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]Goethe Universität Frankfurt, Institut für Pharmazeutische Technologie, Frankfurt am Main, Germany

[f]DataHow AG, Dübendorf, Switzerland

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

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 induction of product formation, to name a few.

In this work we implement a model predictive control (MPC) framework to demonstrate: 1) the challenges in terms of mathematical model structure, state and parameter estimation, and optimization under highly nonlinear and stiff dynamics in biological systems, 2) the adaptations required to enable the application of MPC 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

Production of recombinant proteins using microbial cell factories has seen a dramatic increase over the last decades (Huang et al., 2012). However, finding optimal process conditions for the production of a new protein still remains a challenge, since the number of strains and possible operating conditions to be tested can be very large (Neubauer et al., 2013). The introduction of Mini-Bioreactors (MBR), and in particular their combination with liquid handling stations (LHS), have partially alleviated these problems by enabling high-throughput experiments. Especially when combined with modeling and simulation tools, such platforms were successfully applied for model based experimental re-design (Cruz Bournazou et al., 2017),

54 strain and process characterization (Anane, García, et al., 2019; Anane, Sawatzki, et al., 2019;

55 Sawatzki et al., 2018) or conditional screening of mutants (Hans et al., 2020; Hemmerich et

56 al., 2019).

57 However, these systems still have problems when it comes to scale-up, because of the

58 inhomogeneous cultivation conditions in large-scale bioreactors, which are not as pronounced

59 in such small scale systems (working volume < 20 mL) (Nadal-Rey et al., 2021; Neubauer &

60 Junne, 2016). Hence, proper experiments must be designed so that findings are also

61 applicable in a larger scale. The control of the substrate feeding e.g., offers a simple way to

62 mirror certain heterogeneous process conditions. In this regard, bolus feeding with pulses has

63 proven to be a simple but powerful approach to model the effect of inhomogeneous mixing in

64 large-scale bioreactors (Anane, Sawatzki, et al., 2019). Organisms with a high substrate

65 affinity, as e.g. *Escherichia Coli* (*E. coli*), exhibit high oxygen consumption rates. In systems

66 with a small oxygen transfer coefficient (k_La) such as MBRs, this leads to rapid dynamics of

67 DOT changes, causing large obstacles, that are difficult to overcome.

68 Hence, operating such MBR systems using LHS with limited online and at-line measurements

69 available is still considered a major challenge (Morschett et al., 2021). In this respect, the

70 current contribution builds on our previous work, where we successfully implemented a

71 framework for high-throughput cultivation with conditional screening capabilities (Hans et al.,

72 2020). Avoidance of DOT limitation is a crucial part in optimal operation of such devices, since

73 pulse-based feeding typically leads to drastic stress responses, fast changes in DOT due to

74 fast substrate uptake and elevated levels of corresponding genes (Schweder et al., 1999) as

75 well as elevated secretion of several unwanted byproducts like acetate and reduced biomass

76 yield (Bylund et al., 1998). As already mentioned, with the pulse-based feeding approach used

77 in this study, violation of this constraint might easily happen. After applying a pulse, the DOT

78 drops sharply, as the cells start to consume glucose at a high rate. This may even lead to

79 oxygen limitation and to the induction of anaerobic responses (Schweder et al., 1999). Later,

80 after depletion of the glucose, the DOT rises again to the pre-pulse value. In this bolus-feed

81 based setting, conventional (PID) controllers would fail because they can only react after a

glucose pulse has been added and thus a constraint could have been violated shortly afterwards. This is especially true for strongly nonlinear systems like the one presented in this study. Hence, 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 chemical 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 bacterial 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 (Shin et al., 2019). Therefore, in recent years the application of nonlinear MPC (NMPC) has become more and more prominent (Schwenzer et al., 2021). 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 optimal model parameters are not known beforehand 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. The counterpart of MPC, Moving Horizon Estimation (MHE) is a powerful tool to estimate states and parameters of the model and is an excellent complement to MPC (Hille et al., 2020). Using MHE for state and parameter re-estimation has been proposed for process engineering for some time and various examples can be found in the literature (Hedengren & Eaton, 2017; Jabarivelisdeh et al., 2020; Zavala et al., 2008). The reader is further referred to Elsheikh et al., 2021 for a comprehensive review.

109 We will discuss in this contribution how we tackled several issues which are commonly faced

110 in these constrained and highly perturbed fed-batch cultivations in MBRs as (i) the discontinuity

111 of the feeding regime, i.e., the bolus type addition of glucose to the reactors; (ii) system delay

112 to the input, which make predictive control essential to avoid constraint violation; (iii) the

113 differences in the dynamics of the timescales of the system of differential equations, particularly

114 regarding growth of biomass and the DOT, where the time dynamics differ by orders of

115 magnitudes and thus lead to a very stiff system; (iv) the different measurement frequencies

116 (high for DOT and low for biomass, glucose, and acetate); and (v) the uncertainty in the

117 parameter values of the model, which are unknown prior to the cultivation and might be only

118 based on rough knowledge about the strains. Thus, in a limited amount of time, the MHE needs

119 to solve the highly nonlinear and non-convex parameter estimation problem with sufficient

120 accuracy for the MPC to compute inputs that guide the real process to the expected results.

121 To demonstrate the advantages and challenges of our approach, the production of Elastin Like

122 Proteins (ELPs) in *E. coli* was chosen as an interesting case-study. ELPs are derived from

123 natural tropoelastin and are promising examples of biocompatible, self-assembling and flexible

124 high-performance materials with a great potential for various applications (Huber et al., 2015;

125 Huber et al., 2022; MacEwan & Chilkoti, 2014). The properties of the protein depend on the

126 sequence composition, i.e. the amino acids in the repetitive pentapeptide sequence, as well

127 as the length of the protein (Huber et al., 2014; Schreiber et al., 2019). In order to develop

128 specific characteristics, large clone libraries with different strains are created, for which optimal

129 process conditions for production are yet to be identified. Due to the diverse use of individual

130 amino acids at the fourth position of the repeating sequence and a limited set of core amino

131 acids used (especially proline and valine) the optimization of ELP production depends on

132 multiple parameters such as feed strategies and oxygen supply. Therefore, this case-study is

133 highly interesting to test the MHE/MPC framework to find an optimal feeding trajectory without

134 prior knowledge of the strains.

