

# Adaptive predictive control of bioprocesses with constraint-based modeling and estimation

Banafsheh Jabarivelisdeh<sup>b,c</sup>, Lisa Carius<sup>b</sup>, Rolf Findeisen<sup>b</sup>, and Steffen Waldherr<sup>c</sup>

<sup>b</sup>*Laboratory for Systems Theory and Automatic Control,  
Otto-von-Guericke-Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg,  
Germany*

<sup>c</sup>*KU Leuven, Department of Chemical Engineering, Celestijnenlaan 200F-bus 2424,  
3001 Leuven, Belgium*

**Corresponding author:** Steffen Waldherr, E-mail: [steffen.waldherr@kuleuven.be](mailto:steffen.waldherr@kuleuven.be).  
Phone: +32 16 37 29 25.

This is the authors' version of the paper intended for scholarly archiving at the authors' institutions. The definite version of the paper has been published in *Computers & Chemical Engineering*, Volume 135, 6 April 2020, 106744.  
<https://doi.org/10.1016/j.compchemeng.2020.106744>

## Abstract

Control of biotechnological processes is currently recipe-based with insufficient ability to handle possible uncertainties, which results in sub-optimal production processes. To address this problem, model-based optimization and control approaches can be implemented to derive optimal control strategies. However, for reliable performance of model-based control, it is crucial to use flexible and adaptive control strategies which address biological variability while compensating for uncertainties. In this work, we present an approach for adaptive control of a bioprocess based on dynamic flux balance models. A previously developed bilevel approach for bioprocess optimization is implemented inside a model predictive control (MPC) routine. To account for model uncertainties, a moving horizon estimation algorithm is combined with the MPC in order to estimate uncertain parameters of the underlying model online for different metabolic modes. We apply this method to maximize the productivity of a target metabolite under microaerobic conditions by adapting the degree of oxygen-limitation online.

**Keywords:** Productivity, Constraint-based models, Model predictive control, Moving horizon estimation.

# 1 Introduction

Model-based optimization and control of bioprocesses is classically done using unstructured models based on a very simplistic representation of cellular metabolism. Unstructured models generally involve lumped descriptions of intracellular metabolism in terms of a specific growth rate and constant yield coefficients [1]. While this kind of models are of low complexity, they have limited capability to predict the wide range of cellular behaviour in response to changing environmental conditions. Compared to unstructured models, flux balance analysis (FBA)-based models [2, 3] provide a more detailed description of intracellular metabolism by considering a metabolic network which consists of metabolites and metabolic fluxes connecting the metabolites. These models are defined based on the stoichiometry of the metabolic network and give an optimal metabolic flux distribution by maximizing a relevant biological objective.

FBA models have shown successful applications in biotechnology and bioprocess optimization [4]. They allow to explore potential production capabilities of metabolic networks and can be applied efficiently to identify targets of genetic interventions as well as the optimal time for interventions and/or determining the optimal process condition (such as feeding and aeration level) for improved bioprocess yield and productivity [5, 6]. However, the application of FBA-based models is not yet well studied in bioreactor operation and control.

In our previous work [7], we have implemented the dynamic enzyme-cost FBA (deFBA) model [8] within a bilevel framework. The deFBA-based bilevel approach improves bioprocess productivity by finding optimal strategies for dynamic manipulation of cellular metabolism in both genetic and process levels. However, as parameter variations may lead to changes in the dynamics, it is crucial to have advanced strategies which guarantee the performance of model-based control in the presence of disturbances and plant-model mismatch. In this direction, it was shown that introducing measurement feedback into the open-loop FBA-based optimization allows for improved performance of the bioreactor control [9, 10]. In [10], we have considered a closed-loop implementation of the deFBA-based bilevel optimization method. Repeating the bilevel optimization updated by online measurements resulted in improved controller performance in achieving higher process productivity by compensating plant-model mismatches.

