# Direct control of recombinant protein production rates in *E. coli* fed-batch processes by nonlinear feedback linearization

Julian Kager [a,b,*,1], Johanna Bartlechner [c,1], Christoph Herwig [b], Stefan Jakubek [c]

[a] *Competence Center CHASE GmbH, Linz, Austria*
[b] *Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, 1040 Vienna, Austria*
[c] *Institute of Mechanics and Mechatronics, TU Wien, 1040 Vienna, Austria*

ABSTRACT

Biotechnological cultivation processes aim for high and long lasting productivities. In this paper, a method to directly control the productivity of a recombinantly produced protein in an *E. coli* fed-batch process is introduced. After modeling the process using a nonlinear kinetic model, which includes the degeneration of the product formation capacities, a controller was derived by feedback linearization. The derived control law presents an improved and novel approach to directly influence the process parameter associated with the biomass specific product formation rate. In order to deal with observed model-plant mismatches a Two-Degrees-of-Freedom controller was implemented. A simulation study using different model parameters derived from calibration and validation data sets and two realistic measurement scenarios was carried out to demonstrate the potential of the presented method in comparison to a constant substrate addition. Compared to optimized, constant glycerol feed rates the simulations with controlled productivities led to significantly higher specific titers with the same amount of fed glycerol. The feeding rates given by the developed controller minimize the metabolic load as well as product release and therefore stabilizes the productivity for a prolonged, potentially continuous, production process.

## 1. Introduction

Due to their high needs towards nutrients and environmental factors and the fact that deviations often lead to irreversible changes, biological processes need to be tightly controlled. To do so advanced and predictive control strategies, which are based on process models are highly recommended (Smets et al., 2004; Sommeregger et al., 2017).

The integration of the process behavior by a model, yields in a better and preventive control action compared to classical PID-control or open loop control that are still mainly used in biotechnology (Aguilar-Lopez, 2017; Narayanan et al., 2020).

In recent decades advances control schemes were proposed for different bioprocesses (Mears et al., 2017). Among different control approaches, mainly generic model based (Lee and Sullivan, 1988) and model predictive control (Qin and Badgwell, 2003) were analyzed for different organisms

---

including suspension cultures of microbial host organisms (Abadli et al., 2021; Ulonska et al., 2018b), mammalian cells (Dewasme et al., 2015), algae (Yoo et al., 2016) and fungi (Kager et al., 2020).

These biopharmaceutical processes are still mostly carried out in fed-batch operation (Dewasme et al., 2011; Lee, 1996) as this process mode is advantageous both under process control aspects in comparison to batch processes as well as under process stability aspects in comparison to continuous processes (Henson, 2006).

Dewasme et al. (2015), for example used a nonlinear model predictive control (MPC) strategy in mammalian fed-batch cultures to prevent overflow metabolism by optimizing glucose and glutamine addition. Simulation results show that even in the worst case a production drop can be minimized to < 20%. Similarly, Yoo et al. (2016) used a MPC to optimize biomass and lipid concentrations of microalgae cultivations in a photo-bioreactor. Due to the computational cost of the recursive optimization in a MPC, generic model based control evolved as a valuable alternative for processes with faster growth dynamics, such as *E. coli* bacteria. Within experiments Abadli et al. (2021) showed the ability of generic model based control to control both, the growth rate and the acetate concentration.

To this date, advanced and predictive control was mainly used to control the substrate concentration, the substrate uptake or the biomass growth rate (Abadli et al., 2020; de Battista et al., 2006; Dewasme et al., 2011; Henson, 2006; Hocalar and Turker, 2014; Versyck and van Impe, 1998; Dabros et al., 2010) or to prevent byproduct formation (Abadli et al., 2021; Kager et al., 2020), which are only indirectly correlated with the product formation rates. Although, some works report a relation between nutrient addition, growth and product formation (Zalai et al., 2016; Douma et al., 2010; Wunderlich et al., 2014) and were able to increase product content by the control of associated variables, such as growth (Fan et al., 2005) or substrate uptake (Kager et al., 2020; Picon et al., 2005), a dedicated control strategy to directly control the product formation rates is still lacking for microbial fed-batch processes.

For heterologous protein production in *E. coli* or other microbial host organisms, additional metabolic load arises after induction of the recombinantly introduced genes. This metabolic load leads to a decline in the metabolic capabilities of the host organism (Neubauer et al., 2003), which still hampers the establishment of long lasting and stable production phases. A typical trajectory of the cell specific productivity is displayed in Fig. 1 (grey line). After reaching high production levels, the metabolic load causes a continuous and sharp decrease in productivity. To prevent this drop and to keep the cells productive for a longer period of time (Fig. 1; black line), different genetic and process engineering approaches were followed. On the genetic engineering level tunable (Marschall et al., 2016,2017; Neubauer et al., 1992) or weaker promoter systems, such as the rhamnose inducible rhaBAD system (Wilms et al., 2001), were developed. Recently, Schuller et al. (2020) could show stabilizing effects of the production clones by adaptive evolution. On the process level, it could be shown that a temperature reduction (Pinsach et al, 2008) as well as a switch from an exponentially increasing feed profile to a constant feed (Wechselberger et al., 2012) or decreasing feeds (Ramalingam et al., 2007) can lead to improved productivities. From a process control perspective the strategies used for the control and the
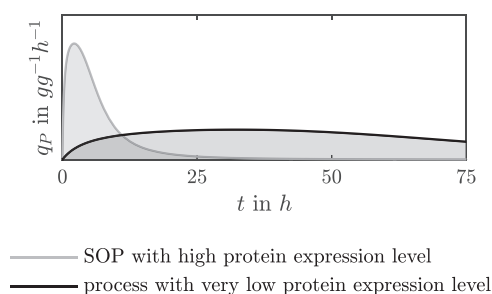


**Fig. 1 – Typical trend of recombinant protein expression level ($q_P$) of an indirectly controlled fed-batch processes (grey line) and a potential process with directly controlled and steady productivity (black line).**

potential prolongation of the production phase of recombinant protein production processes are still very basic (Narayanan et al., 2020; Mears et al., 2017) and, to date, can not fully overcome the problem of production drops due to the additional metabolic load by the forced transcription and translation of the inserted target genes.

Within this study, we tackle this problem on a process control level with the aim to directly control the product formation rates in induced fed-batch processes. In contrast to the above mentioned indirect and static control, direct control enables to define beneficial trajectories of process variables, which potentially lead to long lasting and stable processes as indicated in Fig. 1 (black line).

The control strategy proposed and examined in this paper is, therefore intended to directly control the biomass specific product formation rate of a recombinantly produced protein in an *E. coli* fed-batch process. To do so, the process is modeled with an unstructured growth model with constant parameters, whereas the metabolic load describes the decay of the product formation rate as well as the biomass growth efficiency in the course of the production phase (Neubauer et al., 2003). The parametrized model serves as a basis to derive the control law by the method of nonlinear feedback linearization. The nonlinear feedback linearization ensures the correct incorporation of the nonlinear process dynamics within a linear input-output relationship (Abadli et al., 2020).

A stable feedback control implementation is ensured by a two-degrees-of-freedom (2-DOF) controller (Araki and Taguchi, 2003). This combination of two independent control structures enables a fast reaction to changes in the reference trajectory by a feedforward control term and a stable adjustment of the feed trajectory by the feedback control term, caused by model plant mismatches as well as system disturbances (Baeza, 2017). A similar control structure including a feedback linearized growth model (Abadli et al., 2020), was recently used for an *E. coli* fed-batch to avoid excessive acetate formation by controlling both, the growth rate as well as the acetate concentration (Abadli et al., 2021). The novelty and goal of this study is to directly control the biomass specific protein production rate ($q_P$) by the substrate feed, which to date has not been studied in open literature for recombinant protein production processes. Besides testing the robustness of the novel controller within the calibrated range of operation it was applied on an independent plant model by considering two realistic measurement scenarios.

The paper is organized as follows: In the Material and methods section the experimental data sets, the nonlinear kinetic model and its parametrization as well as the method

of feedback linearization are described. In the Results and discussion section the adequacy of the calibrated model is shown before the control law is derived. After designing an optimal setpoint trajectory two practically relevant measurement scenarios are included into a two-degree-of-freedom controller, which are tested on a plant model, based on a validation data set. The paper is concluded by discussing the robustness and stability of both variants as well as their potential to increase the product amount.

## 2. Material and methods

### 2.1. Calibration and validation data sets

A modified K12 *E. coli* strain was used as host organism. The strain features a rhamnose-inducible, rhaBAD expression system (Wilms et al., 2001). The recombinant protein was a single chain antibody fragment with a transporter sequence to be transported into the periplasm, where under the reductive environment, it can conserve its native formation (Ellis et al., 2017).

Fermentations were conducted in a DASGIP multi-bioreactor system with four parallel reactors with 2.7 L of working volume each (Eppendorf, Germany). The reactors were equipped with baffles and three disk impeller stirrers. Temperature was kept at 35°C, stirrer speed at 1400 rpm and aeration with a L-sparger at 1.4 vvm for the whole process. The pH was controlled at 7.0 with addition of 12.5% $NH_4OH$ solution, which additionally served as nitrogen source. The dissolved oxygen ($DO_2$) was kept over 25% by mixing the pressurized air with pure oxygen.

The pre-culture was incubated at 30°C and 170 rpm for approx. 17 h (OD600 ≈ 1.5). A volume equivalent of pre-culture of 2.5% of the 1.0 L batch volume was used to inoculate the reactors. The exact composition of the used minimal medium can be found in Wilms et al. (2001). After depletion of the C-source (20 g/L glycerol), occurring after approx. 12 h, the pre-induction fed-batch was started. The feed consisted of 70% glycerol with 20 g/L $MgSO_4$ * $4H_2O$ and 4 mL/L trace element solution as given in Wilms et al. (2001). After reaching the predefined biomass concentration of 30 g/L (approx. 9 h with a $\mu$ of 0.14, using a exponential feed forward equation given in (Lee, 1996)), different predefined constant feed rates (*high*= 14 mL/h; *medium* = 10.0 mL/h; *low* = 4.0 mL/h, *very low* = 2.0 mL/h, *validation* = 8.0 mL/h) with the following measured glycerol feed concentrations; *high* = 930 g/L; *medium* = 838 g/L; *low* = 930 g/L, *very low* = 822 g/L, *validation* = 900 g/L were applied. Recombinant protein production was induced by an one-point addition of sterile filtrated rhamnose solution (1.5 g L-rhamnose per 1.0 L batch volume). The processes were stopped after a predefined amount of glycerol (minimum 150 g) was fed to each reactor.