135

2   Materials and Methods

2.1   High throughput bioprocess development facility

All 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 MBRs have a working volume of 8-12 mL and are equipped with fluorometric sensor spots (PreSens Precision Sensing GmbH, Regensburg, Germany) to measure DOT and pH. The LHS performs feeding by adding defined volumes of concentrated glucose solution to the reactors (bolus feeding) in a predefined timeframe. Sampling is automatically performed in regular intervals and the optical density at 600 nm (OD_{600}), fluorescence (as measure for the product concentration) as well as concentrations of glucose and acetate are automatically analyzed at-line on our high-throughput bioprocess development platform. The reader is referred to Haby et al., 2019 for a detailed description of the facility, the sampling and feeding 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). The linkage of the actual target protein, ELP to an eGFP allows a simple non-invasive measurement of the protein concentration. The amount of product, i.e. ELP is calculated based on a conversion factor from the fluorescence measurements which was determined in previous studies. 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 a 125 mL Ultra Yield flask (Thomson Instrument Company, USA) sealed with

164   an AirOtop enhanced flask seal (Thomson Instrument Company, USA) for 5 h at 37°C and 200

165   rpm in an orbital shaker (Adolf Kühner AG, Birsfelden, Switzerland). The second pre-culture

166   was performed with 25 mL EnPresso B (Enpresso GmbH, Berlin, Germany) medium with

167   9 U L^{-1} Reagent A. The composition of the EnPresso B is the same as the main medium used,

168   besides the glucose polymer. This system allows for constant glucose release from the

169   polymer in a fed-batch like manner in a 250 mL Ultra Yield flask, and thus prevents overfeeding

170   even in the preculture. After 12 h, while in exponential growth phase, appropriate volumes of

171   the pre-culture were used to inoculate the MBRs to an OD$_{600}$ of 0.25. The minimal medium in

172   the actual bioreactors consisted as derived from Glazyrina et al., 2010 of mineral salt medium,

173   containing (per L): 2 g Na$_2$SO$_4$, 2.468 g (NH$_4$)$_2$SO$_4$, 0.5 g NH$_4$Cl, 14.6 g K$_2$HPO$_4$, 3.6 g NaH$_2$PO$_4$

174   × 2 H$_2$O, 1 g (NH$_4$)$_2$-H-citrate and 1 mL antifoam (Antifoam 204, Sigma). Before inoculation,

175   the medium was supplemented with 2 mL L^{-1} trace elements solution, 2 mL L^{-1} MgSO4 solution

176   (1.0 M) and kanamycin to a final concentration of 50 mg L^{-1}. The trace element solution

177   comprised (per L): 0.5 g CaCl$_2$ × 2 H$_2$O, 0.18 g ZnSO$_4$ × 7 H$_2$O, 0.1 g MnSO$_4$ × H$_2$O, 20.1 g

178   Na-EDTA, 16.7 g FeCl$_3$ × 6 H$_2$O, 0.16 g CuSO$_4$ × 5 H$_2$O, 0.18 g CoCl$_2$ × 6 H$_2$O, 0.132 g

179   Na$_2$SeO$_3$ × 5 H$_2$O, 0.12 g Na$_2$MoO$_4$ × 2 H$_2$O, 0.725 g Ni(NO$_3$)$_2$ × 6 H$_2$O. In all bioreactor

180   cultivations, the initial glucose concentration for the batch phase was 3 g L^{-1}. At the end of the

181   batch phase, indicated by a sharp rise of DOT, the MHE/MPC controller was started to fit the

182   model to recent available data and start calculating an optimal feeding regime. Feeding was

183   performed by adding glucose pulses (solution with a concentration of 200 g L^{-1} glucose) every

184   10 min by the LHS. This type of feeding exposes the cells to a high glucose concentration for

185   a short time, which is characterized by a steep drop in DOT. After the cells have consumed

186   the glucose, the DOT rises again, resulting in the characteristic oscillating DOT profile. These

187   oscillations come from the fact that DOT drops steeply after the addition of a feeding pulse, as

188   soon as the cells begin to take up glucose. After all glucose is depleted in the pulse period, the

189   DOT rises back to its pre-pulse value. Immediately after the pulse is added, the DOT drops so

190   sharply, that a violation of the constrain of having at least 30 % DOT in the reactors can quickly

191   occur. The pulse feed trajectory for the cultivations which were not controlled by MPC was

192 calculated according to (1) and then integrated over each pulse duration (10 min) to find the

193 volume to be added within a single pulse.

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

194 Here F [L h^{-1}] describes the feed rate over the time t [h], μ_{set} [h^{-1}] the specific growth rate, $Y_{X/S}$

195 [g g^{-1}] the yield coefficient of glucose per biomass, q_m [g g^{-1} h^{-1}] the specific glucose

196 consumption for maintenance (0.02 g g^{-1} h^{-1} were used in this study), S_i [g L^{-1}] the glucose

197 concentration in the feed and X [g L^{-1}] as well as V [L] respectively the biomass concentration

198 and volume at the end of the batch phase. All liquid additions as well as the sampling volumes

199 are stored in the database, so that the current volume and corresponding dilution effects can

200 be always calculated accurately.

201

202 **2.3   Sampling / Analytics**

203 Throughout the cultivation, DOT and pH were measured online, using the photometric sensors

204 at the bottom of the MBRs. Due to the position of the sensors and because the sensors were

205 calibrated under process conditions, gas bubbles in this process do not represent a

206 disturbance of the sensors. For the other state variables, samples were taken every 20 min

207 from one of the replicate columns and directly inactivated with dried 2 M NaOH in 96 well plates

208 and stored at 4°C until further analysis. After collection of 3 columns of samples, the plate was

209 automatically transferred to the Hamilton robot for at-line analysis of OD$_{600}$, fluorescence,

210 glucose, and acetate concentration. The reader is referred to Haby et al., 2019 for a detailed

211 description of the analysis process.

212

213 **2.4   Principles of the MHE/MPC framework**

214 The objective of this study is to find optimal process conditions for strains with little a prior

215 knowledge of their growth behavior. Therefore, a model-based framework was created

216 consisting of an MHE and an MPC part: a moving horizon estimator to estimate the parameters

217 and initial states of the model based on recent measurements; and a model predictive control

218   part to calculate an optimal feeding profile for each condition. Since the strain under

219   investigation was cultivated under 6 different conditions in three replicates, a total of $N_r = 18$

220   mini bioreactors were used. Each of the bioreactors $r \in R = \{1, ..., N_r\}$ can be described by the

221   nonlinear dynamics:

$$\dot{x}_r(t) = f(x_r(t), u_r(t), \theta_r)$$

$$x_r(t_0) = x_{0,r}$$

(2)

222   The dynamic states are denoted by the vector of ODEs $\dot{x}_r$ and include biomass, the substrate

223   glucose, DOT, product (measured via fluorescence), bioreactor volume as well as acetate. The

224   control inputs for each mini bioreactor are $u_r \in R^{N_u}$, while $\theta_r \in R^{N_\theta}$ denotes the unknown

225   parameter vector of the reactors and cultivation conditions and $x_{0,r}$ are the initial conditions for