However, a remaining bottleneck of the model-based control is that the underlying model is often only valid for a limited operational range. In fact, models may be restricted to narrow growth modes of the bioprocess and may not cover all transient modes of the organism during the process operation. If the biological variability is neglected, the performance of model-based control strategies may be limited. To address this aspect, in this work we consider a closed-loop implementation of an adaptive deFBA-based control. For that, we implement the deFBA-based bilevel problem within a model predictive control (MPC) [11] which allows for feedback corrections. Instead of using the nominal value of model parameters, we account for uncertain and time-variable parameters within the control scheme. To do so, MPC is combined with a moving horizon estimation (MHE) algorithm to estimate uncertain modelling param-

eters for different growth modes of the organism online. The approach allows to adapt the model for each transient mode, using the data obtained during the process. Moreover, to update the initial condition of the deFBA model in each MPC iteration, the controller includes a resource balance analysis (RBA) algorithm which predicts the optimal steady state composition of the measured biomass, which is used as an estimate of the biomass components.

As an application, we consider oxygen-limited (microaerobic) batch growth of *Escherichia coli* for ethanol production using a small-scale metabolic-genetic network. Depending on the degree of oxygen limitation in the microaerobic growth, the cells transition to different metabolic modes which lead to changes in dynamics of the system. Through a simulation-based study, we demonstrate that the proposed MPC with the adaptive deFBA model can be a suitable approach to control such a process.

The paper is structured as follows. In Section 2, the deFBA model is reviewed, together with the open-loop optimization problem for maximal productivity. In Section 3, the adaptive MPC algorithm with detailed descriptions of its components is presented. The case study and the implementation of the proposed algorithm on that are discussed in Sections 4 and 5.

## 2 Methods

### 2.1 Review of the deFBA model for metabolic network optimization

In this section, we give a short introduction of the deFBA model [8, 12]. It is formulated as a dynamic optimization problem that includes a stoichiometric part which captures the mass balances of the metabolic-genetic network, and a constraint based part which describes the biophysical constraints such as the metabolic capacity based on cellular resource allocation.

To model the network, different classes of biochemical species are considered: extracellular nutrients and products with the molar amount vector  $z \in \mathbb{R}_{\geq 0}^{n_1}$ , intracellular metabolites with the molar amount vector  $m \in \mathbb{R}_{\geq 0}^{n_2}$ , and macromolecules such as gene products with the molar amount vector  $p \in \mathbb{R}_{\geq 0}^{n_3}$ . The total biomass  $B$  (dry weight in g) is obtained by the scalar product  $b^T p$ , where the vector  $b = (b_1, \dots, b_{n_3})$  contains the molecular weights (in g/mol) of the macromolecules represented in the vector  $p$ . The network reactions are grouped accordingly into three classes:

- Exchange reactions between the inside and the outside of the cell (substrate uptake and product secretion), with reaction flux vector  $V_z \in \mathbb{R}^{r_1}$ .
- Metabolic reactions, converting metabolites into each other, with reaction flux vector  $V_m \in \mathbb{R}^{r_2}$ .
- Biomass production reactions, converting metabolites into macromolecules, with reaction flux vector  $V_p \in \mathbb{R}^{r_3}$ .

We write shortly  $V = (V_z, V_m, V_p) \in \mathbb{R}^{r_1+r_2+r_3}$ , with a unit of molar amount per time. Through the deFBA model, there is no explicit flux for the biomass growth, and instead there are reactions for producing individual biomass components. It was shown previously that with typical reaction stoichiometries and kinetics, the intracellular metabolites  $m$  can be considered to be in quasi-steady state for this system [8]. This yields the algebraic equation

$$\dot{m} = S_m V(t) = 0, \quad (1)$$

with the stoichiometric matrix  $S_m$  which includes the stoichiometric coefficients of the metabolic species represented in the vector  $m$  in the reaction flux vector  $V$ . Collecting other species in vector  $x$ , as  $x = (z, p) \in \mathbb{R}^{n_1+n_3}$ , the mass balances are given by the differential equation

$$\dot{x}(t) = S V(t), \quad (2)$$

where  $S$  is the stoichiometric matrix with the stoichiometric coefficients of the species in  $z$  and  $p$  for reaction fluxes in the vector  $V$ .

