High-throughput screening of optimal process conditions using model predictive control

Niels Krausch[a], ORCID: 0000-0003-2325-6001

Jong Woo Kim[a], ORCID: 0000-0002-4466-8745

Tilman Barz[a,b], ORCID: 0000-0001-7289-3627

Sergio Lucia[c], ORCID: 0000-0002-3347-5593

Sebastian Groß[d], ORCID: 0000-0002-5346-869X

Matthias C. Huber[e], ORCID: 0000-0001-9461-4414

Stefan M. Schiller[e], ORCID: 0000-0001-6864-496X

Peter Neubauer[a], ORCID: 0000-0002-1214-9713

Mariano Nicolas Cruz Bournazou[a,f], ORCID: 0000-0001-9461-4414

[a]Technische Universität Berlin, Department of Biotechnology, Chair of Bioprocess Engineering, Berlin, Germany

[b]AIT Austrian Institute of Technology GmbH, Center for Energy, Vienna, Austria

[c]Technische Universität Dortmund, Process Automation Systems, Dortmund, Germany

[d]WEGA Informatik (Deutschland) GmbH, Weil am Rein, Germany

[e]Albert-Ludwigs-University Freiburg, Center for Biosystems Analysis, Freiburg, Germany & Cluster of Excellence livMatS @ FIT, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Freiburg, Germany

[f]DataHow AG, Dübendorf, Switzerland

Corresponding author: Nicolas Cruz Bournazou, n.cruz@datahow.ch

Keywords: high-throughput, model-predictive control, computer-aided process control

Running title: High-throughput screening using MPC

Abstract

Modern biotechnological laboratories are equipped with advanced parallel mini-bioreactor facilities that can perform sophisticated cultivation strategies (e.g. fed-batch or continuous) and generate significant amounts of measurement data. These systems require not only optimal experimental designs that find the best conditions in very large design spaces, but also algorithms that manage to operate a large number of different cultivations in parallel within a well-defined and tightly constrained operating regime. Existing advanced process control algorithms have to be tailored to tackle the specific issues of such facilities such as: a very complex biological system, constant changes in the metabolic activity and phenotypes, shifts of pH and/or temperature, and metabolic switches, e.g. by product induction, to name a few.

In this work we implement a model-predictive control (MPC) approach based framework to demonstrate: 1) the challenges in terms of mathematical model structure, state and parameter estimation, and optimization under highly nonlinear and stiff constraints in biological systems, 2) the adaptations required to enable its application in High Throughput Bioprocess Development (HTBD), and 3) the added value of MPC implementations when operating parallel mini-bioreactors aiming to maximize the biomass concentration while coping with hard constrains on the Dissolved Oxygen Tension profile.

1 Introduction

The production of recombinant proteins using microbial cell factories has increased dramatically over the last decades (Huang et al., 2012). However, finding optimal process conditions remains a challenge, since the number of strains and possible operating conditions to be tested can be very large (Neubauer et al., 2013). Miniaturization and parallelization with high-throughput liquid handling stations (LHS) can significantly increase throughput and has been mostly performed in microwell plates (MWP) (Knepper et al., 2014). However, cultivation in MWPs has several limitations, including inhomogeneous oxygen transfer rates and limitation regarding sampling, which led to a more widespread use of via mini-bioreactors (MBR) to get more reliable results (Hemmerich et al., 2018). Such platforms have successfully been used

for protein expression studies already 15 years ago (Puskeiler et al., 2005) and computational tools for design and operation of parallel experiments were integrated later on. Cruz Bournazou et al. (Cruz Bournazou et al., 2017) have shown that model based re-design of dynamic experiments can significantly improve information content of the experimental results for the identification of parametric models. Since then, the potential of computer aided high throughput cultivations were demonstrated in the context of process characterization (Sawatzki et al., 2018), strain and process characterization (Anane, García, et al., 2019; Anane, Sawatzki, et al., 2019; Hemmerich et al., 2019) or conditional screening of mutants (Hans et al., 2020; Hemmerich et al., 2019). However, a major problem with such small-scale cultivation is that they do not match the conditions in large-scale bioreactors with respect to various process parameters. While there are several methods that can be used for scaling, there are still major differences between laboratory-scale and industrial-scale cultivations regarding inhomogeneous cultivation conditions (Nadal-Rey et al., 2021; Neubauer & Junne, 2016). Hence, it is essential to ensure that the process behavior and the conclusions drawn from it are as close as possible to the actual process, i.e. also reflect the conditions on a larger scale. In particular, the control of the substrate feeding offers a simple way to mirror certain heterogeneous process conditions. E.g. bolus feeding with pulses has shown to be an easy yet powerful approach to model the effect of inhomogeneous mixing in large-scale bioreactors (Anane, Sawatzki, et al., 2019). An important drawback is still the fact that due to the discrete pulse-based feeding of substrate combined with significant sensor delays, it is difficult to maintain the fastest modes within the desired process constrains. The dissolved oxygen tension (DOT) in cultivations with organisms with a high substrate affinity (e.g. *E. coli*) offers a perfect example. Here we find a number of difficulties caused by the fast dynamics of DOT governed by a large oxygen transfer coefficient (kLa), the sudden changes in the oxygen uptake rate dependent on the substrate availability, and the large delay of the single use DOT sensor-spots on the MBRs.

The operation of robotic facilities for HT conditional screening is still regarded as a major challenge (Morschett et al., 2021). The current contribution builds on our previous work, where

we successfully implemented a framework for high-throughput cultivation with conditional screening capabilities in a milliliter scale (Hans et al., 2020). Avoidance of DOT limitation is a crucial part in optimal operation of such devices, since pulse-based feeding typically leads to drastic stress responses and elevated levels of corresponding genes (Schweder et al., 1999) as well as elevated secretion of several unwanted byproducts like acetate and reduced biomass yield (Bylund et al., 1998). As already mentioned, with the pulse-based feeding approach used in this study, violation of this constraint might easily happen. After applying a pulse, the DOT drops sharply. Conventional controllers though, could only act after a pulse had been given and a constraint violation might have already occurred. Hence, a predictive control algorithms like Model Predictive Controllers (MPC) are required to avoid such conditions.

MPC is an advanced control approach based on a dynamic model of the system which computes the control inputs aiming to minimize a given cost function and satisfy predefined constraints (Rawlings et al., 2017). While widely applied in engineering, MPC has only found relatively few applications in bioprocess engineering (see e.g. the comprehensive review by Mears et al., 2017. One of the first (linear) MPC applications was presented by Kovárova-Kovar et al. to maximize product formation (Kovárová-Kovar et al., 2000). Further examples exist for different cases as e.g. slow growing mammalian cells (Ashoori et al., 2009), yeast (Chang et al., 2016) and microbial cultivations (Del Rio-Chanona et al., 2016; Ulonska et al., 2018). Another approach is to perform set point tracking to follow a predefined trajectory (Craven et al., 2014; Zhang & Lennox, 2004).

The main challenges for the application of linear MPC result from the high nonlinearities and dynamics of biological systems (Biegler, 2010). Therefore, in recent years the application of nonlinear MPC (NMPC) has become more and more prominent (Allgöwer et al., 2004). MPC is a powerful approach but is limited by the accuracy of the model and by the data provided to make optimal decisions. In our specific case, i.e. at the early stage of cultivation, the MPC framework should be able to find an optimal feeding trajectory in real-time time despite high the uncertainty on the model parameter estimates and the scarce data on the strains under

investigation. Hence, it is of great importance to have robust adaptive methods that can perform well under these difficult conditions. A common approach is to use the counterpart of MPC, namely moving horizon estimation (MHE), to estimate the parameters of the process while new data is generated (Rawlings et al., 2017). Using MHE for parameter re-estimation has been used for process engineering for some time and various examples can be found in the literature (Hedengren & Eaton, 2017; Hernández Rodríguez et al., 2021). The reader is referred to Elsheikh et al., 2021 for a comprehensive review. However, none of these mentioned approaches have been applied to systems that present the difficulties mentioned before.