### 2.2. Measurements

Biomass dry weight concentrations were quantified gravimetrically after drying to constant weight (> 72 h) at 105°C. As preparation 2 mL of culture broth were centrifuged (4500 g, for 10 min, 4 °C) in a pre-weighted glass tube and the pellet was washed once with 5 mL normal water. To determine intracellular product concentrations, 2 mL of the fresh culture broth was centrifuged (4500 g, for 10 min, 4°C). The cell pellets were re-suspended in 20 mL of 0.1 M Tris-buffer (pH 7.4) and were disrupted in a high-pressure

homogenizer (Avestin EmulsiFlex; Canada) at 1400 bar in 6 passages. 500 μL of the homogenate were then applied on gel filtration columns (PD MiniTrap g-25, GE Healthcare, USA) and eluted with 1 mL 20 mM phosphate-buffer (pH 7.4). The product titers were measured by a protein G affinity chromatography using a pH gradient on a HPLC system (Thermo Scientific Dionex Ultimate 3000; USA). The column was a HiTrap ProtG (GE Healthcare; USA) with a flow rate of 2 mL/min at 25°C. The detection was at 390 nm and the elution was forced by changing the pH from 7.4 to 2.5 (20 mM phosphate-buffer). In addition, product outside the cell was determined by measuring the filtrated cell broth. Acetate and glycerol concentrations were quantified from the supernatant by enzymatic, photometric principle in a robotic system (BioHT, Roche, Germany). The analysis was used as a quality control to exclude possible acetate formation due to oxygen limitation or glycerol accumulation. Glycerol concentrations in the feed were determined gravimetrically through a feed-density correlation. Under the assumption of a constant rate between two measurement points, the reaction rates were calculated by a simple mass balance approach, where the state change is described by a general material balance including the biomass specific rates. By minimizing the error (fmin: MATLAB) between the balance equation output and the two measurement points, the best constant rate can be determined for the analyzed time-interval (Kroll et al., 2017).

### 2.3. Nonlinear kinetic model

The model of the induction phase consists of five macroscopic reactions:

- Growth on glycerol: $S \xrightarrow{\mu} X$
- Cell maintenance: $X + S \xrightarrow{m_S} X$
- Product formation: $X \xrightarrow{q_P} P$
- Product release: $P \xrightarrow{q_{PR}} P_R$
- Metabolic load: $S \xrightarrow{q_S} S_{met}$

Hereby S, X, P and $P_R$ are the concentrations of glycerol (substrate), biomass, product in the cell (periplasm) and product released to the cell broth. The biomass specific rates governing these macroscopic reactions are the growth rate $\mu$, the glycerol consumption for cell maintenance ($m_S$), the product formation rate ($q_P$) and the product release rate ($q_{PR}$). The metabolic load ($S_{met}$) provoked by the heterologous protein expression was additionally included. The term $S_{met}$ hereby describes the metabolized substrate per biomass content and corresponds to the integral of the biomass specific glycerol uptake rate $q_S$. $S_{met}$ is used as a continuous trigger for changing yields and production rates after induction.

#### 2.3.1. Growth kinetics with changing yields
For substrate uptake a standard Monod kinetic was considered

$$q_S = q_{S,max} \frac{c_S}{c_S + K_S}, \qquad (1)$$

with $q_{S,max}$ being the maximum uptake rate in function of the glycerol concentration $c_S$ and the half saturation constant $K_S$. Biomass formation ($\mu$) from consumed substrate ($q_S$) is linked
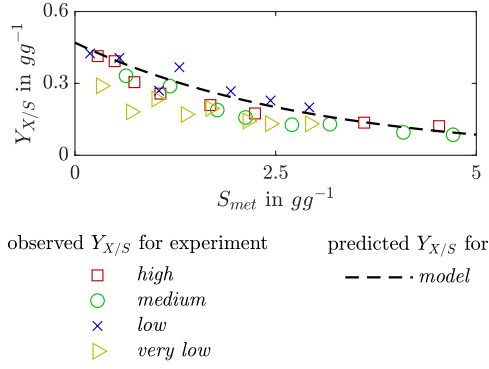
**Fig. 2 – Observed and modeled asymptotic decay of yield coefficient $Y_{X/S}$ as function of metabolic load $S_{met}$.**

via the biomass to substrate conversion yield $Y_{X/S}$ reduced by the substrate needed for cell maintenance $m_S$ as given by:

$$\mu = Y_{X/S}(q_S - m_S). \tag{2}$$

Within Fig. 2 the biomass to substrate yield of the four experiments over the metabolized substrate ($S_{met}$) is shown.

To include the shown yield decay in function of the metabolic load ($S_{met}$), an asymptotic decay was used:

$$Y_{X/S} = Y_{X/S,max} \exp(-S_{met}K_{Y_{X/S}}), \tag{3}$$

where $Y_{X/S,max}$ is the yield at induction time-point, which was determined with data from pre-induction. $K_{X/S}$ is the decay constant, which in function of the metabolized glycerol $S_{met}$ determines the asymptotic decay of $Y_{X/S,max}$.

### 2.3.2. Product formation kinetics

Three main characteristics of product formation kinetics could be deduced from the four experiments by evaluating the trajectories of the biomass specific productivity as well as the maximum productivity (Fig. 3): .

1. a dependency of maximum productivity $q_{P,max}$ and specific glycerol uptake $q_S$
2. a start up phase indicated by a needed amount of metabolized glycerol $S_{met}$ (See Fig. 3; $S_{met} < \approx 0.5$)
3. a decline in productivity in function of overall metabolized glycerol $S_{met}$ (See Fig. 3; $S_{met} > \approx 0.5$)

The dependency of the maximum production rate on the specific substrate uptake rate, as visualized in Fig. 3, includes a hypothesized Monod kinetic with $q_{P,max}$ as the maximum

rate and $K_{Sq_S}$ as the glycerol uptake at $0.5 \cdot q_{P,max}$. To additionally describe the start up and the decline in productivity a Haldane kinetic was included, leading to the following description of the product formation kinetics:

$$q_P = q_{P,max} \frac{q_S}{q_S + K_{Sq_S}} \frac{S_{met}}{\frac{S_{met}^k}{K_{Iq_P}} + S_{met} + K_{Sq_P}}, \tag{4}$$

where the first Monod term describes the dependency of $q_S$ on $q_P$ (Fig. 3, left) and the second Haldane term the underlying start up and decline phase with $S_{met}$ as the time variant trigger, $K_{Sq_P}$ the delay coefficient, $K_{Iq_P}$ the decay coefficient and $k$ the Haldane exponent describing the shape of the decay (Fig. 3, right).

As within the experiments product could be observed both inside and outside the cell, a product release kinetic was added. Under assumption of no active secretion, only growth related product loss due to cell division was included by:

$$q_{PR} = \left(1 - \exp\left(-\frac{S_{met}}{K_{Sps}}\right)\right)p_{rel,max}\frac{q_S - m_S}{K_{q_{S,rel}} + q_S - m_S}, \tag{5}$$

with an activation term $(1 - \exp(-S_{met}/K_{Sps}))$ triggered by $S_{met}$ with the delay coefficient $K_{SpS}$ and a growth associated release rate description with $K_{qS,rel}$ as the half-speed constant and $p_{rel,max}$ the maximum release rate.

### 2.3.3. System differential equations

With the described processes and reaction rates, the system differential equations for the ideally stirred tank reactor in fed-batch mode become

$$\frac{dV_R}{dt} = \dot{V}_{in} \tag{6a}$$

$$\frac{dc_X}{dt} = \mu \, c_X - \frac{\dot{V}_{in}}{V_R} \, c_X \tag{6b}$$

$$\frac{dc_S}{dt} = -q_S \, c_X + \frac{\dot{V}_{in}}{V_R} \, (c_{S,in} - c_S) \tag{6c}$$

$$\frac{dc_P}{dt} = q_P \, c_X - q_{PR} \, c_P - \frac{\dot{V}_{in}}{V_R} \, c_P \tag{6d}$$

$$\frac{dc_{P_R}}{dt} = q_{PR} \, c_P - \frac{\dot{V}_{in}}{V_R} \, c_{P_R} \tag{6e}$$
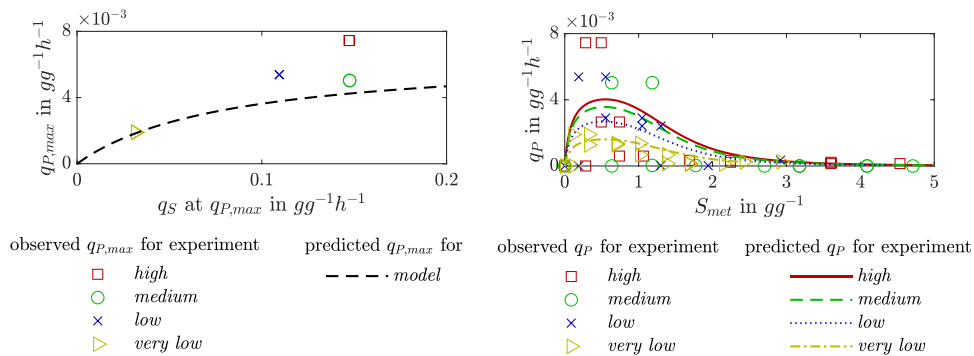
$$\frac{dS_{met}}{dt} = q_S. \tag{6 f}$$



**Fig. 3 – *Left*: Observed and modeled dependency of maximal specific productivity $q_{P,max}$ on substrate uptake rate $q_S$ *Right*: Observed and modeled specific productivity $q_P$ in function of metabolic load $S_{met}$.**

These equations describe the concentration changes over time of the five macroscopic components $[c_X; c_S; c_P; c_{P_R}; S_{met}]$ and the reactor volume $V_R$ with the substrate inflow $\dot{V}_{in}$ with a concentration $c_{S,in}$ and the reaction rates $[\mu; q_S; q_P; q_{PR}]$.

### 2.4. Model parametrization

Four calibration experiments with different constant feed rates, namely *high*, *medium*, *low* and *very low* were used to identify the model parameters. Therefore identifiable combinations of parameters, according to the methodology described by Brun et al. (2002) and Daume et al. (2019), were recursively estimated. In this methodology two selection measures are used:

$$\delta_p^{msqr} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(\frac{\partial x_j}{\partial p} * p\right)^2} \qquad \gamma_K = \frac{1}{\sqrt{\lambda p_k}} \tag{7}$$

The importance index $\delta_k^{msqr}$, is the sum of the local sensitivities of $n$ measured states x and all model parameters $p$. This measure gives a global importance measure of model parameters whereas the collinearity index $\gamma_K$ for a chosen parameter subset $p_k$ gives the linear dependency between the selected parameter sets with $\lambda p_k$ being the smallest eigenvalue of the Fisher Information Matrix. Selected parameter subsets were determined by a simplex optimization algorithm (fmincon; MATLAB) with the objective to minimize the overall sum of the squared residuals between measured and modeled states, as proposed in Ulonska et al. (2018a) and Brun et al. (2002). Hereby biomass, product within the cell and released product were used as measurements, weighted by their average errors (5%, 1% & 8%). Besides determining a generically applicable parameter set for all four calibration experiments together (overall model), optimal parameters for each individual experiment as well as for the validation experiment were determined. All resulting parameters are displayed in Table 1.