The deFBA model also includes several biophysical constraints:

- Enzyme capacity constraints, in which the reaction fluxes are limited by the maximum enzymatic capacity defined by the amount and catalytic constant ( $k_{cat}$ ) of associated enzymes. The capacity constraint for a single enzyme  $p_i$  is given by

$$\sum_{j \in \text{cat}(i)} |V_j / k_{cat,j}| \leq p_i, \quad i \in \mathbb{E} \quad (3)$$

where  $\text{cat}(i)$  indexes the set of reactions catalysed by the enzyme  $p_i$ ,  $i \in \mathbb{E}$  in which  $\mathbb{E}$  is the set of enzymes.

- Enzyme-independent flux bounds, which for example can be used to express the irreversibility of reactions:

$$V_{min} \leq V \leq V_{max}. \quad (4)$$

- Biomass composition constraints: the macromolecules  $p$  also include non-catalytic proteins and other molecules which do not contribute directly to cellular metabolism and growth, but they are required by the cell to keep it working. These are denoted as "quota compounds" [12], and the constraints are expressed by enforcing a minimal fraction  $\varphi_Q \in [0, 1]$  of the total dry weight  $b^T p$  of the cell to be made of a quota compound  $p_Q$

$$\varphi_Q b^T p \leq p_Q, \quad Q \in \mathbb{Q} \quad (5)$$

where  $\mathbb{Q}$  is the set of quota compounds.

Considering the biomass integral over the considered time span as an objective function for the cellular dynamics [8], the deFBA model is described by the following dynamic optimization problem:

$$\begin{aligned}
& \underset{V(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} b^T p(t) dt & (6a) \\
\text{subject to } & \dot{x}(t) = (\dot{z}(t), \dot{p}(t)) = SV(t), & (6b) \\
& x(t_0) = x_0 = (z_0, p_0), \quad x(t) \geq 0 & (6c) \\
& S_m V(t) = 0, & (6d) \\
& V_{min}(t) \leq V(t) \leq V_{max}(t), & (6e) \\
& \varphi_Q b^T p(t) \leq p_Q(t), \quad Q \in \mathbb{Q} & (6f) \\
& \sum_{j \in \text{cat}(i)} |V_j(t)/k_{cat,j}| \leq p_i(t), \quad i \in \mathbb{E} & (6g)
\end{aligned}$$

where fluxes  $V$  are the dynamic optimization variables. Through the model, cellular metabolic-genetic fluxes  $V$  are assigned to maximize the biomass production subject to the given biophysical constraints. In contrast to kinetic models, where regulatory interactions from metabolites on enzyme synthesis are commonly used [13], the deFBA model uses only an optimization principle and no regulatory constraints. However, since the catalytic efficiency of enzymes as well as their biosynthesis costs are accounted for in the optimization, the model can produce behaviours that will typically be realized by regulatory interactions in the actual cells, such as catabolite repression or overflow metabolism [8].

## 2.2 Control problem formulation for maximal productivity

Constraint-based models are based on the hypothesis that the cells are maximizing a reasonable biological objective such as biomass. From a bioprocess point of view, there is often a trade-off between production of biomass and that of a certain target metabolite as a product. To describe the bioprocess optimization which is constrained by a cellular model that includes an optimization problem itself, a bilevel optimization problem is formulated. Thereby, in the outer optimization problem, the process objective is maximized, while the inner problem represents the cellular model where the cellular objective is maximized. Such a problem structure also occurs for example in strain design algorithms in metabolic engineering [14].

Metabolic engineering strategies are often based on static gene deletions to increase the product yield of the process [5, 15]. Such deletions may however decrease productivity by slowing down the biomass growth which is detrimental to the overall processing speed. To regain the bacterial growth but maintain a high yield, dynamic regulation of metabolism in a two-phase process has been proposed [16]. To this end, in [7], we have employed the deFBA model within a bilevel optimization to obtain the optimal temporal manipulation of the cellular metabolism for improved bioprocess productivity. This is implemented by

dynamic manipulation of one or several reaction fluxes such that an optimal balance between biomass growth and product formation is achieved.