We will further discuss in this contribution how we tackled several issues which are commonly faced in these constrained and highly perturbed fed-batch cultivations in MBRs as (i) the discontinuity of the feeding regime, i.e., the bolus type addition of glucose to the reactors; (ii) the delayed response of the most important state variable, the DOT in terms of sensor delay and system delay to the input, which make predictive control essential to avoid constraint violation; (iii) the differences in the dynamics of the timescales of the system of differential equations, particularly regarding growth of biomass and the DOT, where the time dynamics differ by orders of magnitudes and thus lead to a very stiff system; (iv) the different sampling frequencies (high for DOT and low for biomass, glucose, and acetate); and (v) the uncertainty in the parameter values of the model, which are unknown prior to the cultivation and might be only based on rough knowledge about the strains. Thus, in a limited amount of time, the MHE needs to solve the highly nonlinear and non-convex parameter estimation problem with sufficient accuracy for the MPC to compute inputs that guide the real process to the expected results.

We demonstrate the feasibility and added value of such an approach with real experiments aiming to find optimal process conditions for the industrial production of Elastin Like Proteins (ELPs). ELPs are derived from natural tropoelastin and are promising examples of biocompatible, self-assembling and flexible high-performance materials with a great potential for various applications (Huber et al., 2015; Huber et al., 2022; MacEwan & Chilkoti, 2014). In

order to develop specific protein properties, the sequence composition and length must be changed significantly (Schreiber et al., 2019). The choice of the variable amino acid(s) in the repetitive pentapeptide sequence as well as the size of the protein determine the phase transition temperature among other characteristics, encouraging the creation of large clone libraries with different strains (Huber et al., 2014) for which optimal process conditions for production are yet to be identified. Due to the diverse use of individual amino acids at the fourth position of the repeating sequence and a limited set of core amino acids used (especially proline and valine) the optimization of ELP production depends on multiple parameters such as feed strategies and oxygen supply.

2 Materials and Methods

2.1 High throughput bioprocess development facility

Experiments were conducted on our high-throughput bioprocess development platform. The platform comprises two liquid handling stations (Freedom Evo 200, Tecan, Switzerland; Microlab Star, Hamilton, Switzerland), a mini bioreactor system (48 BioReactor, 2mag AG, Munich, Germany) and a Synergy MX microwell plate reader (BioTek Instruments GmbH, Bad Friedrichshall, Germany. The reader is referred to (Haby et al., 2019) for a detailed description of the facility and sampling procedure.

2.2 Strain and cultivation conditions

All experiments were carried out with *E. coli* BL21(DE3), carrying the plasmid pET28-NMBL-eGFP-TEVrec-$(V_2Y)_{15}$-His, expressing a recombinant fusion protein of ELP and EGFP, under the isopropyl-β-D-thiogalactopyranosid (IPTG) inducible *lac*UV5-promoter. Detailed information about the plasmid can be found in Huber (Huber et al., 2014) and Schreiber (Schreiber et al., 2019). All chemicals were purchased from either Roth, VWR or Merck if not stated otherwise. For the first preculture, 10 mL LB medium, containing 16 g L^{-1} tryptone, 10 g L^{-1} yeast extract and 5 g L^{-1} NaCl, were directly inoculated with 100 µL cryostock and cultured in in a 125 mL Ultra Yield flask (Thomson Instrument Company, USA) sealed with an AirOtop

enhanced flask seal (Thomson Instrument Company, USA) for 5 h at 37°C and 200 rpm in an orbital shaker (Adolf Kühner AG, Birsfelden, Switzerland). The second pre-culture was performed with 25 mL EnPresso B (Enpresso GmbH, Berlin, Germany) medium (whose composition is the same as in the main medium used, besides the glucose polymer) with 9 U L^{-1} Reagent A to assure constant glucose release from the polymer in a 250 mL Ultra Yield flask under the same conditions as mentioned before. This allows for continuous glucose release over time and prevents overfeeding. After 12 h, while in exponential growth phase, appropriate volumes of the pre-culture were used to inoculate the MBRs to an OD600 of 0.25. The bioreactor medium consisted as derived from (Glazyrina et al., 2010) of mineral salt medium, containing (per L): 2 g Na_2SO_4, 2.468 g $(NH_4)_2SO_4$, 0.5 g NH_4Cl, 14.6 g K_2HPO_4, 3.6 g $NaH_2PO_4 \times 2\ H_2O$, 1 g $(NH_4)_2$-H-citrate and 1 mL antifoam (Antifoam 204, Sigma). Before inoculation, the medium was supplemented with 2 mL L^{-1} trace elements solution, 2 mL L^{-1} MgSO4 solution (1.0 M) and kanamycin to a final concentration of 50 mg L^{-1}. The trace element solution comprised (per L): 0.5 g $CaCl_2 \times 2\ H_2O$, 0.18 g $ZnSO_4 \times 7\ H_2O$, 0.1 g $MnSO_4 \times H_2O$, 20.1 g Na-EDTA, 16.7 g $FeCl_3 \times 6\ H_2O$, 0.16 g $CuSO_4 \times 5\ H_2O$, 0.18 g $CoCl_2 \times 6\ H_2O$, 0.132 g $Na_2SeO_3 \times 5\ H_2O$, 0.12 g $Na_2MoO_4 \times 2\ H_2O$, 0.725 g $Ni(NO_3)_2 \times 6\ H_2O$. The composition is also derived from (Glazyrina et al., 2010). In all bioreactor cultivations, the initial glucose concentration for the batch phase was 3 g L^{-1}. At the end of the batch phase, indicated by a sharp rise of DOT, the MHE/MPC controller was started to fit the model to recent available data and start calculating an optimal feeding regime. Feeding was performed with pulses every 10 min. This feeding regime exposes the cells to a short time high glucose concentration, so that they consume a lot of oxygen, which in turn could easily lead to violation of the constraint. The pulse feed for the cultivations which were not controlled by MPC was calculated according to (1) and then integrated over the pulse duration to find the volume to be added within a single pulse.

$$F(t) = \frac{\left(\frac{\mu_{set}}{Y_{X/S}} + q_m\right) * X * V}{S_i} * exp(\mu_{set} * t) \quad (1)$$

Where F describes the feed rate over the time t, μ_{set} the specific growth rate, $Y_{X/S}$ the yield coefficient of glucose per biomass, q_m the glucose consumption for maintenance, S_i the glucose concentration in the feed and X as well as V respectively the biomass and volume at the end of the batch phase.

2.3 Sampling / Analytics

Throughout the cultivation, DOT and pH were measured online, using photometric sensors at the bottom of the MBRs. For the other state variables, samples were taken every 20 min from one of the replicate columns and directly inactivated with dried 2 M NaOH in 96 well plates and stored at 4°C until further analysis. After collection of 3 columns of samples, the plate was automatically transferred to the Hamilton robot for at-line analysis to measure OD600, GFP, glucose and acetate concentration. The reader is referred to (Haby et al., 2019) for a detailed description of the analysis process.