### 2.5. Feedback linearization

Feedback linearization aims to derive a control law that fully incorporates the nonlinearities of an input affine system

$$\dot{x} = f(x) + g(x)u$$
$$y = h(x)$$

with the time derivative $\dot{x}$ of the state vector $x$, the continuously differentiable system vector function $f(x)$, the continuously differentiable input vector function $g(x)$, the input variable $u$ and the output variable $y$ described by the continuously differentiable output function $h(x)$.

This linearization can be achieved by transforming the nonlinear system in $x$ into a linear system in $z$ by a suitable, unique and nonlinear state transformation $z = t(x)$. For this resulting linear system a controller can be designed, whereas the new synthetic input $v$ and the output $y$ are correlated by an integrator linear system. The inverse transformation of the linear system and the linear synthetic input yields the nonlinear control law $u = \alpha(x) + \beta(x)v$ that considers the nonlinearities of the initial system. $\alpha(x)$ and $\beta(x)$ will later be referred to as maintenance and driving term. A thorough description and discussion of the method of feedback linearization can be found in Isidori (1995). For completeness a brief outline of the method of feedback linearization is given below:

By defining the **Lie derivative**

$$L_a c(x) = \frac{\partial c(x)}{\partial x} a(x) = \text{grad}^T c(x) a(x)$$
$$L_a^i c(x) = \frac{\partial L_a^{i-1} c(x)}{\partial x} a(x) \tag{8}$$

with a vector function $a(x)$ and a scalar function $c(x)$, the **relative degree** $\delta$ of a system can be defined as

$$y^{(i)} = L_g L_f^i h(x) = 0 \quad \text{for } i = [0, \delta - 2]$$
$$y^{(\delta)} = L_g L_f^{\delta-1} h(x) \neq 0. \tag{9}$$

**Table 1 – Model parameters and feed settings of all experimental data-sets. The overall model parameters were determined for the four calibration data-sets (*high, medium, low, very low*). In addition model parameters were also determined for every single experiment. The overall model parameters were used to develop the controller and to design the set-point trajectory, whereas plant behavior was simulated by the parameters of the single experiments, including the parameters of the independent validation data-set.**

| | overall model | high ☐ | medium ○ | low ✕ | very low ▷ | validation | Dimension | Description |
|---|---|---|---|---|---|---|---|---|
| $\dot{V}_{in}$ | — | 0.014 | 0.010 | 0.004 | 0.002 | 0.009 | $Lh^{-1}$ | constant feed rate |
| $c_{S,in}$ | 850 | 930 | 838 | 930 | 822 | 900 | $gL^{-1}$ | substrate concentration in feed |
| $q_{S,max}$[a] | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | $gg^{-1}h^{-1}$ | maximal specific substrate uptake |
| $K_S$[a] | 0.0050 | 0.0050 | 0.0050 | 0.0050 | 0.0050 | 0.0050 | $gL^{-1}$ | half saturation constant |
| $m_S$ | 0.02 | 0.015 | 0.023 | 0.02 | 0.016 | 0.029 | $h^{-1}$ | maintenance constant |
| $Y_{X/S,max}$[a] | 0.47 | 0.47 | 0.47 | 0.47 | 0.47 | 0.47 | $gg^{-1}$ | maximal growth yield |
| $K_{Y_{X/S}}$ | 0.34 | 0.34 | 0.41 | 0.17 | 0.27 | 0.50 | $gg^{-1}$ | growth yield decay |
| $q_{P,max}$ | 0.0066 | 0.0086 | 0.0067 | 0.0100 | 0.0053 | 0.0068 | $gg^{-1}h^{-1}$ | maximal specific production rate |
| $K_{Sq_S}$ | 0.082 | 0.066 | 0.040 | 0.079 | 0.076 | 0.089 | $gg^{-1}h^{-1}$ | $q_P$ affinity to $q_S$ |
| $k$ | 4.6 | 3.9 | 5.7 | 4.7 | 5.3 | 3.9 | – | Haldane exponent |
| $K_{Iq_P}$ | 3.8 | 4.5 | 4.8 | 3.8 | 3.4 | 3.2 | $(gg^{-1})^{k-1}$ | $S_{met}$ dependent product decay |
| $K_{Sq_P}$ | 0.096 | 0.111 | 0.148 | 0.099 | 0.089 | 0.123 | $gg^{-1}$ | $S_{met}$ dependent product formation delay |
| $p_{rel,max}$ | 0.090 | 0.224 | 0.335 | 0.220 | 0.049 | 0.114 | $gg^{-1}h^{-1}$ | maximal product release |
| $K_{q_{S,rel}}$ | 0.113 | 0.076 | 0.090 | 0.076 | 0.168 | 0.098 | $h^{-1}$ | product release affinity to growth |
| $K_{Sps}$ | 2.1 | 1.7 | 2.1 | 2.1 | 2.4 | 2.1 | $gg^{-1}$ | $S_{met}$ dependent product release delay |

[a] Parameter values were determined before induction and kept constant during parametrization.

The relative degree $\delta$ corresponds to the number of differentiations of the output $y = h(x)$ until the input $u$ appears explicitly. If the relative degree $\delta$ is smaller than the system order $n = \dim \Gamma$, the system can not be fully linearized and consequently has to be split into the external linearizable states $\xi(x)$ and the internal nonlinearizable states $\eta(x)$, hence

$$\Gamma = \mathbf{W}(\Gamma) = \begin{bmatrix} \xi(\Gamma) \\ \eta(\Gamma) \end{bmatrix} = \begin{bmatrix} h(\Gamma) \\ L_I h(\Gamma) \\ \vdots \\ L_I^{\delta-1} h(\Gamma) \\ \hline \eta_1(\Gamma) \\ \vdots \\ \eta_{n-\delta}(\Gamma) \end{bmatrix} = \begin{bmatrix} y \\ \dot{y} \\ \vdots \\ y^{(\delta-1)} \\ \hline \eta_1(\Gamma) \\ \vdots \\ \eta_{n-\delta}(\Gamma) \end{bmatrix} \qquad (10)$$

whereas the nonlinearizable states $\eta(x)$ are chosen so that their time derivatives are independent of the input $u$, which leads to a subsystem that evolves independent of the input (Henson and Seborg, 1992). This can be achieved by solving the partial differential equation

$$L_J \eta_i(\Gamma) = \frac{\partial \eta_i(\Gamma)}{\partial \Gamma} \mathbf{J}(\Gamma) = 0. \qquad (11)$$

On the basis of the previous subdivision, the external and internal dynamics are determined by deriving the states with respect to time. For the linear and controllable **external dynamics**

$$\dot{\xi}(\Gamma) = \begin{bmatrix} \dot{z}_1 \\ \vdots \\ \dot{z}_{\delta-1} \\ \dot{z}_\delta \end{bmatrix} = \begin{bmatrix} z_2 \\ \vdots \\ z_\delta \\ L_I^{\delta} h(\Gamma) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ L_J L_I^{\delta-1} h(\Gamma) \end{bmatrix} u \qquad (12)$$

a linear controller can be designed. The **internal dynamics**

$$\dot{\eta}(\Gamma) = \begin{bmatrix} \dot{\eta}_1 \\ \vdots \\ \dot{\eta}_{n-\delta} \end{bmatrix} = \begin{bmatrix} L_I \eta_1(\Gamma) \\ \vdots \\ L_I \eta_{n-\delta}(\Gamma) \end{bmatrix} \qquad (13)$$

do not influence the output $y$ and consequently can not be observed by the output $y$. The stability of the internal dynamics is a necessity to obtain overall system stability and has to be examined. This verification varies for every system depending on the structure and dimension of the internal dynamics.

By defining the **synthetic input**

$$v = y^{(\delta)} = L_I^{\delta} h(\Gamma) + L_J L_I^{\delta-1} h(\Gamma) u, \qquad (14)$$

a linear correlation of the input $v$ and the output $y$ is obtained. Thus, any linear controller for the control of the external dynamics by the synthetic input $v$ can be designed. Rearranging Eq. (14) leads to the final nonlinear control law for the real input

$$u = -\frac{L_I^{\delta} h(\Gamma)}{L_J L_I^{\delta-1} h(\Gamma)} + \frac{1}{L_J L_I^{\delta-1} h(\Gamma)} v. \qquad (15)$$

The maintenance term $-L_I^{\delta} h(\Gamma)/L_J L_I^{\delta-1} h(\Gamma)$ describes the changes of the system due to nonlinear dynamics. The driving term $v/L_J L_I^{\delta-1} h(\Gamma)$ represents the influence of the output variable $y$ on the input $u$.

### 2.6. Two degree of freedom control

Based on the nonlinear correlation between the real input $u$ with the output $y$ given in Eq. (15) any linear control structures can be adopted as for example a PI controller:

$$u = \frac{\alpha_P e_* + \alpha_I \int_0^t e_* dt - L_I^{\delta} h(\Gamma)}{L_J L_I^{\delta-1} h(\Gamma)} + \frac{1}{L_J L_I^{\delta-1} h(\Gamma)} v \qquad (16)$$

with $e_*$ the tracking error and the PI tuning parameters $\alpha_P$ and $\alpha_I$ and the above described maintenance and driving term derived from the feedback linearization. Besides a feedback control law a feedforward control action which precalculates the response based on a perfect model assumption according to Eq. (15) can be defined. Based on this combination also referred as two-degree-of-freedom control (Araki and Taguchi, 2003) fast changes in the reference trajectory can be compensated by the feedforward term whereas the feedback control term in Eq. (16) accounts for model plant mismatches and reacts on random process deviations.

To quantify the accuracy of the model and the control strategies, root mean square error (RMSE) was used

$$\text{RMSE} = \sqrt{\frac{\sum_k^m (e_*)^2}{k}} \qquad (17)$$

with $k$ evaluation points of the distance between model simulation and measurement ($e_* = x - x_{real}$) or the tracking error ($e_* = w - w_{real}$). For the model error the RMSE was normalized by the range ($x_{max,real} - x_{min,real}$) whereas the control error was normalized by the average control output (mean $w_{real}$).