226   each reactor. The inputs are applied as time-discrete bolus-type pulses. This leads to a highly

227   discontinuous operation with jumps in the volume and concentrations of the other state

228   variables. Thus, after each pulse, the concentrations are recalculated based on the previous

229   concentrations and the pulse volume. The time-series evolution of the denoted states can be

230   described by a system of ordinary differential equations (ODE). The ODE system exhibits

231   dynamics in very different timescales, especially regarding biomass growth and DOT, leading

232   to a very stiff system. Since the dynamics of DOT are usually very fast compared to the other

233   dynamics, they can be expressed in a reduced form as an algebraic equation and thereby

234   reduce the stiffness of the system significantly building a differential-algebraic system of

235   equations (DAE) (Duan et al., 2020). Since the actual DOT (DOT_a) can be only measured with

236   a first order delay, the measured DOT (DOT_m) is also considered as a state variable, taking

237   the response time of the sensor into account. The model has 6 differential states, 1 control

238   input and 18 parameters in total. A complete overview about the equations of the macro kinetic

239   growth model and the meaning of the respective parameters can be found in Kim et al., 2022.
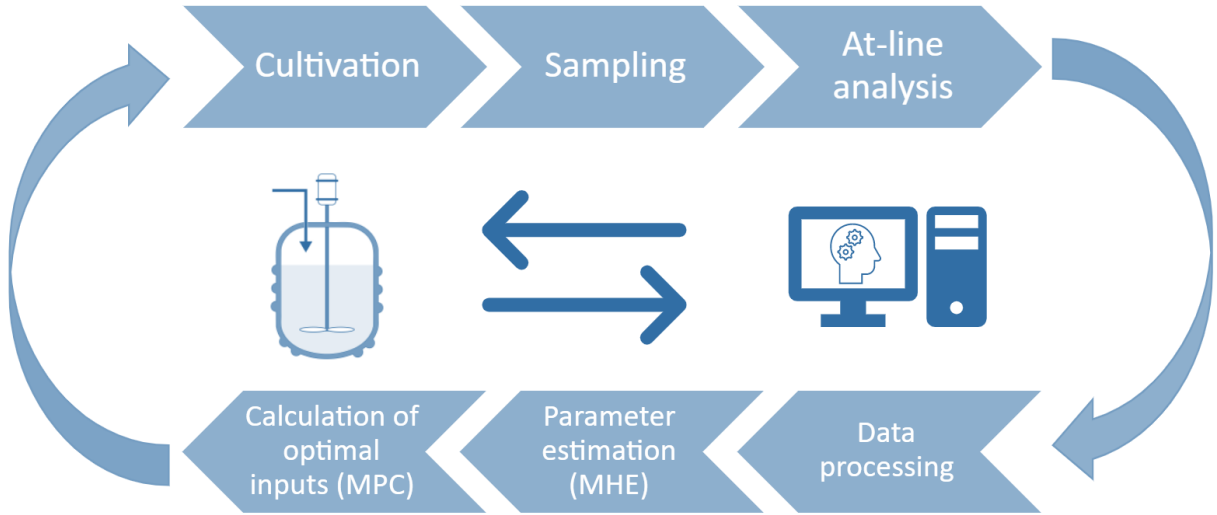
Figure 1: Flowchart of the MHE/MPC framework. During the cultivation, samplings are taken in regular intervals, processed for at-line analysis and used for subsequent parameter estimation and MPC calculations.

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 initial states of the new horizon window can be written as:

$$\min_{\theta, x_{0,r}} \frac{1}{2} \left\| x_{0,r} - x_{0,r,old} \right\|_{W_x}^2 + \frac{1}{2} \| \theta - \theta_{old} \|_{W_P}^2$$

$$+ \sum_{k=0}^{N_{MHE}} \frac{1}{2} \| h(x_r(t), u_r(t), \theta) - y_{meas}(t) \|_{W_y}^2 \tag{3}$$

$$\text{s.t.}$$

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

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

The objective function is composed of the following parts: The estimate for the states at the initial point of the window $x_{0,r}$ and the prior estimate for that state $x_{0,r,old}$ as well as the difference between the current parameter vector θ and the previous parameter estimate vector θ_{old}. The final optimal parameter set is denoted as $\hat{\theta}$. The last term is the summed difference between the predicted outputs $h(\cdot)$ as function of the states $x_r(t)$, the inputs $u_r(t)$ and parameters θ and the available measurements $y_{meas}(t)$. $\|x\|_{W_i}^2 = x^T W_i x$ denotes the squared norm, weighted by the matrix W_i. The subscript r indicates the respective set of MBR

254 replicates. θ_{min} and θ_{max} refer to the lower and upper boundaries of the parameter vector. The

255 penalty on the parameter changes in the objective function $(\theta - \theta_{old})$ assures that, in each

256 iteration, the parameters are not adapted too much considering their previous values.

257 The MPC calculates optimal inputs to maximize biomass at the end of the feeding phase,

258 considering that the DOT should not drop below a predefined threshold of 30 %. A detailed

259 description of the MPC and its mathematical formulation can be found in Kim et al., 2022. The

260 general problem can be written as follows:

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

s.t.

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

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

$$DOT_r(t) \geq 30\ \%, \qquad u_r(t) \geq 3\ \mu L$$

261 The optimization problem is composed of two parts: the terminal cost term (also called Mayer

262 term) $W_M X_r(t + N_{MPC}\varDelta t)$ and the stage-cost term (also called Lagrangian term)

263 $W_L \sum_{k=0}^{N_{MPC}-1} X_r(t + k\varDelta t)$. W_M and W_L denote the weighting matrices for the respective terms.

264 $\hat{x}_{0,r}$ refers to the last point of the previous MHE timeframe, which is in turn the first element of

265 the new MPC frame. X_r is the biomass, which should be maximized in the control horizon N_{MPC}

266 and $\varDelta t$ is the timeframe between two pulses. The system is subject to the constraints of

267 keeping the DOT above 30 % and to pipette at least 3 µL in every pulse. In every cycle, the

268 MHE fits the model to the recent measured values by updating the parameter values and

269 predicting new values for the initial state of the MPC. With the updated parameters, the MPC

270 is started and calculates new inputs until the end of the feeding phase and beginning of

271 induction. By using an efficient nonlinear program solver (IPOPT) and parallelization, the total

272 calculation time for MPC for 24 bioreactors does not exceed the 10 minutes control interval. A