2.4 MHE/MPC framework

The strain was cultivated under 6 and 8 different conditions, respectively, in three replicates each consisting of a batch and an exponential fed-batch phase using a total of $n_r = 18$ mini bioreactors. Each one of the bioreactors $r \in R = \{1, \dots, n_r\}$ can be described by the nonlinear dynamics:

$$\dot{x}_r(t) = f(x_r(t), u_r(t), \theta_r)$$
$$x_r(t_0) = x_{0,r} \tag{2}$$

The dynamic states are denoted by the vector of ODEs $\dot{x}_r$ and include biomass, glucose, DOT, product, volume as well as acetate. The control inputs for each mini bioreactor are $u_r \in R^{n_u}$, while $\theta_r \in R^{n_\theta}$ denotes the unknown parameter vector of the reactors and cultivation conditions and $x_{0,r}$ is the initial condition for each reactor. The inputs are applied as bolus-type pulses. This leads to a highly discontinuous operation with jumps in the volume and concentrations. Thus, after each pulse, the concentrations are recalculated based on the previous

concentrations and the pulse volume. Time-series evolution of the denoted states can be described by a system of ordinary differential equations (ODE) and an algebraic equation, a differential-algebraic system of equations (DAE). The ODE system exhibits dynamics in very different timescales, especially regarding biomass growth and DOT, leading to a very stiff system. Since the dynamics of DOT are usually very fast compared to the other dynamics, they can be expressed in a reduced form as an algebraic equation and thereby reduce the stiffness of the system significantly (Duan et al., 2020). Since the actual DOT (DOT_a) can be only measured with a first order delay, the measured DOT (DOT_m) is also considered as a state variable, taking the response time of the sensor into account. The model has 6 differential states, 1 control input and 18 parameters in total. A complete overview about the equations of the macro kinetic growth model and the meaning of the respective parameters can be found in (Kim et al., 2021).

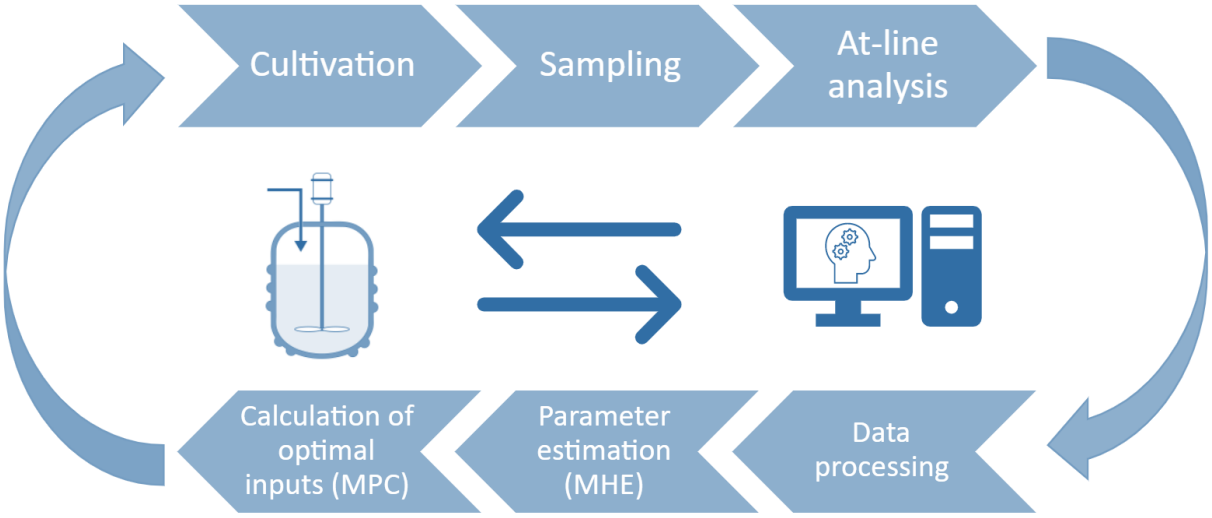

Figure 1: Flowchart of the MHE/MPC framework. During the cultivation, samplings are taken in regular intervals, processed for at-line analysis and used for subsequent parameter estimation and MPC calculations.

The framework used in the study comprised of two parts: a moving horizon estimator for estimating the parameters of the model based on the recent measurements and a model predictive control part, for calculating an optimal feeding profile for each condition. An overview about the workflow is depicted in **Figure *1***. Following this procedure, the parameter set is continuously updated and used for MPC calculations. Considering the N_{MHE} last

measurements, the optimization problem for obtaining a new set of parameters (and in the first iteration also estimates for the initial conditions $x_{0,r}$) can be written as:

$$\min_{\theta, x_{0,r}} \frac{1}{2}\left\|x_{0,r} - x_{0,r,old}\right\|^2_{W_x} + \frac{1}{2}\|\theta - \theta_{old}\|^2_{W_P} + \sum_{k=0}^{N_{MHE}} \frac{1}{2}\|h(x_r(t), u_r(t), \theta) - y_{meas}(t)\|^2_{W_y} \tag{3}$$

s.t.

$$\dot{x}_r(t) = f(x_r(t), u_r(t), \theta) \tag{4}$$

$$\theta_{min} \leq \theta \leq \theta_{max}$$

The estimate for the states at the initial point of the window are then denoted by $x_{0,r}$, the final optimal parameter set by θ and the previous parameter estimate is given by θ_{old}, while the final optimal parameter set is then called $\hat{\theta}$. It further considers the change in the parameters and the difference in the predicted and measured values by the squared norm. Each norm is further weighted by the weighting factors W_x, W_p, W_y. The penalty on the parameter changes assures that the parameters do not drift too much and consider their previous values.

The MPC calculates optimal inputs to maximize biomass at the end of the feeding phase, considering that the DOT should not drop below a predefined threshold of 30%. A detailed description of the MPC and its mathematical formulation can be found in (Kim et al., 2021). The general problem can be written as follows:

$$\min_{u_r} - W_M X_r(t + N_{MPC}\Delta t) - W_L \sum_{k=0}^{N_{MPC}-1} X_r(t + k\Delta t) \tag{5}$$

s.t.

$$\dot{x}_r(t) = f\left(x_r(t), u_r(t), \hat{\theta}\right)$$

$$x_r(t_0) = \hat{x}_{0,r} \tag{6}$$

$$DOT_r(t) \geq z_{lb}, z_{lb} = 30\%, u_r(t) \geq 3$$

Where W_M and W_L denote the weightings for the terminal- and the stage-cost, respectively. $\hat{x}_{0,r}$ refers to the last point of the previous MHE timeframe, which is in turn the first element of the new MHE frame. In every cycle, the MHE fits the model to the recent measured values by

updating the parameter values. With the updated parameters, the MPC is started and calculates new inputs until the end of the feeding phase and beginning of induction. A schematic overview about the workflow is depicted in **Figure *2***.