## 3. Results and discussion

### 3.1. Calibrated production phase model

Within Fig. 4, model outputs are plotted against the experimental observations of the four calibration experiments with *high*, *medium*, *low* and *very low* feed rates during induction. Overall, biomass and product within the cell with errors of around 10% and their associated rates ($\mu$, $q_S$ and $q_P$) with overall errors below 25% are well described by the model. For the released product $P_R$ higher errors (35.6%) are occurring. Although the release rate shows a clear dependency on overall growth as described by Eq. (5), the overestimated product release of the *very low* experiment in Fig. 4 shows that almost no product is released if $\mu$ is small enough. This is also indicated in Table 1 by a very small $p_{rel,max}$ for the *very low* experiment compared to higher $p_{rel,max}$ of all other experiments. The biomass specific productivity $q_P$ being the targeted control variable has an overall error of 11.7%. Highest model mismatches can be observed at high productivites, which can also be seen in Fig. 3 were the model underestimates the maximum productivity. As potential control is only attainable at lower $q_P$ levels this model inaccuracy at high $q_P$ levels can be regarded as acceptable.

The model parameters of Table 1 contain model parameters obtained by including all four calibration experiments as well as best parameters for the single experiments with *high*, *medium*, *low* and *very low* feed rates as well as the parameter set obtained from the validation experiment. Besides the release rate $p_{rel,max}$, which strongly varies between the processes (0.049–0.34 $gg^{-1}h^{-1}$) the other model parameters show a lower spread. For example $q_{P,max}$ resulted between 0.0053 and 0.01 $gg^{-1}h^{-1}$ for the single experiments, with a relative standard deviation of 24%. Altogether the standard deviation ranges from 17% to 38% (59% for $p_{rel,max}$), which indicates stable parameters with acceptable parameter uncertainties. According to Brun et al. (2002) these relative uncertainties corresponds to class two out of three classes (low, moderate & poor uncertainty). Together with
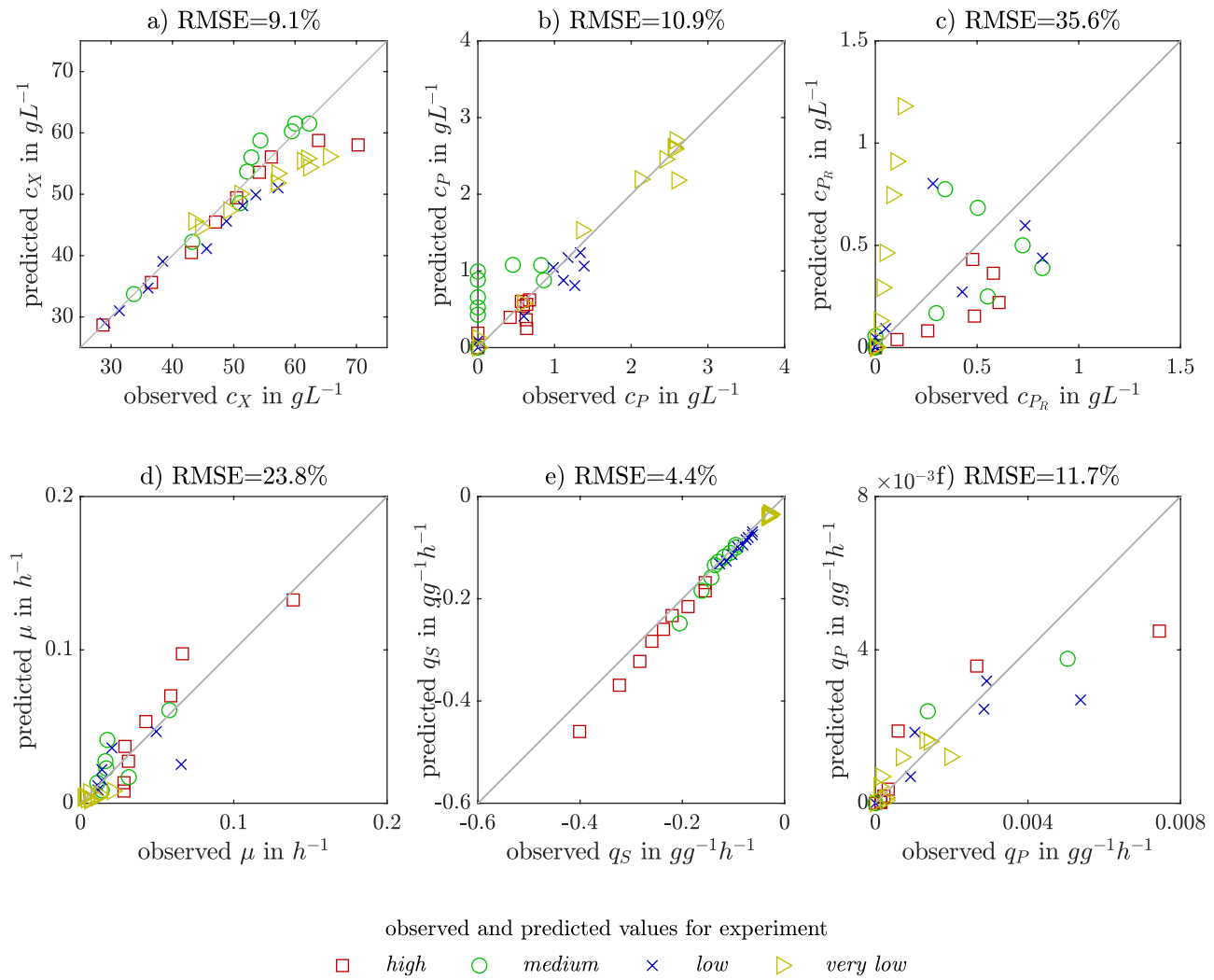
**Fig. 4 – Observation vs. prediction of fitted states biomass ($c_X$), product within the cell ($c_P$) and released product ($c_{P_R}$) and the corresponding biomass specific rates for growth ($\mu$), glycerol consumption ($q_S$) and heterologous protein production ($q_P$) with the corresponding root mean square error (RMSE) normalized by the observed range. The predictions were obtained under usage of the overall model parameter set of Table 1.**

the low state error, the presented kinetic model with the overall model parameter set (first column in Table 1) describes sufficiently well growth and formation of the recombinant protein under different feed regimes and occurring metabolic loads and is therefore suited for further tasks.

### 3.2. Feedback linearization of the modeled system

The method of feedback linearization described in 2.5 is applied to the calibrated production phase model (Eqn. (1)-(6)) with the state vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} V_R \\ c_X \\ c_S \\ c_P \\ c_{P_R} \\ S_{met} \end{bmatrix},$$

the manipulable input variable

$$u = \dot{V}_{in}$$

and the biomass specific productivity $q_P$ as the controlled output variable

$$y = h(\mathbf{x}) = q_P = q_{P,max} \frac{q_S}{q_S + K_{Sq_S}} \frac{S_{met}}{\frac{S_{met}^k}{KI_{q_P}} + S_{met} + K_{Sq_P}}.$$

The **relative degree** $\delta$ of the modeled system is one, because

$$0 \neq L_J h(\mathbf{x}) = -\frac{c_{S,in} - c_S}{V_R} \frac{q_P}{K_S + c_S} \left( \frac{q_{S,max} - q_S}{K_{Sq_S} + q_S} - \frac{K_S}{c_S} \right). \quad (18)$$

Note that with increasing $c_S$, $L_g h(\mathbf{x})$ approaches zero resulting in a loss of controllability. However, since the substrate concentration $c_S$ is relatively small in the suggested modes of operation and cannot exceed the input concentration $c_{S,in}$, the relative degree $\delta$ is well defined for the following considerations. With a system order of $n = 6$, the relative degree $\delta$ is not full and the transformed system has to be split into external and internal dynamics.

The system can be transformed into the new coordinates by solving the partial differential Eq. (11) so that the time derivatives of the transformed states $(\delta + 1)$ to $n$ are

independent of the input $\dot{V}_{in}$. With the solution of the partial differential equation

$$\eta_i = \phi\left(V_R\,c_X;\; V_R\,(c_{S,in} - c_S);\; V_R\,c_P;\; V_R\,c_{P_R};\; S_{met}\right),$$

where $\eta_i$ can be any function $\phi$ as long as all states are linear independent of each other, the **transformed states** become

$$
\mathbf{l} = \mathbf{W}(\mathbf{l}) = \begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} \frac{q_P}{c_X V_R} \\ (c_{S,in} - c_S)\,V_R \\ c_P\,V_R \\ c_{P_R}\,V_R \\ S_{met} \end{bmatrix}.
\tag{19}
$$

The **external dynamics**

$$\dot{\xi} = \dot{z}_1 = \dot{q}_P = L_I\,h(\mathbf{l}) + L_J\,h(\mathbf{l})\,\dot{V}_{in}
\tag{20}$$

with

$$
\begin{aligned}
L_I\,h(\mathbf{l}) = q_S\,q_P\Bigg( &\frac{1}{K_S + c_S}\left(\frac{q_{S,max} - q_S}{K_{Sq_S} + q_S} - \frac{K_S}{c_S}\right)c_X \\
&+ \frac{\left(K_{Sq_P} - \frac{(k-1)S_{met}^k}{K_{Iq_P}}\right)}{S_{met}\left(K_{Sq_P} + S_{met} + \frac{S_{met}^k}{K_{Iq_P}}\right)}\Bigg),
\end{aligned}
\tag{21}
$$

and Eq. (18) are linear in the $\mathbf{z}$ coordinates. The **internal dynamics** of the system

$$
\dot{\eta}(q_P, t) = \begin{bmatrix} \mu c_X V_R \\ q_S c_X V_R \\ (q_P c_X - q_{PR} c_P) V_R \\ q_{PR} c_X V_R \\ q_S \end{bmatrix}
\tag{22}
$$

have the order $(n - \delta) = 5$. For systems with relative degree $\delta = 1$, the stability of the zero dynamics $\dot{\eta}(q_P = 0, t)$ is a necessary and sufficient condition to prove the stability of the internal dynamics (Henson, 1997). According to Eq. (4),

$$q_P = 0 \leftrightarrow (q_S = 0) \vee (S_{met} = 0) \vee \left(\lim_{t \to \infty} S_{met} = \infty\right).
\tag{23}$$

Because the maximal reactor volume $V_R$ is limited, $S_{met}$ can not go to infinity. Since $S_{met} = 0$ is a special case of $q_S = 0$, only the case $q_S = 0$ has to be examined. It can easily be shown that, if $q_S = 0$, all five states and, therefore, the internal dynamics of the system are stable. This can also be explained physically: if $q_S = 0$, no substrate is within the system and all reactions stop, resulting in a dead but stable state of the system.