Here, we are looking for a time course of the manipulated flux which leads to an increased productivity. The formulated dynamic bilevel problem for open-loop optimal control of the productivity is as follows:

$$\begin{aligned}
& \underset{t_f, u(\cdot)}{\text{maximize}} && x_t(t_f)/t_f \\
& \text{s.t.} && \underset{V(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} b^T p(t) dt \\
& && \text{s.t. (6b) to (6g),} \\
& && V_{reg}(t) = u(t).
\end{aligned} \tag{7}$$

In the outer optimization, the optimal batch time  $t_f$  and the optimal profile of the manipulated flux  $u$ , are determined to get the maximum productivity  $x_t(t_f)/t_f$  for the target metabolite  $x_t$ . The outer optimization is subject to an inner optimization (the deFBA model) which addresses maximizing the biomass. In the inner problem,  $V_{reg}$  is the manipulated flux, which is set to the control profile  $u$  determined in the outer optimization through dynamic metabolic control. The inner problem forces the unregulated fluxes to be distributed such that the biomass formation is maximized. For a specific application, dynamic manipulation of any pathway (either gene expression pathways or metabolic fluxes) can be considered through the defined problem.

### 3 MPC based on adapted metabolic-genetic model

The proposed control approach is composed of MPC and an estimation part including MHE and RBA. The MHE is used to estimate uncertain parameters of the underlying model (deFBA) and RBA to estimate the state of biomass components  $p$ . Estimated states and parameters are then used to initialize and correct the deFBA model employed within MPC. The control scheme is visualized in Figure 1 and explained in detail in the following sections.

#### 3.1 Model predictive control

Model predictive control is an optimization-based control approach which uses a model to predict the future behaviour of a dynamic system. MPC computes a trajectory of control inputs by solving an optimization problem at each sampling time [11]. In this work, we consider a closed-loop control of the bioprocess by introducing feedback to the open-loop problem (7), based on the MPC. The optimal trajectory of the control inputs is determined by repeating the bilevel optimization (7) while updating the state  $x_0$  based on online measurements. For the MPC optimization, the batch time is fixed equal to the optimal batch time  $t_f$  obtained from the open-loop optimization. As the batch runs for a finite

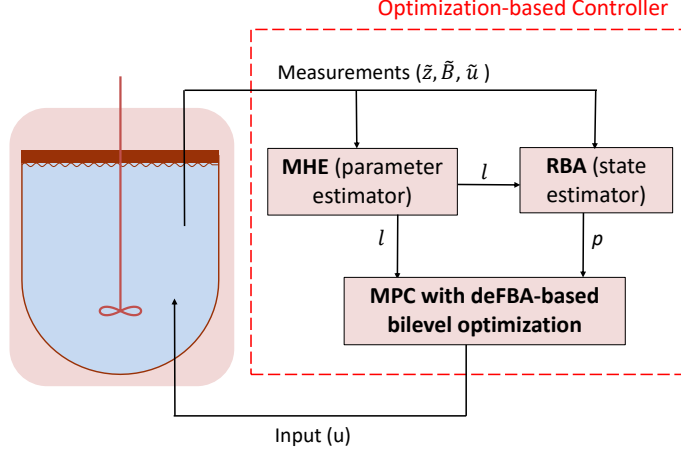


Figure 1: Adaptive MPC control scheme.

time, a shrinking horizon of length  $t_f - T$ , where  $T$  is the current time, is used for the time the optimization is performed over. A step size  $h$  is defined to discretize the batch time:

$$h = t_f/n, \quad T = ih, \quad i = 0, \dots, n-1.$$

The MPC problem has a general form of

$$\underset{\{u_k(\cdot)\}_T^{t_f}}{\text{maximize}} \quad \mathcal{F}(\{u_k(\cdot)\}_T^{t_f}, \{x_k(\cdot)\}_T^{t_f}), \quad (8)$$

where  $\{u_k(\cdot)\}_T^{t_f}$  and  $\{x_k(\cdot)\}_T^{t_f}$  stand for the sequences of control input vector and state vector  $(z, p)$ , respectively, from the current time  $T$  to the final time  $t_f$  at the  $k$ -th iteration of the MPC scheme.  $\mathcal{F}$  is the objective function.