273 schematic overview about the workflow is depicted in Figure 2.
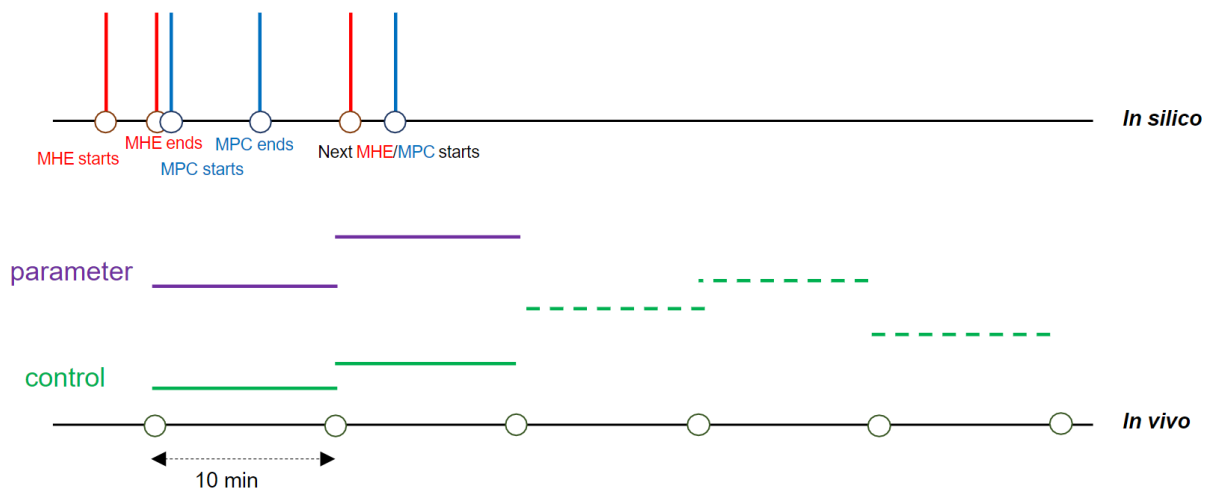
274
275 **Figure 2: Overview about the MPC workflow.** Glucose pulses (the inputs) are given every 10 min as indicated by
276 the circles. The current control inputs for each interval are represented by the green solid lines. Every 10 min, the
277 MHE updates the model parameter (purple lines) by fitting the model to the most recent data. The updated model
278 is used for the MPC to calculate new feeding inputs until induction. The updated inputs are represented by the
279 dashed green lines.

280

281 The MHE/MPC framework implemented using an adapted version of do-mpc (Lucia et al.,

282 2017) was compared with a conventional screening approach, which tested the boundaries of

283 the design space to identify optimal cultivation conditions as shown in Table 1 (A-D). The

284 growth rates and respective induction strengths chosen for the conventional approach are

285 based on initial screening experiments and indicate that a possible optimum is in this range

286 (data not shown).

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

Exp. setting	MPC (DOT constraint)	μ_{set} [h^{-1}]	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

291

3 Results

3.1 Identifying optimal process conditions and avoiding adverse DOT limitations

Finding optimal cultivation conditions is a significant task during the development of a new biotechnological process. Many biotechnological processes depend on aerobic conditions, since oxygen limitation would lead to a substantial change of the internal metabolism and lead to a considerable stress response of the cells (Schweder et al., 1999). To reduce the number of necessary experiments until optimal process conditions are found, our previously available HT cultivation system has been extended by an innovative MPC approach. The MPC framework tries to find an optimal feeding rate according to the last generated data, but it has to take into account that the DOT does not drop below 30 % and the system cannot pipette less than 3 μL. Considering these constraints, an optimal feeding profile was found, which maximizes the biomass at the end. Figure 3 shows such an optimal trajectory at one iteration, where the color-coded constraints were considered.
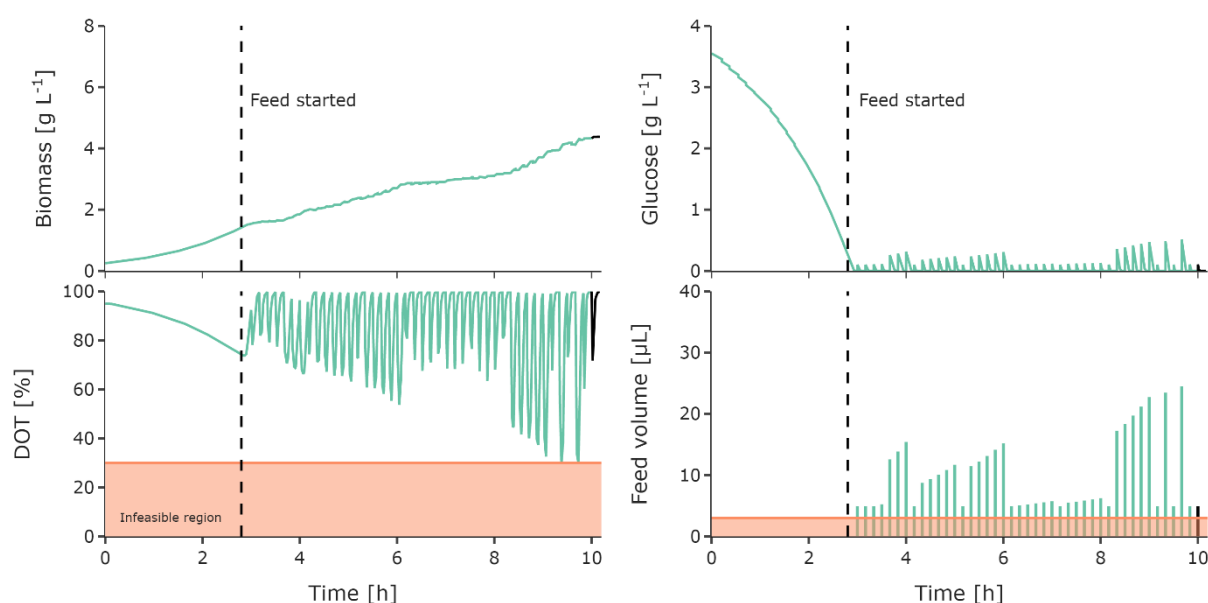
Figure 3: Optimal trajectory avoiding infeasible regions. Shown is a possible trajectory calculated by the MPC framework to obtain high biomass with a pulsed based feeding. Indicated are the infeasible regions (colored areas) as are low levels of oxygen (< 30 %) or low pipetting volumes (< 3 μL). The new suggested input from the MPC is indicated in black.