According to Eq. (14), the synthetic input becomes

$$v = \dot{q}_P = L_I\,h(\mathbf{l}) + L_J\,h(\mathbf{l})\,\dot{V}_{in},
\tag{24}$$

for whom a controller can be designed. Consequently the nonlinear control law is

$$\dot{V}_{in} = \frac{v - L_I\,h(\mathbf{l})}{L_J\,h(\mathbf{l})}
\tag{25}$$

For example, the nonlinear control law for a simple proportional feedback controller, as used in the following trajectory planning, with the linear control law $v = \alpha(w - y)$ is

$$\dot{V}_{in} = \frac{\alpha(w - q_P) - L_I\,h(\mathbf{l})}{L_J\,h(\mathbf{l})}.
\tag{26}$$

where $w$ is the reference trajectory for the desired behavior of $q_P$. The impact of the output variable $q_P$ on the input $\dot{V}_{in}$ is

described by the driving term $v/L_g h(\mathbf{x}) = \alpha(w - q_P)/L_g h(\mathbf{x})$. The changes of the system due to nonlinear dynamics are accounted for by the maintenance term $L_f h(\mathbf{x})/L_g h(\mathbf{x})$.

### 3.3. Trajectory planning

The derived control law (Eq. (26)) enables to track a predefined static setpoint $w$ or a dynamic reference trajectory $q_{P*}$. In comparison to a static setpoint $w$ a reference setpoint trajectory $q_{P*}$ is better suited to account for different aspects of the underlying process. Hereby, especially the time variant biological limits of the system, can be considered to guarantee a robust, long lasting and overall productive process. To do so, the induction phase of the analyzed recombinant protein production process was separated into three parts: .

1. a protective start up phase (see Fig. 5),
2. a productive phase with constant $q_P$'s at different levels and
3. a safe shut down phase (see Fig. 6)

For the start up phase different control scenarios were tested, which are displayed in Fig. 5. Dotted lines represent the aimed reference productivities as well as the control limits, whereas the resulting feeds ($\dot{V}_{in}$) and the reached productivities are shown as solid lines. A constant setpoint $w$ from the beginning (Fig. 5; bright grey line) leads to high initial feed rates (upper $\dot{V}_{in}$ limit) before it converges to a lower and more steady feed rate. In case of an increasing reference $q_{P*}$ trajectory (Fig. 5; grey line) the aimed setpoint is slowly approached by starting with very low feed rates.



simulated plant behavior for

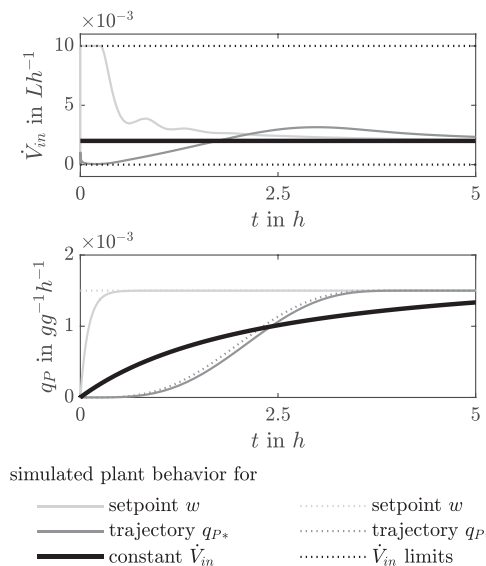| | |
|---|---|
| —— setpoint $w$ | ⋯⋯ setpoint $w$ |
| —— trajectory $q_{P*}$ | ⋯⋯ trajectory $q_{P*}$ |
| —— constant $\dot{V}_{in}$ | ⋯⋯ $\dot{V}_{in}$ limits |

**Fig. 5 – Three different control scenarios for the start up phase (0–5 h) with dotted lines the aimed setpoints and solid lines the resulting control responses of: 1. A static setpoint $w$ (bright grey line) leading to a large and rapid change of the feed, starting with very high feed rates. 2. An increasing $q_{P*}$ trajectory (grey line) resulting in very small feed rates at the beginning. Both variants could lead to potentially negative and irreversible effects on the living microorganisms. 3. A simple, constant feed $\dot{V}_{in}$ (black line) in the beginning ensures a safe and harmless start up of the production process, although $q_P$ only reaches slowly the aimed value.**
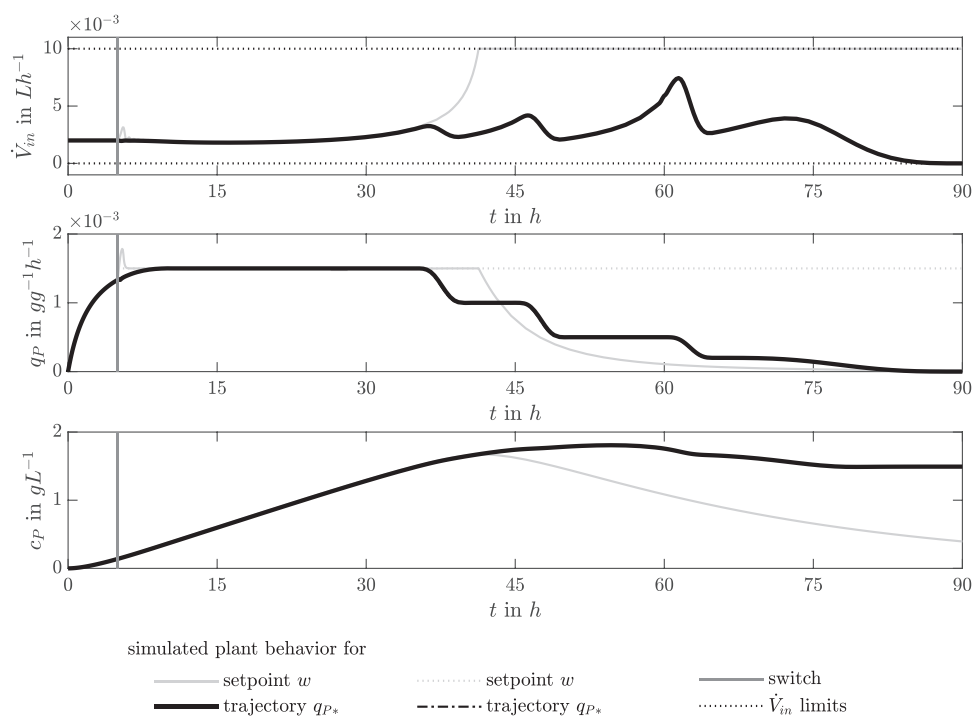
**Fig. 6 – Designed setpoint trajectory $q_{P*}$ (dotted black line) compared to constant setpoint $w$ (dotted grey line). Both controllers were started (grey vertical line) after the start up with a constant feed (0–5 h). In the first phase (5–40 h) both controller can follow the highest $q_p$ setpoint of $1.5 \cdot 10^{-3}$ $gg^{-1}h^{-1}$ by adapting $\dot{V}_{in}$. Beyond 36 h to maintain the constant setpoint $w$ strong $\dot{V}_{in}$ adaptions are needed, driving $\dot{V}_{in}$ to the upper limit, followed by a reduction of $q_P$ and product concentration $c_P$ (solid grey line). The stepwise decreasing $q_{P*}$ trajectory avoids this effect by a transition to a lower setpoint (solid black line). By doing so the $\dot{V}_{in}$ stays within the limits and higher product concentrations can be reached after approx. 55 h. As the controller can reproduce the suggested trajectory $q_{P*}$ almost perfectly, $q_{P*}$ (dotted black line) and the corresponding simulated plant behavior (solid black line) are congruent.**

A constant feed rate $\dot{V}_{in}$ (Fig. 5; black line) leads to a steady increase of $q_P$. In comparison to the constant setpoint $w$ and $q_{P*}$ trajectory responses, a constant feed during the first hours prevents rapid variations of the feed rates or respectively extremely low and high feed rates, which can have potentially negative and irreversible effects on the micro-organisms. Therefore, to ensure a protective start up of the production, the feed is held constant after induction (from $t = 0\,h$ to $t = 5\,h$).

In Fig. 6 the full production phase for a constant setpoint $w$ (grey line) and the proposed setpoint trajectory $q_{P*}$ (black line) is displayed. When the production phase has been safely started by applying a constant feed ($\dot{V}_{in}$), feedback control is activated (Fig. 6; grey vertical line). Dotted lines represent the control input and solid lines the obtained outputs. Due to the usage of a perfect model both lines are partially overlaying as the controller perfectly tracks the given setpoints ($w$ and $q_{P*}$).

In course of the process, the first and highest setpoint ($w$ and $q_{P*} = 1.5 \cdot 10^{-3}\,gg^{-1}h^{-1}$) can only be maintained for a certain time window before strong control actions, visible in a sharp increase in $\dot{V}_{in}$, are needed to keep the setpoint (see Fig. 6; setpoint $w$ between $t = 30\,h$ and $t = 45\,h$). To counteract a subsequent sharp decrease in $q_P$ and to prolong the production phase, a transition to a smaller $q_{P*}$ setpoint was included in the $q_{P*}$ trajectory (Fig. 6, trajectory $q_{P*}$). This ensures that the process stays productive for a longer time and the product concentration $c_P$ is kept at high levels with a maximum product concentration $c_P$ of $1.8\,gL^{-1}$ at $t = 55\,h$ compared to $1.7\,gL^{-1}$ for the constant setpoint $w$. The transitions

between the different $q_P$ levels were realized under usage of continuous polynomials by incorporating the productivity trajectory and its derivatives, which leads to smooth transitions.

Although highest product concentrations are reached after approx. 60 h a slow shutdown avoids additional stress and keeps high product concentrations, which enables increased flexibility for the harvest. The shutdown procedure was realized by a polynomial for the reference trajectory $q_{P*}$ that slowly approaches zero. Based on the feedback linearized control law (Eqn. 26), it was possible to design a reference trajectory $q_{P*}$ that includes all prerequisites for a stable and safe operation as well as to enable a long lasting productive phase to reach high product concentrations.