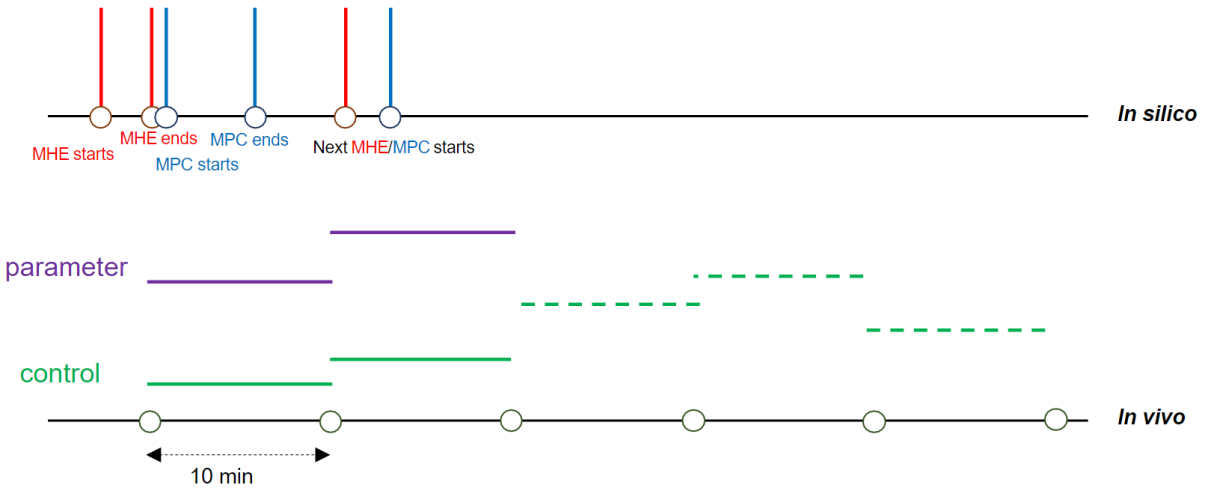

Figure 2: Overview about the MPC workflow. Glucose pulses (the inputs) are given every 10 min as indicated by the circles. The current control inputs for each interval are represented by the green solid lines. Every 10 min, the MHE updates the model parameter (purple lines) by fitting the model to the most recent data. The updated model is used for the MPC to calculate new feeding inputs until induction. The updated inputs are represented by the dashed green lines.

The MHE/MPC framework was compared with a conventional screening approach, which tested the boundaries of the design space to identify optimal cultivation conditions as shown in **Table *1*** (A-D). The values for the growth rates and the respective induction strengths used for the conventional screening approach were identified by initial screening experiments which showed that these ranges might be promising boundaries of the design space and the possible optimum is somewhere in between (data not shown).

Table 1: Overview about the experimental layout. Depicted are the 6 experimental layouts, stating if MPC was applied (+) or not (-) and in case the DOT constraint, the growth rate, and the induction strengths. The first 4 designs comprise the boundaries of the design space and are based on early screening results, while the latter 2 were controlled by MPC.

Exp. setting	MPC (DOT constraint)	μ_{set}	IPTG [mM]
A	-	0.15	0.05
B	-	0.30	0.05
C	-	0.15	2.00
D	-	0.30	2.00

E	+ (30 %)	Controlled by MPC	0.05
F	+ (30 %)	Controlled by MPC	2.00

3 Results

3.1 Coping with sensor delay

The cultivation is fed via bolus feeding of glucose pulses. Even though continuous glucose release systems exist to be used on the robotic platform like EnBase (Krause et al., 2016), these systems would not reach high cell densities and cannot be controlled externally. These high concentrated glucose pulses lead to rapid oxygen uptake of the cells in the process and constraint violation, i.e., oxygen limitation, is easily possible. In such situations it is of great importance to consider the response time of the DOT sensor in the control of the process. This is especially the case when the inverse of the probe response time τ is significantly smaller than the inverse of the k_La ($\frac{1}{\tau} \ll \frac{1}{k_L a}$) (Tribe et al., 1995). This is a major problem in the control of a bioprocess, where avoidance of DOT limitation is a crucial process aspect. Moreover, the DOT drops after a pulse is given, i.e., a change is applied to the system which cannot be reverted. Since the MPC framework is able to predict the actual DOT considering the sensor delay, it was possible to find an optimal feeding regime which satisfies the constraint of the actual DOT not going below 30% or 20% respectively. **Figure *3*** shows the necessity to also simulate the actual DOT, since the measured DOT can be only measured with a delay and will never drop as much as the actual value. Therefore, constraint violation might be possible, even if the sensor shows that there is no oxygen limitation.

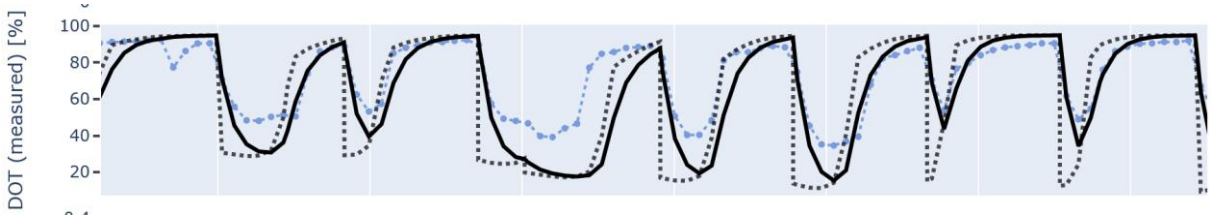

Figure 3: DOT sensor delay. Depicted are the simulated measured DOT (DOTm, solid black line) with the first order delay as well as the simulated actual DOT (DOTa, dashed line) and the real DOT measurements (blue circles and line). As can be seen in the figure, the actual DOT is declining much faster than the measured DOT and could actually violate the constraint if only the measured DOT was considered for control.

3.2 Identifying optimal process conditions and avoiding adverse DOT limitations

Finding optimal process conditions is a significant task in process development of a new biotechnological process. One of the critical parts is to avoid oxygen limitation in a process, since this would lead to a substantial change of the internal metabolism and lead to a significant stress response of the cells (Schweder et al., 1999). To reduce the number of experiments needed to find optimal process conditions while at the same time avoiding anoxic conditions, a HT station was combined with an innovative MPC approach. However, the MPC controller needed to cope with several restrictions in the optimization. Besides the constraint of 30 % oxygen in the reactors, the MPC had to consider that the LHS was not able to pipette less than 3 µL accurately. Under these restrictions it was possible to find a trajectory which satisfies the constrains and leads to high final biomass at the end of the feeding phase. **Figure *4*** shows such a possible optimal feeding trajectory at one iteration step, indicating that the infeasible regions, that is, the regions that do not satisfy some of the constraints, are properly avoided.

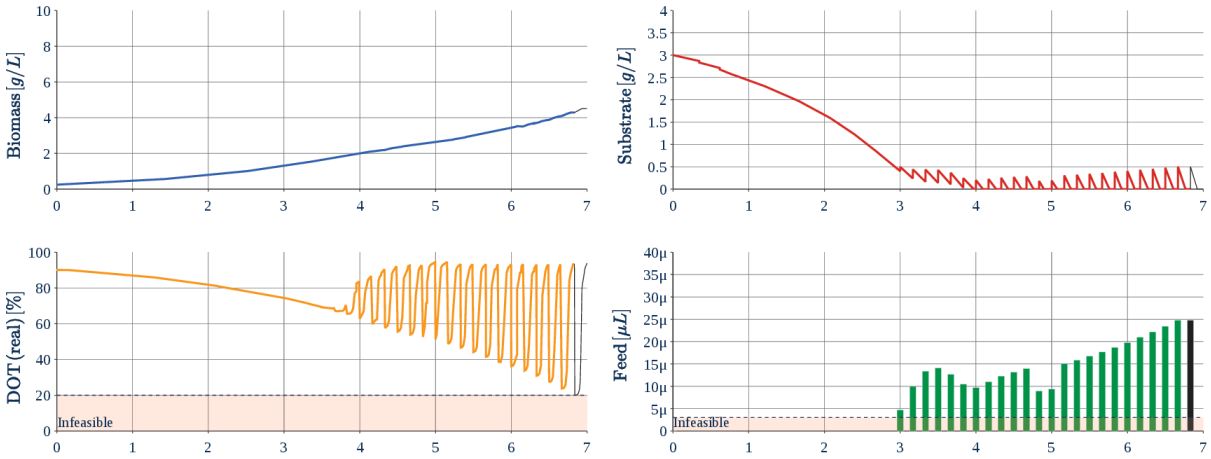

Figure 4: Optimal trajectory avoiding infeasible regions. Shown is a possible trajectory calculated by the MPC framework to obtain high biomass with a pulsed based feeding. Indicated are the infeasible regions as are low levels of oxygen (<20%) or low pipetting volumes (< 3 µL).