The MPC framework optimized the feeding trajectory to maximize biomass at the end of the feeding rate while complying with constraints, using the parameters obtained from fitting the model to the data which are measured. Accurate estimates for the parameters of the underlying

12

313 dynamical model are essential to ensure truly optimal inputs for the real process. The MHE

314 updated the parameter values every 10 min via fitting the model to the most recent 4 h of the

315 process. Figure 4 shows a parameter estimation which was performed after the experiment to

316 show the capabilities of the model to describe the data and find good parameter values. This

317 emphasizes that the model and framework used are capable of estimating good parameter

318 values that can be used in the MPC framework to calculate optimal feeding. While biomass is

319 slightly overestimated by the model during the fed-batch phase, there is good agreement for

320 substrate and the measured DOT signal in the batch phase, even though the fitting accuracy

321 deteriorated during the induction phase. Acetic acid is underestimated by the model, especially

322 in the beginning of the feeding phase, but the measured values are still in a low range and the

323 prediction error is small. Underestimating the acetate could lead to wrong predictions of the

324 substrate, since acetate is inhibiting biomass growth.
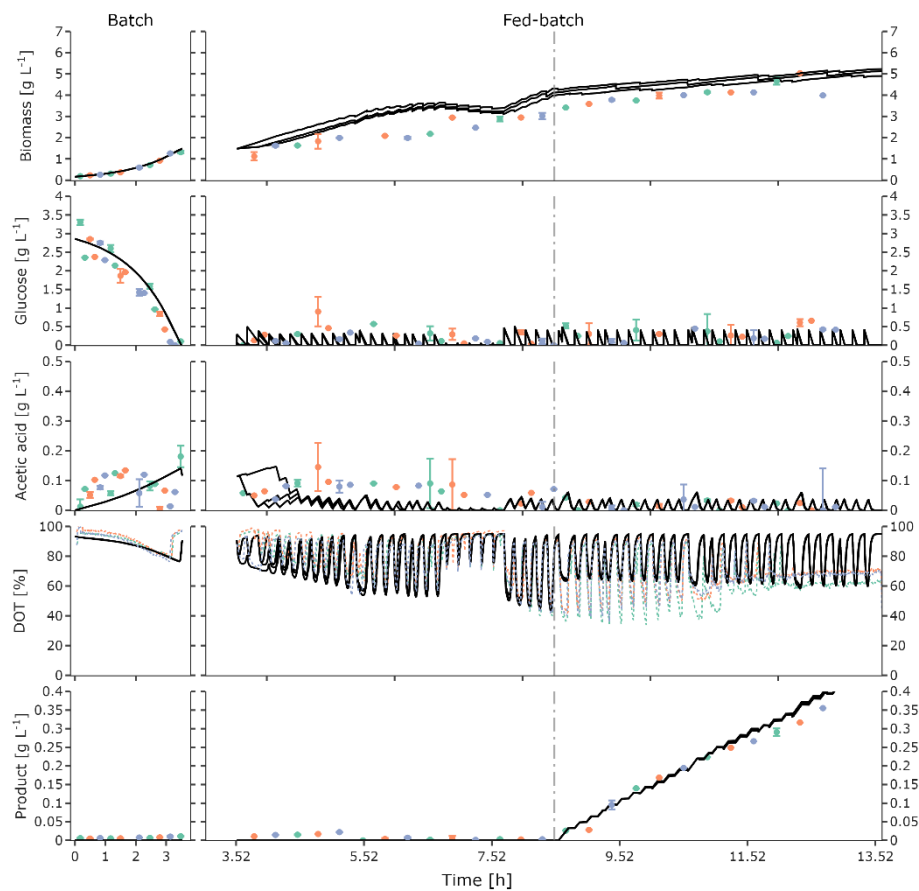
325
326 **Figure 4: *A posteriori* parameter estimation.** Shown is the output of the parameter estimation after the process
327 was performed from 3 replicate reactors (colored dots, each color representing one of the triplicate bioreactors).
328 Note the differences in time scales between batch and fed-batch phases. In this setting, product refers to the ELP-
329 eGFP fusion protein (which was measured via Fluorescence and converted to g L^{-1} with previously calibrated
330 conversion factor). The dash-dotted line at around 8 h indicates the point of induction.

13

331

The MPC framework generated a good feeding trajectory to reach a high biomass, while considering that the constraints are not violated. The results of the experiments following the layout of Table 1 are depicted in Figure 5. After a batch phase of around 4 h, typically detected by the sudden increase in the DOT signal, the feed and MPC controller were started. For the conventional approach, 4 experiments (in triplicates) according to the experiments A-D in Table 1 were fed with a predefined feed at a μ_{set} of 0.30 h^{-1} or 0.15 h^{-1}. The other two experiments (also in triplicates) were fed with individual feeds (Figure 6 E and F) which were calculated from the MPC controller and updated every 10 min. The reactors with the higher feed rate reached higher biomass values at the end of the process compared to the reactors with the lower feed rate (Figure 6 A and C) and therefore also higher values for the product concentration as depicted in Figure 5. However, especially after induction, the DOT signal drops below the 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 product concentration levels normalized to the biomass than the cultivations with lower IPTG. In the reactors, which were controlled by the MPC framework (Figure 6 E and F), the biomass reached comparable levels between the high and the low predefined feeding rate as shown in Figure 6A and B. All reactors which were controlled by the MPC framework 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 remained almost constant during the course of the cultivation. Product concentration 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.