### 3.4. Feedback control and measurement scenarios

In order to deal with model inaccuracies and to act on random process deviations, a two-degrees-of-freedom controller (2-DOF) consisting of a model-based feedforward controller and a PI feedback controller as displayed in Fig. 7 was implemented. Hereby the model-based feedforward controller precalculates based on the feedback linearized model Eq. (26) the feed trajectory so that the predefined reference trajectory $q_{P*}$ is tracked, assuming a perfect model with no other deviations ($q_{P*} - q_p = 0$). Model-plant mismatches, namely the deviation of the real output $y$ from the desired trajectory $q_{P*}$ as well as random deviations, are detected by measurements and taken into account by the PI feedback controller in Eq. (27).
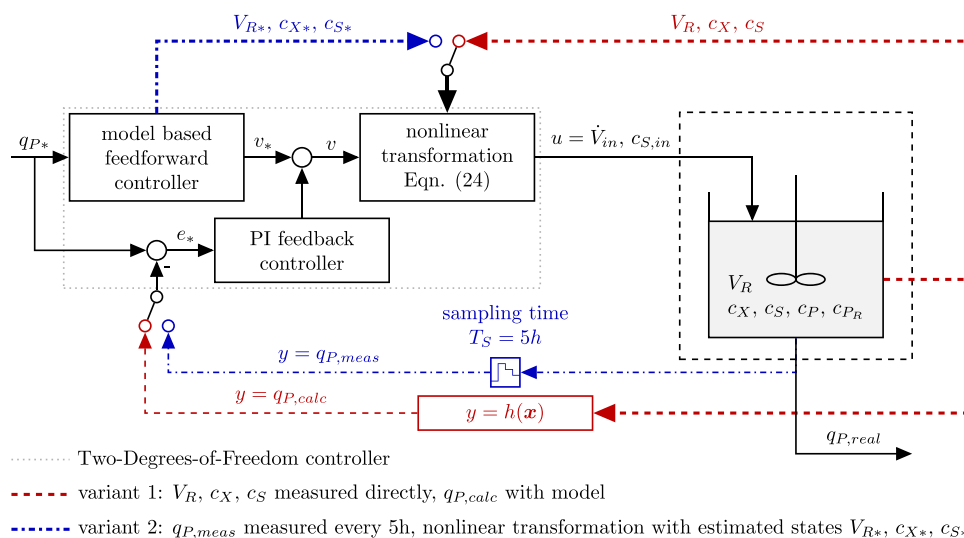
Fig. 7 – Block diagram of the controller setup: Two-Degrees-of-Freedom controller with Feedback Linearization (nonlinear transformation). The two examined variants, based on different measurement scenarios, are displayed. For variant 1 (red) time continuous measurements of $V_R$, $c_X$ and $c_S$ were assumed and $q_{P,calc}$ reconstructed via the model equation ($y = h(x)$). Variant 2 (blue) included a direct measurement of the productivity ($q_{P,meas}$), which is only available every 5 h. No additional measurements were assumed for variant 2.

The control law for the synthetic input $v$ (Eq. (14)) with a PI feedback controller becomes

$$v = v_* + \alpha_P (q_{P*} - y) + \alpha_I \int_0^t (q_{P*} - y)\, dt \qquad (27)$$

with the synthetic input $v_*$ for a perfect model, the desired dynamic setpoint $q_{P*}$, the actual output $y$ of the system and the control parameters $\alpha_P = 100$ and $\alpha_I = 600$. By measuring $y$ directly ($y = q_P$) or indirectly with correlated key variables ($y = h(x)$), the feedback controller adjusts the feed trajectory in order to minimize the control deviation $q_{P*} - y$. Based on different realistic measurement availabilities **two different variants**, were implemented. The block diagram of the control scheme is shown in Fig. 7 where the two variants are indicated by two different colors.

For **variant 1**, the reactor volume $V_R$ and the concentrations of biomass $c_X$ and glycerol $c_S$ are assumed to be continuously measurable. Since the specific productivity $q_{P,real}$ is not measured directly, it is calculated ($q_{P,calc}$) by inserting the continuous measurements in the model function ($y = h(x)$; Eq. (4)). This calculated output ($y = q_{P,calc}$) is fed into the feedback controller. The physical input $\dot{V}_{in}$ is obtained by the nonlinear transformation of the synthetic input $v$ evaluated with the measurements of $V_R$, $c_X$ and $c_S$. According to Eqs. (25) and (27), the nonlinear control law for variant 1 becomes

$$\dot{V}_{in} = \frac{v_* + \alpha_P e_* + \alpha_I \int_0^t e_*\, dt - L_f h(\boldsymbol{x})}{L_g h(\boldsymbol{x})}, \qquad (28)$$

with the control error $e_* = (q_{P*} - q_{P,calc})$ and the Lie-derivatives $L_f h(\boldsymbol{x})$ and $L_g h(\boldsymbol{x})$ evaluated with the measurements $V_R$, $c_X$ and $c_S$. Note that $V_R$, $c_X$ and $c_S$ appear explicitly in the control law.

**Variant 2** is based on the assumption that the specific productivity $q_{P,real}$ can be measured directly. Since this is a difficult quantity to determine, it is assumed that a measurement is only available every five hours as $q_{P,meas}$. In contrast to variant 1, the nonlinear transformation is performed with the estimated reactor volume $V_{R*}$ and the estimated concentrations of biomass $c_{X*}$ and glycerol $c_{S*}$ to

reduce additional measurement expenses. Based on Eqs. (25) and (27), the nonlinear control law for variant 2 becomes

$$\dot{V}_{in} = \frac{v_* + \alpha_P e_* + \alpha_I \int_0^t e_*\, dt - L_f h(\boldsymbol{x}_*)}{L_g h(\boldsymbol{x}_*)}, \qquad (29)$$

with $e_* = (q_{P*} - q_{P,meas})$ and the Lie-derivatives $L_f h(\boldsymbol{x}_*)$ and $L_g h(\boldsymbol{x}_*)$ evaluated with the estimated states $V_{R*}$, $c_{X*}$ and $c_{S*}$. Since the nonlinear transformation is carried out with the estimated reactor volume $V_{R*}$ and the estimated concentrations $c_{X*}$ and $c_{S*}$, the real reactor volume $V_R$ as well as biomass concentration $c_X$ and glycerol concentration $c_S$ do not occur in the control law explicitly.

### 3.5. Controller performance on validation data set

The two implemented control variants were applied on the validation plant model (see last column of Table 1). The plant behavior is shown by a black dotted line in Fig. 8 for variant 1 and in Fig. 9 for variant 2. The solid line represents the reference behavior with perfect model parameters and no additional disturbances. The average control errors are given as normalized RMSE in Table 2. The grey shaded area of Figs. 8 and 9 represents the range of the expected behavior under the usage of the single parameter sets of the fitting experiments (see parameters in Table 1). With an average error of 7% both control variants were able to follow the $q_P*$ reference trajectory under a realistic model-plant mismatch. With a perfectly known plant model and full measurement information the error was < 0.05%.

For **variant 1** (See Fig. 8) the absence of direct information on the real productivity $q_{P,real}$ leads to only marginal adjustments in the feed rate. This adjustments originate from slightly different growth behavior measurable by $c_X$ and $c_S$. The deviations in variant 1, were mainly due to a constant offset from $q_{P,real}$ to $q_{P*}$. Besides this offset, the simulated process showed a steady and long lasting $q_P$ trajectory and reached the expected product amount which is already a
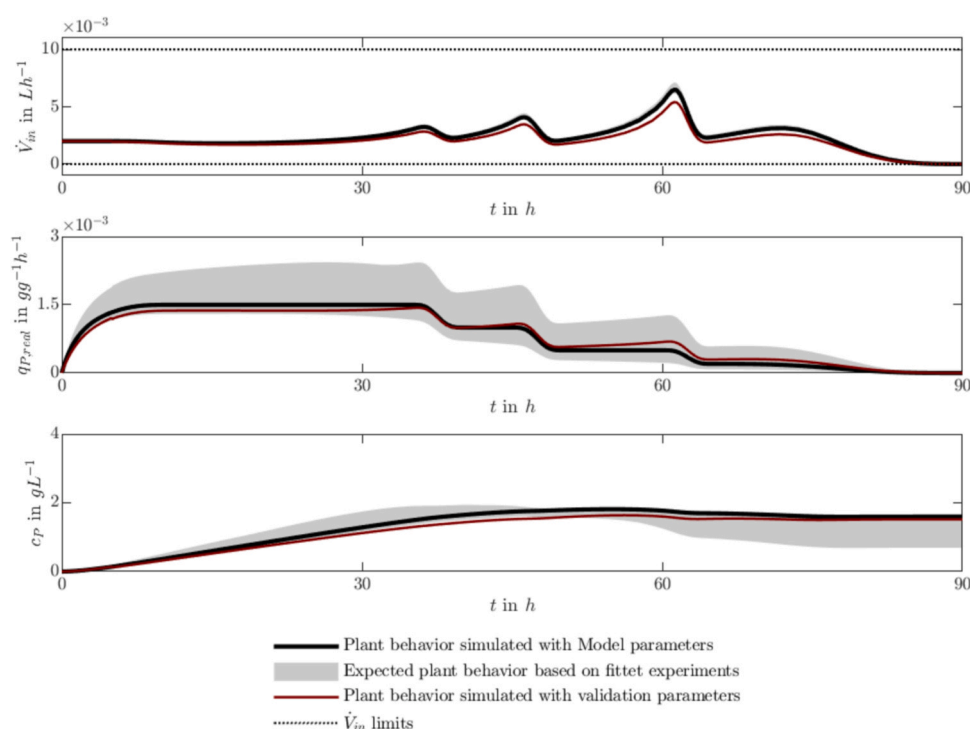
**Fig. 8 – Controller behavior of variant 1 with continuous $V_R$, $c_X$, $c_S$ measurements. $q_{P,real}$ is not known to the controller. The feedback control action $\dot{V}_{in}$ is therefore based the calculated $q_P$ according to Eq. (4). The ideal plant behavior is simulated with the overall model parameters (black line). Simulating the plant behavior with the validation parameters leads to the red dashed line. The expected plant behavior is obtained by simulations with the four individual calibration parameter sets (*high, medium, low* and *very low*) (grey area). For model parameters see Table 1 and Fig. 7 for the block diagram of the controller.**

significant improvement to standard fed-batch production processes with only short productive phases as reported in (Marisch et al., 2013; Wechselberger et al., 2012). From a process analytical technology (PAT) perspective real-time measurements can be established for biomass by dielectric spectroscopy (Dabros et al., 2010) or state estimation techniques (Dewasme et al., 2013; Ulonska et al., 2018b) and glycerol by near infrared spectroscopy (Macaloney et al., 1997).

Variant 2 assumed a direct measurement of $q_P$, which in a realistic setting can occur every five hours (Kager et al., 2018), also showed a relatively low error of 7%. This time-discrete behavior of variant 2 can be seen in Fig. 9) where every 5 h the feedback term corrects the feed rate, resulting in a good tracking of the aimed productivity ($q_{P,real}$).

Overall the observed control error of 7% is comparable to other published feedback controllers used to control either the specific growth rate (Dabros et al., 2010) or the biomass specific substrate uptake rate (Kager et al., 2020) in biotechnological processes. This makes the direct control of $q_P$ an interesting option to ensure long lasting recombinant protein production processes by a tailored and optimized addition of substrate during the induction phase.

