, and G. Stephanopoulos, “Use of Various Measurements for Biomass Estimation,” *Biotechnol. Prog.*, vol. 8, no. 1, pp. 81–84, 1992.

[41] D. Paulsson, R. Gustavsson, and C.-F. Mandenius, “A Soft Sensor for Bioprocess Control Based on Sequential Filtering of Metabolic Heat Signals,” *Sensors*, vol. 14, no. 10, pp. 17864–17882, Oct. 2014.
[42] A. Vicente, J. I. Castrillo, J. A. Teixeira, and U. Ugalde, “On-line estimation of biomass through pH control analysis in aerobic yeast fermentation systems,” *Biotechnol. Bioeng.*, vol. 58, no. 4, pp. 445–450, May 1998.
[43] P. Harms, Y. Kostov, and G. Rao, “Bioprocess monitoring,” *Curr. Opin. Biotechnol.*, vol. 13, no. 2, pp. 124–127, Apr. 2002.
[44] A. Bren, Y. Hart, E. Dekel, D. Koster, and U. Alon, “The last generation of bacterial growth in limiting nutrient,” *BMC Syst. Biol.*, vol. 7, p. 27, Mar. 2013.
[45] M. R. Warnes, J. Glassey, G. A. Montague, and B. Kara, “On data-based modelling techniques for fermentation processes,” *Process Biochem.*, vol. 31, no. 2, pp. 147–155, Jan. 1996.
[46] J. Schubert, R. Simutis, M. Dors, I. Havlik, and A. Lübbert, “Bioprocess optimization and control: Application of hybrid modelling,” *J. Biotechnol.*, vol. 35, no. 1, pp. 51–68, Jun. 1994.
[47] M. von Stosch, R. Oliveira, J. Peres, and S. Feyo de Azevedo, “Hybrid semi-parametric modeling in process systems engineering: Past, present and future,” *Comput. Chem. Eng.*, vol. 60, pp. 86–101, Jan. 2014.