The MPC framework had to calculate an optimal feeding profile based on the parameters generated from the data which are measured. A critical step in the framework and for the MPC to work is the estimation of reliable parameter values for the underlying model. If the values, e.g. for maximum substrate uptake capacity were wrong, the MPC would also suggest a wrong feeding regime, which in turn could even lead to constraint violation. The MHE updated the parameter values every 10 min via fitting the model to the most recent 4 h of the process. As

shown in **Figure 5**, there is a good agreement between the model predictions (black solid line) and the measurements from the replicate reactors. While biomass is slightly overestimated by the model, there is good agreement for substrate and the measured DOT signal. Acetic acid is underestimated by the model, especially in the beginning of the feeding phase, but the measured values are still in a low range and the prediction error is small. Underestimating the acetate could lead to wrong predictions of the substrate, since acetate is inhibiting biomass growth.

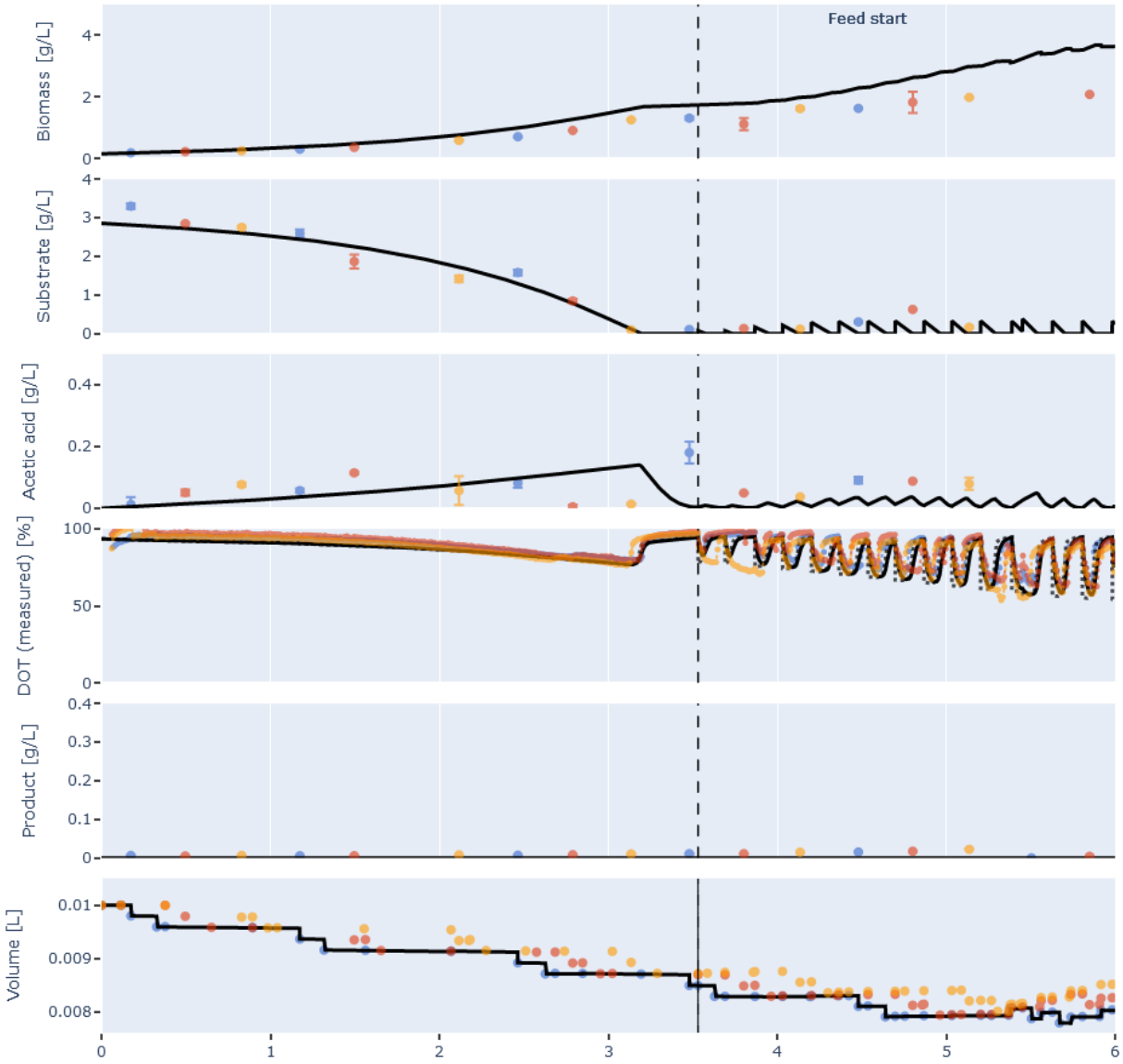

Figure 5: Fitting the model to the data. Shown are the simulated data with the most recent parameters (black sold line) and measurements from 3 replicate reactors (colored dots, each color representing one of the triplicate bioreactors) at a process time of around 6 h.

The MPC controller was able to calculate appropriate inputs for the cultivation. The results of the experiments following the layout of **Table *1*** are depicted in **Figure *6***. After a batch phase of around 4 h, which was indicated by a sharp increase in the DOT signal, the feed was started. Four reactor rows (running in triplicate) were fed with a predefined feed at a μ_{set} of 0.30 h^{-1} or 0.15 h^{-1}, respectively. The other two reactors were fed with individual feeds which were calculated from the MPC controller and updated every 10 min. The reactors with the higher predefined feed rate reached higher biomass values at the end of the process compared to the reactors with the lower feed rate and therefore also higher values for the product concentration as indicated by higher fluorescence as depicted in **Figure *6***. However, especially after induction, the DOT signal falls below the preset threshold of 30 % in those reactors and cells entered overflow metabolism, which is also indicated by glucose accumulation and higher levels of acetate. Induction strength has only minor impact on the production, the cultivations with the higher IPTG concentration showed slightly higher GFP levels normalized to the biomass than the cultivations with lower IPTG. In the reactors which were under control of the MPC framework, the biomass reached comparable levels between the high and the low predefined feedrate, as also indicated by the feeding rate as shown in **Figure *7***. All reactors which were operated using MPC satisfied the constraint of having oxygen levels over 30 %. Glucose accumulation was only observed after induction in those reactors with the high induction level and acetate kept roughly constant during the course of the cultivation. GFP levels were also as high as in the cultivations with the predefined feed. As a result, the biomass obtained was similar to the high μ_{set} but without violating the DOT constrains. This is an increase of approx. 50 % compared to the non-controlled cultivations that stayed within bounds was achieved.