### 3.6. Controller uncertainty and stability

To obtain information regarding the stability and the uncertainty of the two control variants, a robustness analysis was performed. Therefore the plant behavior was simulated with the parameter sets of the four individual calibration

experiments *high, medium, low* and *very low* (See Table 1). The controller scheme for both variants is displayed in Fig. 7 and described in Section 3.4. The range of the obtained plant behavior is indicated with a grey shaded area in Fig. 8 for variant 1 and in Fig. 9 for variant 2. The average control errors are given in Table 2.

For **variant 1** it can be seen in Fig. 8, that there are considerable deviations of the real productivity trajectory $q_{P,real}$ (grey area) from the targeted $q_P^*$ trajectory, displayed as black solid line. With an average accuracy of 28% this variant has the highest uncertainty to track $q_P$. For this control variant, direct information on the real productivity is not accessible and therefore, the control action is only slightly influenced. Consequently, the feed trajectories of the fitting experiments (grey shaded area) differ only slightly from each other, as the control action is manly determined by the feedforward part. However, only model-plant mismatches are considered for this analysis and possible system disturbances would not be detected and acted on by open loop control. Although, the controller shows the lowest accuracy, stable trajectories in terms of the input feed $\dot{V}_{in}$, the output $q_{P,real}$ as well as the product concentrations $c_P$ can be expected as none of the simulations showed an unstable behavior and was able to run for over 60 h.

In Fig. 9, the simulations for the robustness analysis of **variant 2** are displayed. Since the real productivity $q_{P,real}$ is measured discretely every five hours and is therefore partly known to the controller as measured productivity $q_{P,meas}$, the feedback controller intervenes more compared to variant 1. This stronger feedback controller intervention can be seen in the feed rate $\dot{V}_{in}$ of Fig. 9 were the expected plant behavior
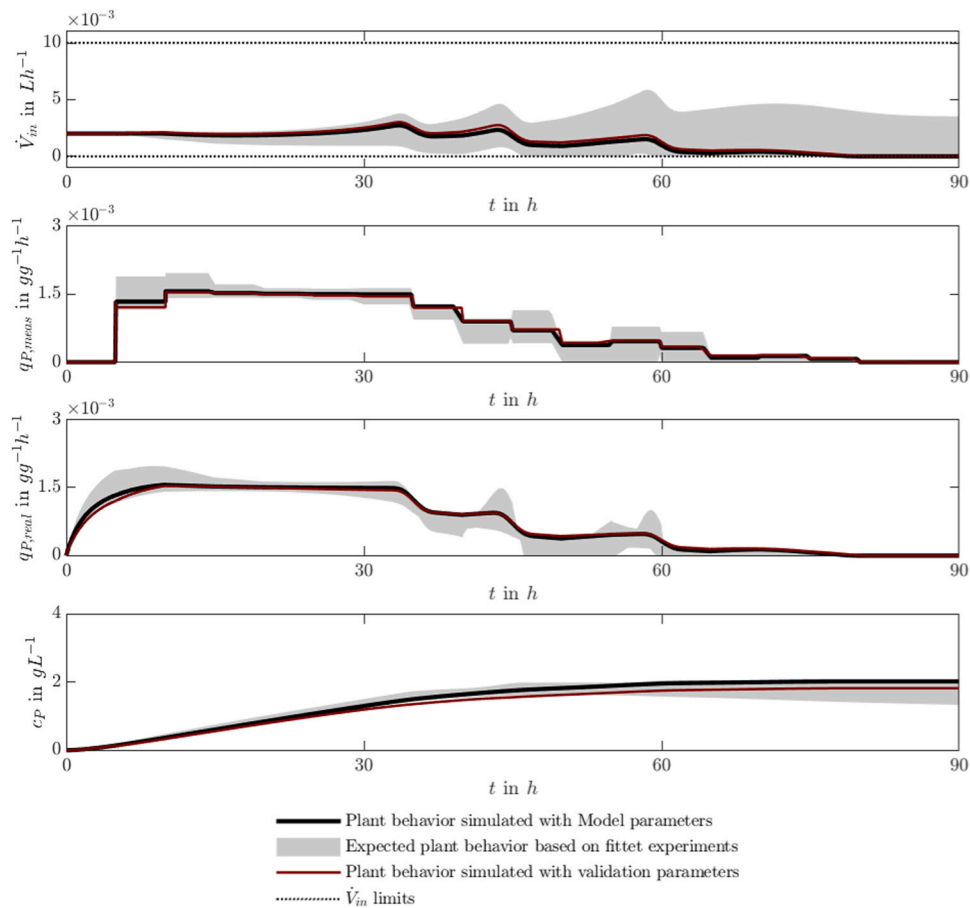
Fig. 9 – Controller behavior of variant 2 with a direct measurement of $q_P$ every 5 h ($q_{P,meas}$). The feedback control action $\dot{V}_{in}$ is based on this discontinuous measurement leading to ta continuous plant behavior ($q_{P,real}$). The ideal plant behavior is hereby simulated with the overall model parameters (black line). Simulating the plant behavior with the validation parameters leads to the red dashed line. The expected plant behavior is obtained by simulations with the four individual calibration parameter sets (*high*, *medium*, *low* and *very low*) (grey area). For model parameters see Table 1 and Fig. 7 for the block diagram of the controller.

**Table 2 – RMSE and normalized NRMSE of $q_{P^*} - q_{P,real}$ for the two examined controller variants applied on the validation parameter set and on the calibration parameter sets.**

| | Paramater set | | Parameter set | |
|---|---|---|---|---|
| | validation | | calibration | |
| | RMSE $gg^{-1}h^{-1}$ | NRMSE % | RMSE $gg^{-1}h^{-1}$ | NRMSE % |
| Variant 1 | $1.05 \cdot 10^{-4}$ | 0.07 | $4.28 \cdot 10^{-4}$ | 0.28 |
| Variant 2 | $1.05 \cdot 10^{-4}$ | 0.07 | $2.38 \cdot 10^{-4}$ | 0.16 |

based on the fitting experiments covers a bigger area. These stronger control actions are effective indicated by the narrow grey shaded area of $q_{P,real}$ and the overall reduced average error of 16%. Especially, the highest $q_P$ level ($q_{P^*} = 1.5\,gg^{-1}h^{-1}$) can be reached and kept very accurately under the present model-plant mismatches. A direct measurement of $q_P$, which realistically is available only every 5 h is an effective way to improve the feedback behavior of the implemented 2-DOF controller. Compared to the time continuous measurements of $V_R$, $c_X$ and $c_S$ for variant 1, variant 2 with time discrete $q_P$ measurements showed a 40% better tracking accuracy.

### 3.7.    *Potential product increase*

For a fair comparison, the biomass specific product titer was evaluated at maximal product concentration $c_P$ for the five experiments (*high*, *medium*, *low*, *very low and* validation) as well as the simulation results of the two analyzed control variants (controlled $q_P$ V1; controlled $q_P$ V2) applied on the validation parameter set. The biomass specific product titers as well as the corresponding feed profiles are displayed in Fig. 10. As uncertainty for the controller simulations, the standard deviation of the robustness analysis was added.

Within Fig. 10, the potential of the proposed control strategy becomes clearly visible. Within the experiments, characterized by different constant feed rates, highest specific titers were reached with a *very low* constant feed of 0.002 L/h. Compared to that, both control variants applied on the validation data set show their potential to further increase this titer by + 17.1% for variant 1 and + 40.8% for variant 2. Based on one-sided t-tests with a significance level of $\alpha = 0.05$ the potential increase of variant 1 is not significant (p = 0.076), whereas variant 2 shows a significant improvement compared to *very low* (p = 0.012) as well as to variant 1 (p = 0.009).

Even if the missing experimental verification makes it difficult to give an ultimate statement, the potential increase in titer and the more stable and prolonged production are
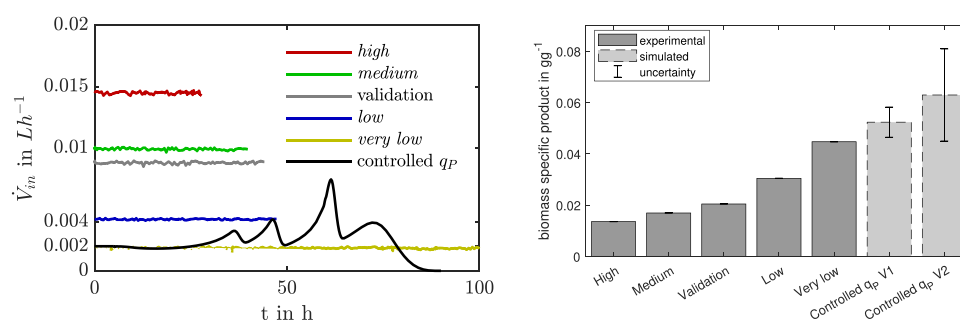
**Fig. 10 –** *Left*: Production phase feed profiles of the four calibration experiments, the validation experiment and the resulting feed trajectory of the controlled $q_P$, obtained with the perfect model *Right*: Maximal, biomass specific titers of the four calibration experiments, the validation experiment and the two control variants containing the simulation results of the validation parameter set, with the uncertainty (single standard deviation), derived from the robustness analysis.

strong arguments in favor of the proposed direct control of $q_P$.

## 4.    Conclusion

Within this contribution, a novel model-based control strategy for the direct control of the biomass specific productivity $q_P$ in recombinant *E. coli* fed-batch processes was established under usage of nonlinear feedback linearization and the implementation of a Two-Degrees-of-Freedom controller. For this purpose a nonlinear biological process model was developed describing product formation in function of feed addition and metabolic stress. The kinetic model fitted well the production phases of four analyzed processes with different constant feeds. From the model fits, it could be observed that with *very low* feed addition growth related product release was minimized and a long lasting productivity at lower levels was achieved. Therefore higher errors of the model at elevated productivities can be accepted as for the analyzed recombinant process stable and prolonged control can only be achieved at low production levels.

Based on the feedback linearized model, an optimal setpoint trajectory $q_{P^*}$ was designed to ensure a prolonged production phase lasting up to 75 h. The setpoint trajectory was characterized by a short and safe start up phase with a constant feed, a subsequent control of $q_P$ at different productivity levels and a smooth shut down to preserve high product concentrations.

The controller was tested for two realistic measurement scenarios under different model-plant mismatches. This analysis revealed that an at-line determination of $q_P$ every five hours is sufficient for effective feedback control with comparable control errors to other works aiming to control cell specific rates such as $\mu$ or $q_S$. For a practical application the inclusion of occurring time delays of the at-line $q_P$ measurements would be necessary.