357

14
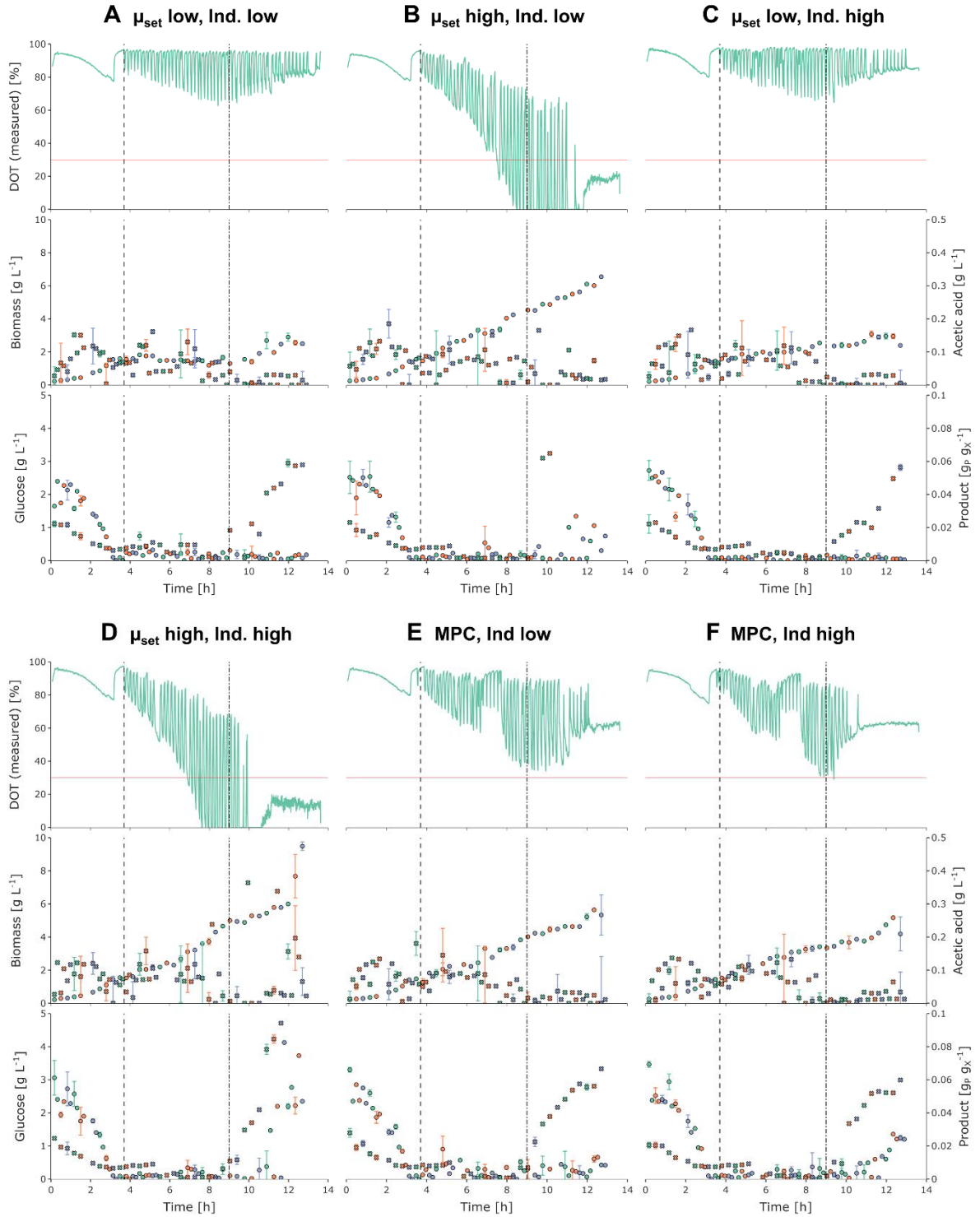
**Figure 5 Results from the first cultivation.** In the figures A-D are the cultivations depicted with low (0.15 h^{-1}) and high (0.3 h^{-1}) feeding rate as well with low (0.02 mM) and high (2 mM) induction with IPTG. The part figures E and F show the comparison of the processes which are controlled by MPC, again with the low and high induction. Depicted are the measurements for measured DOT, biomass (circles), glucose (circles), acetate (x) and product per biomass (x). The dashed vertical line indicates the start of the feed and the dash-dotted vertical line the start of the induction.
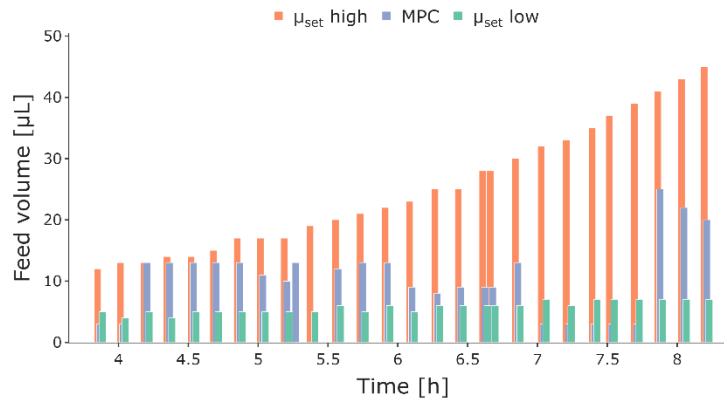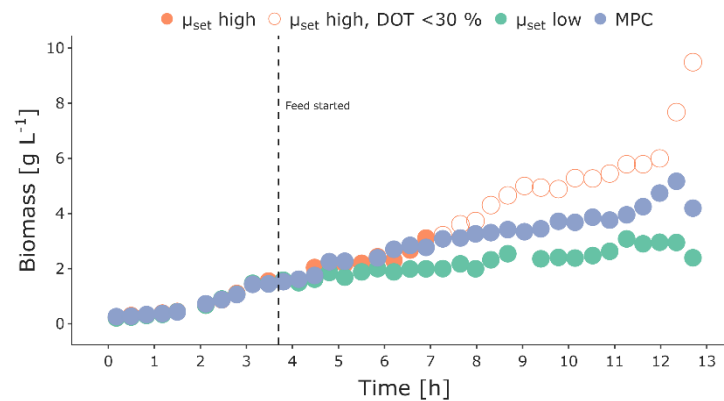
Figure 6: Comparison of the feed profiles and the biomass. Depicted are the different bolus feed volumes at each feeding time during the exponential feeding phase (A) and the measured cell dry weight for the high and the low predefined growth rate as well as for the cultivation which was controlled by the MPC framework (B). The dashed line indicates the start of the feeding. Open circles indicate when a cultivation has violated the constraint of having at least 30 % DOT.

4 Discussion

4.1 Optimal process control with limited a priori knowledge

In this study, we have extended our existing automated high-throughput bioprocess development platform with an MPC framework that allows new *E. coli* strains, about which little prior knowledge is available, to be cultured at their maximum growth capacities. By using online and various at-line measurements, it is possible to measure the key state variables at high frequency and generate sufficient data to fit our mechanistic model of the organism to these data. Unlike previous examples of MPC in bioprocesses, the parameter values of the model do not need to be known in advance, but are adaptively fitted to the measured values during the model (Jabarivelisdeh et al., 2020). This made it possible to determine better cultivation conditions in a single run than would be the case with classical feeding profiles. However, further tuning of the framework is still needed to further optimize the optimal feeding trajectory. Furthermore, we show that this MPC based control of the process is necessary to meet the

16

383 constraints (DOT > 30 %) even though a bolus-based feeding is used. A classical PID

384 controller, on the other hand, could not respond until a glucose pulse was given, which could

385 lead to a violation of the constraint in this system (Santos et al., 2012). In addition, Kager et al.

386 compared stability and performance of a PID controller with MPC and found that the PID

387 controller often cannot cope with the nonlinear dynamics and is unstable, and MPC furthermore

388 achieves better performance. e.g., higher yield (Kager et al., 2020). In addition, a PID controller

389 cannot handle nonlinear process constraints such as oxygen limitation. These constraints can

390 only be met with the help of model knowledge in the form of mathematical optimization.