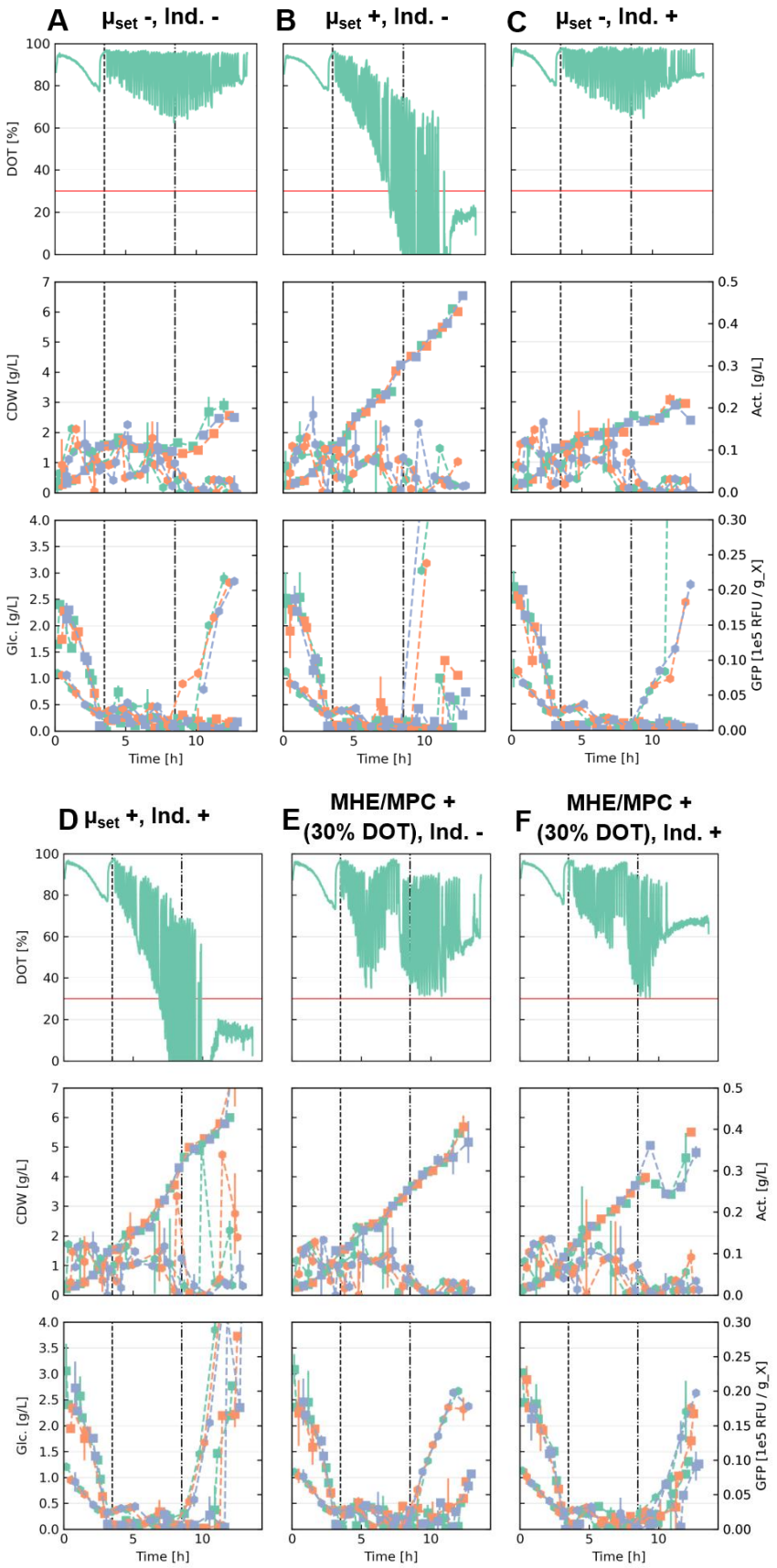

Figure 6: Results from the first cultivation. In the figures A-D are the cultivations depicted with low (0.15) and high (0.3) feeding rate as well with low (0.02 mM) and high (2 mM) induction with IPTG. The part figures E and show the comparison of the processes which are controlled by MPC, again with the low and high induction. Depicted are the measurements for DOT, cell dry weight (squares), glucose (squares), acetate (hexagons) and normalized GFP (hexagons).

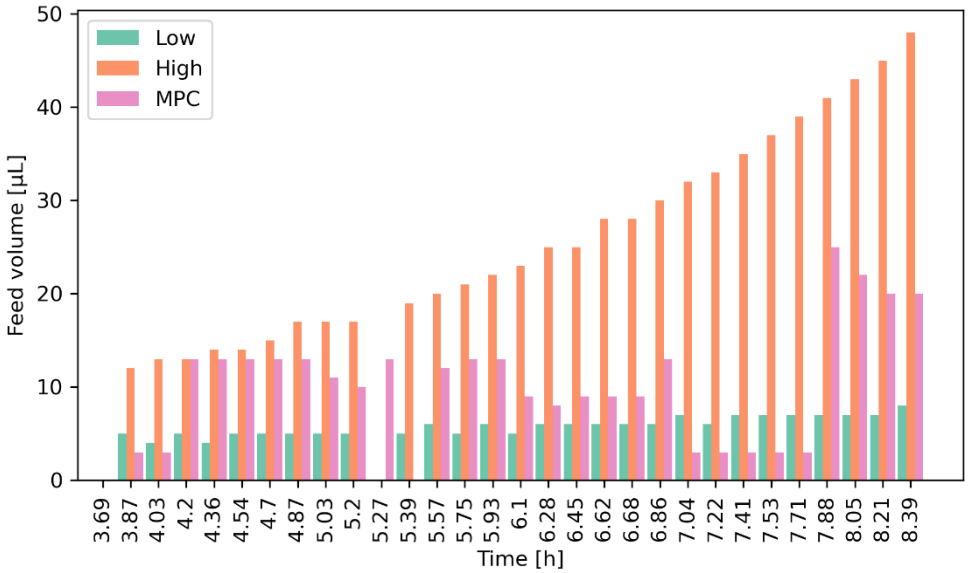

Figure 7: Comparison of the different feeding profiles. Shown are the different pulse volumes at each feeding time during the exponential feeding phase.

4 Discussion

4.1 Successful automated optimal operation of the facility

In this study, we extended our already existing platform for automated high-throughput bioprocess operation with an MHE/MPC framework. The implementation of MHE/MPC now enables automatic optimal operation of the high-throughput facility with only limited a priori knowledge of the process under investigation. Only with the parallel cultivation setup with automated at-line analysis of the most important state variables it is possible to generate enough data to allow a sufficient online model fitting to previously unknown strains or conditions. In contrast to the few previous examples of MPC in bioprocess engineering as mentioned in the introduction, exact parameter values do not need to be known before the experiment but are estimated by the MHE during the process. Hence, in a single run, it was possible to identify possibly optimal cultivation conditions, although some further tuning is necessary to find the optimal trajectory. Moreover, we demonstrate that the control needs to be tackled with an MPC to fulfill the DOT constraints despite the bolus feeding strategy. A classical PID controller would only react after a certain glucose pulse had already been added and a subsequent constraint violation had occurred.

4.2 MHE/MPC guides to optimal process conditions

Operating a high-throughput MBR system is a challenging task and violation of several process constraints might easily happen (Hemmerich et al., 2018). This is especially true when screening a new strain for optimal process conditions, where the biological parameters are unknown before the experiment. The MPC controller successfully managed to maintain the process within the predefined bounds. The approach was compared to a classical approach with predefined feeding rates: Two different feed rates were applied to the process which are often applied in bioprocesses: $\mu_{set} = 0.15$ h^{-1} or $\mu_{set} = 0.3$ h^{-1}, respectively. The lower feed rate did not operate the cultivation at its maximum capabilities. On the other hand, cultivating the cells with the higher feed-rate led to significant oxygen limitation as can be seen in **Figure *6***. An adaptive computation of the optimal profile was necessary to maximize biomass concentration without violating process constraints.