Overall, the direct control of $q_P$ revealed high potentials to stabilize and prolong the induction phase of recombinant protein production processes and to increase biomass specific titers of up to 40.8%. Future experimental verification will concretize these expected positive effects and the potential of the control strategy towards the establishment of continuous and more productive recombinant production processes.

## CRediT authorship contribution statement

**Julian Kager**: Conceptualization, Model establishment, Data visualization, Writing – original draft. **Johanna Bartlechner**: Methodology, Control application, Visualization, Writing – original draft. **Christoph Herwig**: Supervision, Writing – review & editing. **Stefan Jakubek**: Conceptualization, Supervision, Methodology, Project administration, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

## References

Abadli, M., Dewasme, L., Tebbani, S., Dumur, D., Vande Wouwer, A., 2020. Generic model control applied to e. coli bl21(de3) fed-batch cultures. Processes 8, 772. https://doi.org/10.3390/pr8070772

Abadli, M., Dewasme, L., Tebbani, S., Dumur, D., Wouwer, A.V., 2021. An experimental assessment of robust control and estimation of acetate concentration in Escherichia coli bl21 (de3) fed-batch cultures. Biochem. Eng. J., 108103.

Aguilar-López, R., 2017. Input-output linearizing-type controller design with application to continuous bioreactor. Comptes rendus de l'Académie bulgare des Sci. 70, 321–328.

Araki, M., Taguchi, H., 2003. Two-degree-of-freedom pid controllers. Int. J. Control, Automation, Syst. 1, 401–411.

Baeza, J.A., 2017. Principles of bioprocess control. In: Current Developments in Biotechnology and Bioengineering. Elsevier, pp. 527–561. https://doi.org/10.1016/B978-0-444-63663-8.00018-5

de Battista, H., Picó, J., Picó-Marco, E., 2006. Globally stabilizing control of fed-batch processes with haldane kinetics using growth rate estimation feedback. J. Process Control 16, 865–875. https://doi.org/10.1016/j.jprocont.2006.02.001

Brun, R., Kühni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of asm2d parameters–systematic selection and tuning of parameter subsets. Water Res. 36, 4113–4127.

Dabros, M., Schuler, M.M., Marison, I.W., 2010. Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. Bioprocess Biosyst. Eng. 33, 1109–1118.

Daume, S., Kager, J., Herwig, C., 2019. Time resolved sensitivity & identifiability analysis for directed parametrization of highly dynamic models. In: Computer Aided Chemical Engineering 46. Elsevier, pp. 1111–1116.

Dewasme, L., Coutinho, D., Wouwer, A.V., 2011, Adaptive and robust linearizing control strategies for fed-batch cultures of microorganisms exhibiting overflow metabolism, in: Cetto, J. A., Ferrier, J.L., Filipe, J., (Eds.), Informatics in Control, Automation and Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg. volume 89 of Lecture Notes in Electrical Engineering, 283–305.

Dewasme, L., Fernandes, S., Amribt, Z., Santos, L., Bogaerts, P., Wouwer, A.V., 2015. State estimation and predictive control of fed-batch cultures of hybridoma cells. J. Process Control 30, 50–57.

Dewasme, L., Goffaux, G., Hantson, A.L., Wouwer, A.V., 2013. Experimental validation of an extended kalman filter estimating acetate concentration in e. coli cultures. J. Process Control 23, 148–157.

Douma, R.D., Verheijen, P.J., de Laat, W.T., Heijnen, J.J., van Gulik, W.M., 2010. Dynamic gene expression regulation model for growth and penicillin production in penicillium chrysogenum. Biotechnol. Boeng. 106, 608–618.

Ellis, M., Patel, P., Edon, M., Ramage, W., Dickinson, R., Humphreys, D.P., 2017. Development of a high yielding e. coli periplasmic expression system for the production of humanized fabafragments. Biotechnol. Progress 33, 212–220.

Fan, D.D., Luo, Y., Mi, Y., Ma, X.X., Shang, L., 2005. Characteristics of fed-batch cultures of recombinant Escherichia coli containing human-like collagen cdna at different specific growth rates. Biotechnol. Lett. 27, 865–870.

Henson, M.A. (Ed.), 1997. Nonlinear Process Control. Prentice-Hall, Upper Saddle River, NJ.

Henson, M.A., 2006. Biochemical reactor modeling and control. IEEE Control Syst. 26, 54–62. https://doi.org/10.1109/MCS.2006.1657876

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

Hocalar, A., Türker, M., 2014. Nonlinear control of large-scale fed-batch yeast fermentation: control of the specific growth rate. Turkish J. Eng. Environ. Sci. 38, 338–353. https://doi.org/10.3906/muh-1411-14

Isidori, A., 1995. Nonlinear Control Systems: [1]. 3. ed. ed. Springer,, London ([u.a.]).

Kager, J., Herwig, C., Stelzer, I.V., 2018. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. Chem. Eng. Sci. 177, 234–244.

Kager, J., Tuveri, A., Ulonska, S., Kroll, P., Herwig, C., 2020. Experimental verification and comparison of model predictive, pid and model inversion control in a penicillium chrysogenum fed-batch process. Process Biochem. 90, 1–11.

Kroll, P., Hofer, A., Stelzer, I.V., Herwig, C., 2017. Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. Process Biochem. 62, 24–36.

Lee, P., Sullivan, G., 1988. Generic model control–theory and applications. IFAC Proc. Volumes 21, 111–119.

Lee, S.Y., 1996. High cell-density culture of Escherichia coli. Trends Biotechnol. 14, 98–105.

Macaloney, G., Hall, J., Rollins, M., Draper, I., Anderson, K., Preston, J., Thompson, B., McNeil, B., 1997. The utility and performance of near-infra red spectroscopy in simultaneous monitoring of multiple components in a high cell density recombinant escherichiacoli production process. Bioprocess Eng. 17, 157–167.

Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M., Striedner, G., 2013. Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level sod protein production. Microbial Cell Factories 12, 58.

Marschall, L., Sagmeister, P., Herwig, C., 2016. Tunable recombinant protein expression in e. coli: enabler for continuous processing? Appl. Microbiol. Biotechnol. 100, 5719–5728.

Marschall, L., Sagmeister, P., Herwig, C., 2017. Tunable recombinant protein expression in e. coli: promoter systems and genetic constraints. Appl. Microbiol. Biotechnol. 101, 501–512.

Mears, L., Stocks, S.M., Sin, G., Gernaey, K.V., 2017. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. J. Biotechnol. 245, 34–46.

Narayanan, H., Luna, M.F., von Stosch, M., Cruz Bournazou, M.N., Polotti, G., Morbidelli, M., Butté, A., Sokolov, M., 2020. Bioprocessing in the digital age: The role of process models. Biotechnol. J. 15, e1900172. https://doi.org/10.1002/biot.201900172

Neubauer, P., Hofmann, K., Holst, O., Mattiasson, B., Kruschke, P., 1992. Maximizing the expression of a recombinant gene in Escherichia coli by manipulation of induction time using lactose as inducer. Appl. Microbiol. Biotechnol. 36, 739–744.

Neubauer, P., Lin, H., Mathiszik, B., 2003. Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in Escherichia coli. Biotechnol. Bioeng. 83, 53–64.

Picon, A., Teixeira de Mattos, M., Postma, P., 2005. Reducing the glucose uptake rate in Escherichia coli affects growth rate but not protein production. Biotechnol. Bioeng. 90, 191–200.

Pinsach, J., de Mas, C., López-Santín, J., Striedner, G., Bayer, K., 2008. Influence of process temperature on recombinant enzyme activity in Escherichia coli fed-batch cultures. Enzyme Microbial Technol. 43, 507–512.

Qin, S.J., Badgwell, T.A., 2003. A survey of industrial model predictive control technology. Control Eng. Pr. 11, 733–764.

Ramalingam, S., Gautam, P., Mukherjee, K.J., Jayaraman, G., 2007. Effects of post-induction feed strategies on secretory production of recombinant streptokinase in Escherichia coli. Biochem. Eng. J. 33, 34–41.

Schuller, A., Cserjan-Puschmann, M., Köppl, C., Grabherr, R., Wagenknecht, M., Schiavinato, M., Dohm, J.C., Himmelbauer, H., Striedner, G., 2020. Adaptive evolution in producing microtiter cultivations generates genetically stable Escherichia coli production hosts for continuous bioprocessing. Biotechnol. J., 2000376.

Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P., Van Impe, J.F., 2004. Optimal adaptive control of (bio) chemical reactors: past, present and future. J. Process Control 14, 795–805.

Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., Striedner, G., 2017. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. Biotechnol. J. 12, 1600546.

Ulonska, S., Kroll, P., Fricke, J., Clemens, C., Voges, R., Müller, M.M., Herwig, C., 2018a. Workflow for target-oriented parametrization of an enhanced mechanistic cell culture model. Biotechnol. J. 13, 1700395.

Ulonska, S., Waldschitz, D., Kager, J., Herwig, C., 2018b. Model predictive control in comparison to elemental balance control in an e. coli fed-batch. Chem. Eng. Sci. 191, 459–467.

Versyck, K.J., van Impe, J.F., 1998. Feedback linearizing controllers for optimal parameter estimation of haldane kinetics. IFAC Proc. Volumes 31, 129–134. https://doi.org/10.1016/S1474-6670(17)40172-8

Wechselberger, P., Sagmeister, P., Engelking, H., Schmidt, T., Wenger, J., Herwig, C., 2012. Efficient feeding profile optimization for recombinant protein production using physiological information. Bioprocess Biosyst. Eng. 35, 1637–1649.

Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M., Altenbuchner, J., 2001. High-cell-density fermentation for production of l-n-carbamoylase using an expression system based on the Escherichia coli rhabad promoter. Biotechnol. Bioeng. 73, 95–103.

Wunderlich, M., Taymaz-Nikerel, H., Gosset, G., Ramírez, O.T., Lara, A.R., 2014. Effect of growth rate on plasmid dna production and metabolic performance of engineered Escherichia coli strains. J. Biosci. Bioeng. 117, 336–342.

Yoo, S.J., Jeong, D.H., Kim, J.H., Lee, J.M., 2016. Optimization of microalgal photobioreactor system using model predictive control with experimental validation. Bioprocess Biosyst. Eng. 39, 1235–1246.

Zalai, D., Hevér, H., Lovász, K., Molnár, D., Wechselberger, P., Hofer, A., Párta, L., Putics, Á, Herwig, C., 2016. A control strategy to investigate the relationship between specific productivity and high-mannose glycoforms in cho cells. Appl. Microbiol. Biotechnol. 100, 7011–7024.