391

392 **4.2   MHE/MPC guides to optimal process conditions**

393 Operating a high-throughput MBR system is a challenging task and violation of several process

394 constraints might easily happen (Hemmerich et al., 2018). This is especially true when

395 screening a new strain for optimal process conditions, where the biological parameters are

396 unknown before the experiment. The MPC controller successfully managed to maintain the

397 process within the predefined bounds. The approach was compared to a classical approach

398 with predefined feeding rates: Two different feed rates were applied to the process which are

399 often applied in bioprocesses of *E. coli*: μ_{set} = 0.15 h^{-1} or μ_{set} = 0.3 h^{-1}, respectively. The low

400 feeding rate did not achieve the high biomass outputs that would be possible with the strain.

401 On the other hand, cultivating the cells with the higher feed-rate led to significant oxygen

402 limitation as can be seen in Figure 5 and Figure 6. An adaptive computation of the optimal

403 profile was necessary to maximize biomass concentration without violating process

404 constraints.

405 Even though the feeding calculated with the MPC led to significantly better results than with

406 the predefined feed, the optimal feeding profile was not achieved. This is mainly due to plant-

407 model mismatches and inaccuracies of the measurements, which have great influence on the

408 simulation outcome (Nagy & Braatz, 2004). Due to the uncertainties of the parameters which

409 are currently not considered in the nominal MPC, the actual optimal feeding rate could have

410 been higher. Further tuning of the MPC framework, which would make it more aggressive and

411 penalize constraint violation less, could lead to higher yields.

412

4.3 Control under uncertainty

414 In particular, uncertainties inherent in the model as well as uncertainties in the parameters lead

415 to sub-optimal feeding profiles. Especially after induction, the model is less accurate to

416 describe the process. The use of hybrid models could improve model predictions and reduce

417 dependence on individual parameter values (Stosch et al., 2014). However, this requires very

418 large data sets to train such models well. In addition, they are sometimes worse at generalizing

419 for unknown strains. Furthermore, the use of data-driven approaches such as PCA (Thombre

420 et al., 2019) could be supported. In contrast, other approaches in MPC such as multi-stage

421 MPC or stochastic MPC would likely predict more cautious feeding rates so that they do not

422 violate constraints even in the presence of large uncertainties (Lucia et al., 2013).

423

5 Conclusion and outlook

425 Finding optimal experimental conditions in early bioprocess development is time consuming

426 and laborious. Even though the combination of liquid handling stations and MBR have

427 decreased the bottleneck in the screening phase, it is still not easy to find optimal process

428 conditions which yield e.g. high biomass or product concentration without violating predefined

429 constraints which might be adverse to the process under investigation. However, cultivating

430 bacterial strains at their maximum capabilities while fulfilling the constraints is essential for a

431 fast and robust bioprocess development framework. We have described how an MPC

432 approach based on a macro-kinetic growth model can be successful to maintain DOT

433 constraints while maximizing biomass production in the exponential growth phase. Hence,

434 within a single parallel run it is possible to identify close to optimal process conditions. Using

435 an adaptive approach like MHE to estimate states and parameters can support the MPC to

436 deliver optimal control inputs. However, the current framework is limited by the uncertainties

437 in the parameters, the model structure, and the time evolution of the system dynamics. Other

438 implementations are suggested, as e.g. also consider a Kalman Filter, to deal with these

439 uncertainties and plant-model mismatches to ensure a sufficiently accurate parameter

440 estimation and optimal control.

441

453

454 **Conflict of Interest**

455 The authors declare that there is no conflict of interests.

456

457

458

References

Anane, E., García, Á. C., Haby, B., Hans, S., Krausch, N., Krewinkel, M., Hauptmann, P., Neubauer, P., & Cruz Bournazou, M. N. (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., Sawatzki, A., Neubauer, P., & 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

Bylund, F., Collet, E., Enfors, 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., Barz, T., Nickel, D. B., Lopez Cárdenas, D. C., Glauche, F., Knepper, A., & Neubauer, 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., Kravaris, C., & Cruz Bournazou, M. N. (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., Tătulea-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

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

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

498 Hans, S., Haby, B., Krausch, N., Barz, T., Neubauer, P., & Cruz Bournazou, M. N. (2020). Automated Conditional
499   Screening of Multiple Escherichia coli Strains in Parallel Adaptive Fed-Batch Cultivations. *Bioengineering*,
500   *7*(4). https://doi.org/10.3390/bioengineering7040145

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

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

506 Hemmerich, J., Tenhaef, N., Steffens, C., Kappelmann, J., Weiske, M., Reich, S. J., Wiechert, W., Oldiges, M., &
507   Noack, S. (2019). Less Sacrifice, More Insight: Repeated Low-Volume Sampling of Microbioreactor
508   Cultivations Enables Accelerated Deep Phenotyping of Microbial Strain Libraries. *Biotechnology Journal*,
509   *14*(9), e1800428. https://doi.org/10.1002/biot.201800428

510 Hille, R., Brandt, H., Colditz, V., Classen, J., Hebing, L., Langer, M., Kreye, S., Neymann, T., Krämer, S.,
511   Tränkle, J., Brod, H., & Jockwer, A. (2020). Application of Model-based Online Monitoring and Robust
512   Optimizing Control to Fed-Batch Bioprocesses. *IFAC-PapersOnLine*, *53*(2), 16846–16851.
513   https://doi.org/10.1016/j.ifacol.2020.12.1204

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

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

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

523 Huber, M. C., Schreiber, A., Wild, W., Benz, K., & Schiller, S. M. (2014). Introducing a combinatorial DNA-toolbox

524   platform constituting defined protein-based biohybrid-materials. *Biomaterials*, *35*(31), 8767–8779.

525   https://doi.org/10.1016/j.biomaterials.2014.06.048

526 Jabarivelisdeh, B., Carius, L., Findeisen, R., & Waldherr, S. (2020). Adaptive predictive control of bioprocesses with

527   constraint-based modeling and estimation. *Computers & Chemical Engineering*, *135*, 106744.

528   https://doi.org/10.1016/j.compchemeng.2020.106744

529 Kager, J., Tuveri, A., Ulonska, S., Kroll, P., & Herwig, C. (2020). Experimental verification and comparison of model

530   predictive, PID and model inversion control in a Penicillium chrysogenum fed-batch process. *Process*

531   *Biochemistry*, *90*, 1–11. https://doi.org/10.1016/j.procbio.2019.11.023

532 Kim, J. W., Krausch, N., Aizpuru, J., Barz, T., Lucia, S., Neubauer, P., & Cruz Bournazou, M. N. (2022). *Model*

533   *predictive control and moving horizon estimation for adaptive optimal bolus feeding in high-throughput*

534   *cultivation of E. coli.* https://doi.org/10.48550/arXiv.2203.07211

535 Kovárová-Kovar, K., Gehlen, S., Kunze, A., Keller, T., Däniken, R. von, Kolb, M., & van Loon, A. P. (2000).

536   Application of model-predictive control based on artificial neural networks to optimize the fed-batch process

537   for riboflavin production. *Journal of Biotechnology*, *79*(1), 39–52. https://doi.org/10.1016/S0168-

538   1656(00)00211-X

539 Lucia, S., Finkler, T., & Engell, S. (2013). Multi-stage nonlinear model predictive control applied to a semi-batch

540   polymerization reactor under uncertainty. *Journal of Process Control*, *23*(9), 1306–1319.