Even though the feeding calculated with the MPC led to significantly better results than with the predefined feed, the optimal feeding profile was not achieved. This is mainly due to plant-model mismatches and inaccuracies of the measurements, which have great influence on the simulation outcome (Anane, López C, et al., 2019). Due to the uncertainties of the parameters which are currently not considered in the MPC, the actual optimal feeding rate could have been higher. This problem could be tackled if for example robust MPC strategies would be applied, which was not of the scope of this work, but will be applied in later stages.

4.3 Dealing with uncertainties

Uncertainties in the model parameter estimates affect the optimization of the MPC. Using an approach which is less sensitive to constraint violation like Multi-Stage MPC could give better results but is also computational more expensive (Lucia et al., 2013). This method could be supported by using a subset selection method with sensitivity analysis or further data-driven approaches like PCA (Thombre et al., 2019). However, only the combination of the MPC framework with the high-throughput cultivation platform is able to generate enough data to obtain good model parameters and ensure an optimal control trajectory. Due to the model

inaccuracies, the optimal feeding trajectory could be even better, but our study shows that avoidance of oxygen limitation is possible while at the same time retaining a high growth rate.

5 Conclusion and outlook

Finding optimal experimental conditions in early bioprocess development is time consuming and laborious. Even though the combination of liquid handling stations and MBR have decreased the bottleneck in the screening phase, it is still not easy to find process conditions which yield high biomass without violating predefined constraints which might be adverse to the process under investigation. However, operating MBRs at their maximum capabilities while fulfilling the constraints is essential for a fast and robust bioprocess development framework. We have described how an MPC approach based on a macro-kinetic growth model can be successful to maintain DOT constraints while maximizing biomass production in the exponential growth phase. Hence, within a single parallel run it is possible to identify close to optimal process conditions. Using an adaptive approach like MHE can alleviate this issue and yield good parameters for the subsequent MPC. However, the current framework is limited by the uncertainties in the parameters. More robust implementations are suggested to deal with these uncertainties to ensure a sufficiently accurate parameter estimation.

Acknowledgements

We thank Felix Fiedler for his contribution to the do-mpc software and Irmgard Schäffl for the early screening of the strains. We further thank labforward GmbH and DataHow AG for their support within the project BioProBot. SMS & MCH gratefully acknowledge the support of the BMBF (31A490 and Research Prize Next Generation of Biotechnological Processes 2014 Biotechnology2020+, 031A550), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2193/1 – 390951807. We further gratefully acknowledge the financial support of the German Federal Ministry of Education and Research (BMBF) (project no. 01QE1957C – BioProBot and 01DD20002A –

KIWI Biolab) and support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

Conflict of Interest

The authors declare that there is no conflict of interests.

References

Allgöwer, F., Findeisen, R., & Nagy, Z. (2004). Nonlinear model predictive control: From theory to application. *Journal of the Chinese Institute of Chemical Engineers*, *35*(3).

Anane, E., García, Á. C., Haby, B., Hans, S., Krausch, N., Krewinkel, M., Hauptmann, P., Neubauer, P [Peter], & Cruz Bournazou, M. N [Mariano Nicolas] (2019). A model-based framework for parallel scale-down fed-batch cultivations in mini-bioreactors for accelerated phenotyping. *Biotechnology and Bioengineering*, *116*(11), 2906–2918. https://doi.org/10.1002/bit.27116

Anane, E., López C, D. C., Barz, T [Tilman], Sin, G [Gurkan], Gernaey, K. V., Neubauer, P [Peter], & Cruz Bournazou, M. N [Mariano Nicolas] (2019). Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. *Biochemical Engineering Journal*, *150*, 107247. https://doi.org/10.1016/j.bej.2019.107247

Anane, E., Sawatzki, A., Neubauer, P [Peter], & Cruz-Bournazou, M. N. (2019). Modelling concentration gradients in fed-batch cultivations of E. coli - towards the flexible design of scale-down experiments. *Journal of Chemical Technology & Biotechnology*, *94*(2), 516–526. https://doi.org/10.1002/jctb.5798

Ashoori, A., Moshiri, B., Khaki-Sedigh, A., & Bakhtiari, M. R. (2009). Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. *Journal of Process Control*, *19*(7), 1162–1173. https://doi.org/10.1016/j.jprocont.2009.03.006

Biegler, L. T. (2010). *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898719383

Bylund, F., Collet, E., Enfors, S.-O [S.-O.], & Larsson, G. (1998). Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Engineering*, *18*(3), 171. https://doi.org/10.1007/s004490050427

Chang, L., Liu, X., & Henson, M. A. (2016). Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. *Journal of Process Control*, *42*, 137–149. https://doi.org/10.1016/j.jprocont.2016.04.012

Craven, S., Whelan, J., & Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *Journal of Process Control*, *24*(4), 344–357. https://doi.org/10.1016/j.jprocont.2014.02.007

Cruz Bournazou, M. N [M. N.], Barz, T [T.], Nickel, D. B., Lopez Cárdenas, D. C., Glauche, F [F.], Knepper, A [A.], & Neubauer, P [P.] (2017). Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. *Biotechnology and Bioengineering*, *114*(3), 610–619. https://doi.org/10.1002/bit.26192

Del Rio-Chanona, E. A., Zhang, D., & Vassiliadis, V. S. (2016). Model-based real-time optimisation of a fed-batch cyanobacterial hydrogen production process using economic model predictive control strategy. *Chemical Engineering Science*, *142*, 289–298. https://doi.org/10.1016/j.ces.2015.11.043

Duan, Z., Wilms, T., Neubauer, P [Peter], Kravaris, C., & Cruz Bournazou, M. N [Mariano Nicolas] (2020). Model reduction of aerobic bioprocess models for efficient simulation. *Chemical Engineering Science*, *217*, 115512. https://doi.org/10.1016/j.ces.2020.115512

Elsheikh, M., Hille, R., Tatulea-Codrean, A., & Krämer, S. (2021). A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Computers & Chemical Engineering*, *146*, 107219. https://doi.org/10.1016/j.compchemeng.2020.107219

Glazyrina, J., Materne, E.-M., Dreher, T., Storm, D., Junne, S., Adams, T., Greller, G., & Neubauer, P [Peter] (2010). High cell density cultivation and recombinant protein production with Escherichia coli in a rocking-motion-type bioreactor. *Microbial Cell Factories*, *9*, 42. https://doi.org/10.1186/1475-2859-9-42

Haby, B., Hans, S., Anane, E., Sawatzki, A., Krausch, N., Neubauer, P [Peter], & Cruz Bournazou, M. N [Mariano Nicolas] (2019). Integrated Robotic Mini Bioreactor Platform for Automated, Parallel Microbial Cultivation With Online Data Handling and Process Control. *SLAS Technology*, *24*(6), 569–582. https://doi.org/10.1177/2472630319860775

Hans, S., Haby, B., Krausch, N., Barz, T [Tilman], Neubauer, P [Peter], & Cruz-Bournazou, M. N. (2020). Automated Conditional Screening of Multiple Escherichia coli Strains in Parallel Adaptive Fed-Batch Cultivations. *Bioengineering*, *7*(4). https://doi.org/10.3390/bioengineering7040145

Hedengren, J. D., & Eaton, A. N. (2017). Overview of estimation methods for industrial dynamic systems. *Optimization and Engineering*, *18*(1), 155–178. https://doi.org/10.1007/s11081-015-9295-9

Hemmerich, J., Noack, S., Wiechert, W., & Oldiges, M. (2018). Microbioreactor Systems for Accelerated Bioprocess Development. *Biotechnology Journal*, *13*(4), e1700141. https://doi.org/10.1002/biot.201700141