541   https://doi.org/10.1016/j.jprocont.2013.08.008

542 Lucia, S., Tătulea-Codrean, A., Schoppmeyer, C., & Engell, S. (2017). Rapid development of modular and

543   sustainable nonlinear model predictive control solutions. *Control Engineering Practice*, *60*, 51–62.

544   https://doi.org/10.1016/j.conengprac.2016.12.009

545 MacEwan, S. R., & Chilkoti, A. (2014). Applications of elastin-like polypeptides in drug delivery. *Journal of*

546   *Controlled Release : Official Journal of the Controlled Release Society*, *190*, 314–330.

547   https://doi.org/10.1016/j.jconrel.2014.06.028

548 Mears, L., Stocks, S. M., Sin, G., & Gernaey, K. V. (2017). A review of control strategies for manipulating the feed

549   rate in fed-batch fermentation processes. *Journal of Biotechnology*, *245*, 34–46.

550   https://doi.org/10.1016/j.jbiotec.2017.01.008

551 Morschett, H., Tenhaef, N., Hemmerich, J., Herbst, L., Spiertz, M., Dogan, D., Wiechert, W., Noack, S., &

552   Oldiges, M. (2021). Robotic integration enables autonomous operation of laboratory scale stirred tank

553   bioreactors with model-driven process analysis. *Biotechnology and Bioengineering*, *118*(7), 2759–2769.

554   https://doi.org/10.1002/bit.27795

555   Nadal-Rey, G., McClure, D. D., Kavanagh, J. M., Cornelissen, S., Fletcher, D. F., & Gernaey, K. V. (2021).

556   Understanding gradients in industrial bioreactors. *Biotechnology Advances*, *46*, 107660.

557   https://doi.org/10.1016/j.biotechadv.2020.107660

558   Nagy, Z. K., & Braatz, R. D. (2004). Open-loop and closed-loop robust optimal control of batch processes using

559   distributional and worst-case analysis. *Journal of Process Control*, *14*(4), 411–422.

560   https://doi.org/10.1016/j.jprocont.2003.07.004

561   Neubauer, P., Cruz Bournazou, M. N., Glauche, F., Junne, S., Knepper, A., & Raven, M. (2013). Consistent

562   development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*,

563   *13*(3), 224–238. https://doi.org/10.1002/elsc.201200021

564   Neubauer, P., & Junne, S. (2016). Scale-Up and Scale-Down Methodologies for Bioreactors. In C.-F. Mandenius

565   (Ed.), *Bioreactors* (Vol. 13, pp. 323–354). Wiley-VCH Verlag GmbH & Co. KGaA.

566   https://doi.org/10.1002/9783527683369.ch11

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

568   edition). Nob Hill Publishing.

569   Santos, L. O., Dewasme, L., Coutinho, D., & Wouwer, A. V. (2012). Nonlinear model predictive control of fed-batch

570   cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. *Computers &*

571   *Chemical Engineering*, *39*, 143–151. https://doi.org/10.1016/j.compchemeng.2011.12.010

572   Sawatzki, A., Hans, S., Narayanan, H., Haby, B., Krausch, N., Sokolov, M., Glauche, F., Riedel, S. L.,

573   Neubauer, P., & Cruz Bournazou, M. N. (2018). Accelerated Bioprocess Development of

574   Endopolygalacturonase-Production with Saccharomyces cerevisiae Using Multivariate Prediction in a 48

575   Mini-Bioreactor Automated Platform. *Bioengineering (Basel, Switzerland)*, *5*(4).

576   https://doi.org/10.3390/bioengineering5040101

577   Schreiber, A., Stühn, L. G., Huber, M. C., Geissinger, S. E., Rao, A., & Schiller, S. M. (2019). Self-Assembly

578   Toolbox of Tailored Supramolecular Architectures Based on an Amphiphilic Protein Library. *Small*, *15*(30),

579   e1900163. https://doi.org/10.1002/smll.201900163

580   Schweder, T., Krüger, E., Xu, B., Jürgen, B., Blomsten, G., Enfors, S.-O., & Hecker, M. (1999). Monitoring of genes

581   that respond to process-related stress in large-scale bioprocesses. *Biotechnology and Bioengineering*,

582   *65*(2), 151–159. https://doi.org/10.1002/(SICI)1097-0290(19991020)65:2<151::AID-BIT4>3.0.CO;2-V

583   Schwenzer, M., Ay, M., Bergs, T., & Abel, D. (2021). Review on model predictive control: an engineering

584   perspective. *The International Journal of Advanced Manufacturing Technology*, *117*(5-6), 1327–1349.

585   https://doi.org/10.1007/s00170-021-07682-3

586   Shin, S., Venturelli, O. S., & Zavala, V. M. (2019). Scalable nonlinear programming framework for parameter

587   estimation in dynamic biological system models. *PLoS Computational Biology*, *15*(3), e1006828.

588   https://doi.org/10.1371/journal.pcbi.1006828

589 Stosch, M. von, Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process

590   systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101.

591   https://doi.org/10.1016/j.compchemeng.2013.08.008

592 Thombre, M., Krishnamoorthy, D., & Jäschke, J. (2019). Data-driven Online Adaptation of the Scenario-tree in

593   Multistage   Model   Predictive   Control.   *IFAC-PapersOnLine*,   *52*(1),   461–467.

594   https://doi.org/10.1016/j.ifacol.2019.06.105

595 Ulonska, S., Waldschitz, D., Kager, J., & Herwig, C. (2018). Model predictive control in comparison to elemental

596   balance   control   in   an   E.   coli   fed-batch. *Chemical   Engineering   Science*,   *191*,   459–467.

597   https://doi.org/10.1016/j.ces.2018.06.074

598 Zavala, V. M., Laird, C. D., & Biegler, L. T. (2008). A fast moving horizon estimation algorithm based on nonlinear

599   programming   sensitivity.   *Journal   of   Process   Control*,   *18*(9),   876–884.

600   https://doi.org/10.1016/j.jprocont.2008.06.003

601 Zhang, H., & Lennox, B. (2004). Integrated condition monitoring and control of fed-batch fermentation processes.

602   *Journal of Process Control*, *14*(1), 41–50. https://doi.org/10.1016/S0959-1524(03)00044-1

603