Hemmerich, J., Tenhaef, N., Steffens, C., Kappelmann, J., Weiske, M., Reich, S. J., Wiechert, W., Oldiges, M., & Noack, S. (2019). Less Sacrifice, More Insight: Repeated Low-Volume Sampling of Microbioreactor Cultivations Enables Accelerated Deep Phenotyping of Microbial Strain Libraries. *Biotechnology Journal*, *14*(9), e1800428. https://doi.org/10.1002/biot.201800428

Hernández Rodríguez, T., Posch, C., Pörtner, R., & Frahm, B. (2021). Dynamic parameter estimation and prediction over consecutive scales, based on moving horizon estimation: Applied to an industrial cell culture seed train. *Bioprocess and Biosystems Engineering*, *44*(4), 793–808. https://doi.org/10.1007/s00449-020-02488-1

Huang, C.-J., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *Journal of Industrial Microbiology & Biotechnology*, *39*(3), 383–399. https://doi.org/10.1007/s10295-011-1082-9

Huber, M. C., Jonas, U., & Schiller, S. M. (2022). An Autonomous Chemically Fueled Artificial Protein Muscle. *Advanced Intelligent Systems*, 2100189. https://doi.org/10.1002/aisy.202100189

Huber, M. C., Schreiber, A., Olshausen, P. von, Varga, B. R., Kretz, O., Joch, B., Barnert, S., Schubert, R., Eimer, S., Kele, P., & Schiller, S. M. (2015). Designer amphiphilic proteins as building blocks for the intracellular formation of organelle-like compartments. *Nature Materials*, *14*(1), 125–132. https://doi.org/10.1038/NMAT4118

Huber, M. C., Schreiber, A., Wild, W., Benz, K., & Schiller, S. M. (2014). Introducing a combinatorial DNA-toolbox platform constituting defined protein-based biohybrid-materials. *Biomaterials*, *35*(31), 8767–8779. https://doi.org/10.1016/j.biomaterials.2014.06.048

Kim, J. W., Krausch, N., Aizpuru, J., Barz, T [Tilman], Lucia, S., Neubauer, P [Peter], & Bournazou, M. N. C. (2021). Model predictive control and moving horizon estimation for the pulse-based high-throughput cultivation. *Preprint*.

Knepper, A [Andreas], Heiser, M., Glauche, F [Florian], & Neubauer, P [Peter] (2014). Robotic platform for parallelized cultivation and monitoring of microbial growth parameters in microwell plates. *Journal of Laboratory Automation*, *19*(6), 593–601. https://doi.org/10.1177/2211068214547231

Krause, M., Neubauer, A., & Neubauer, P [Peter] (2016). The fed-batch principle for the molecular biology lab: Controlled nutrient diets in ready-made media improve production of recombinant proteins in Escherichia coli. *Microbial Cell Factories*, *15*(1), 110. https://doi.org/10.1186/s12934-016-0513-8

Lucia, S., Finkler, T., & Engell, S. (2013). Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *Journal of Process Control*, *23*(9), 1306–1319. https://doi.org/10.1016/j.jprocont.2013.08.008

MacEwan, S. R., & Chilkoti, A. (2014). Applications of elastin-like polypeptides in drug delivery. *Journal of Controlled Release : Official Journal of the Controlled Release Society*, *190*, 314–330. https://doi.org/10.1016/j.jconrel.2014.06.028

Mears, L., Stocks, S. M., Sin, G [Gürkan], & Gernaey, K. V. (2017). A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *Journal of Biotechnology*, *245*, 34–46. https://doi.org/10.1016/j.jbiotec.2017.01.008

Morschett, H., Tenhaef, N., Hemmerich, J., Herbst, L., Spiertz, M., Dogan, D., Wiechert, W., Noack, S., & Oldiges, M. (2021). Robotic integration enables autonomous operation of laboratory scale stirred tank bioreactors with model-driven process analysis. *Biotechnology and Bioengineering*, *118*(7), 2759–2769. https://doi.org/10.1002/bit.27795

Nadal-Rey, G., McClure, D. D., Kavanagh, J. M., Cornelissen, S., Fletcher, D. F., & Gernaey, K. V. (2021). Understanding gradients in industrial bioreactors. *Biotechnology Advances*, *46*, 107660. https://doi.org/10.1016/j.biotechadv.2020.107660

Neubauer, P [Peter], Cruz, N., Glauche, F [Florian], Junne, S., Knepper, A [Andreas], & Raven, M. (2013). Consistent development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*, *13*(3), 224–238. https://doi.org/10.1002/elsc.201200021

Neubauer, P [Peter], & Junne, S. (2016). Scale-Up and Scale-Down Methodologies for Bioreactors. In C.-F. Mandenius (Ed.), *Bioreactors* (Vol. 13, pp. 323–354). Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527683369.ch11

Puskeiler, R., Kaufmann, K., & Weuster-Botz, D. (2005). Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD). *Biotechnology and Bioengineering*, *89*(5), 512–523. https://doi.org/10.1002/bit.20352

Rawlings, J. B., Mayne, D. Q., & Diehl, M. (2017). *Model predictive control: Theory, computation, and design* (2nd edition). Nob Hill Publishing.

Sawatzki, A., Hans, S., Narayanan, H., Haby, B., Krausch, N., Sokolov, M., Glauche, F [Florian], Riedel, S. L., Neubauer, P [Peter], & Cruz Bournazou, M. N [Mariano Nicolas] (2018). Accelerated Bioprocess Development of Endopolygalacturonase-Production with Saccharomyces cerevisiae Using Multivariate Prediction in a 48 Mini-Bioreactor Automated Platform. *Bioengineering (Basel, Switzerland)*, *5*(4). https://doi.org/10.3390/bioengineering5040101

Schreiber, A., Stühn, L. G., Huber, M. C., Geissinger, S. E., Rao, A., & Schiller, S. M. (2019). Self-Assembly Toolbox of Tailored Supramolecular Architectures Based on an Amphiphilic Protein Library. *Small*, *15*(30), e1900163. https://doi.org/10.1002/smll.201900163

Schweder, T., Krger, E., Xu, B., Jrgen, B., Blomsten, G., Enfors, S.-O [Sven-Olof], & Hecker, M. (1999). Monitoring of genes that respond to process-related stress in large-scale bioprocesses. *Biotechnology and Bioengineering*, *65*(2), 151–159. https://doi.org/10.1002/(SICI)1097-0290(19991020)65:2<151::AID-BIT4>3.0.CO;2-V

Thombre, M., Krishnamoorthy, D., & Jäschke, J. (2019). Data-driven Online Adaptation of the Scenario-tree in Multistage Model Predictive Control. *IFAC-PapersOnLine*, *52*(1), 461–467. https://doi.org/10.1016/j.ifacol.2019.06.105

Tribe, L. A., Briens, C. L., & Margaritis, A. (1995). Determination of the volumetric mass transfer coefficient (kLa) using the dynamic "gas out-gas in" method: Analysis of errors caused by dissolved oxygen probes. *Biotechnology and Bioengineering*, *46*(4), 388–392. https://doi.org/10.1002/bit.260460412

Ulonska, S., Waldschitz, D., Kager, J., & Herwig, C. (2018). Model predictive control in comparison to elemental balance control in an E. coli fed-batch. *Chemical Engineering Science*, *191*, 459–467. https://doi.org/10.1016/j.ces.2018.06.074

Zhang, H., & Lennox, B. (2004). Integrated condition monitoring and control of fed-batch fermentation processes. *Journal of Process Control*, *14*(1), 41–50. https://doi.org/10.1016/S0959-1524(03)00044-1