In our case, each individual optimization problem for  $k = 1$  to  $n$  over the shrinking batch time  $[T, t_f]$  is defined as:

$$\begin{aligned} & \underset{\{u_k(\cdot)\}_T^{t_f}}{\text{maximize}} \quad x_t(t_f) \\ & \text{s.t.} \quad \underset{V(\cdot)}{\text{maximize}} \quad \int_T^{t_f} b^T p_k(t) dt \\ & \quad \text{s.t.} \quad (6b) \text{ to } (6g), \\ & \quad \quad V_{reg}(t) = u_k(t). \end{aligned} \quad (9)$$

As the batch time  $t_f$  is taken from the open-loop optimization, MPC maximizes the final concentration of the target metabolite  $x_t$  in order to have maximal productivity. Based on the problem (9), the feedback is applied by resetting the controller initial conditions using measured and estimated state variables (substrate, products and biomass components concentrations) at each iteration.

Generally, the standard MPC scheme uses nominal values of the parameters, while an adaptive MPC scheme allows to adjust the model online to compensate for time-varying process characteristics [17]. Due to the very dynamic nature of biological systems, adaptive MPC can be a suitable approach for a flexible bioprocess control. In this work we consider an adaptive MPC scheme which addresses uncertain parameters of the metabolic-genetic model within the control scheme to take biological variabilities into account; besides updating the state based on the online data, the uncertain parameters are also estimated and updated in each iteration of the MPC scheme. However, as the underlying model is relatively complex with many parameters inside, the prerequisite of the model adaptation is to identify which parameters should be considered as inputs for adjusting network dynamics. Such a parameter can be selected by sensitivity/observability analyses for any particular application and growth mode. Later, we show how the proposed adaptive MPC is capable to identify target parameters for model adjustment in each metabolic mode of the bioprocess.

For the parameter estimation, a moving horizon estimation algorithm is applied as explained in the next section.

### 3.2 Moving horizon estimation

In order to estimate unknown or uncertain parameters of the model based on the information obtained through online measurements, we implement moving horizon estimation [18, 19]. MHE minimizes the mismatch between the model outputs and the measurements over an estimation window and can therefore give an estimate of initial conditions or unknown parameters.

Uncertain parameters of the deFBA model are usually due to variations in the catalytic constants  $k_{cat}$  of metabolic or gene expression reactions, but could also be the stoichiometry of an elementary reaction or other model elements. To represent changing biological conditions, we assume that the parameters are slowly time-varying. Thus, while in each MHE problem constant parameter values are estimated for the considered horizon, the MHE problems at different iterations may result in different parameter estimates.

Here, we implement MHE to estimate the uncertain parameters of the deFBA model denoted by  $l_k$  during each MPC iteration  $k$ . From  $k = 1$  to  $n$ , each individual MHE problem can be given by the following bilevel problem:



$$\begin{aligned}
& \underset{l_k}{\text{minimize}} && \sum_{i=0}^{N_{mhe}} (B_k(T - ih) - \tilde{B}_k(T - ih))^2 + (z_k(T - ih) - \tilde{z}_k(T - ih))^2 \\
& \text{s.t.} && l_k^{min} \leq l_k \leq l_k^{max}, \\
& && \underset{V(\cdot)}{\text{maximize}} \int_{T-N_{mhe}h}^T b^T p_k(t) dt \\
& \text{s.t.} && (6b) \text{ to } (6g), \\
& && V_{reg}(t) = \tilde{u}_k(t).
\end{aligned} \tag{10}$$

In the estimation problem (10), available measurements on biomass  $\tilde{B}$ , concentrations  $\tilde{z}$  of extracellular metabolites, and the manipulated flux  $\tilde{u}$  are used to estimate uncertain deFBA parameters  $l$ . In principle the measured control input  $\tilde{u}_k$  should be the same as the input  $u_k$  determined in the MPC problem, but to account the possibility of input disturbances which can be measured we use a more general formulation here. Note that in (10), the constraints (6b)–(6g) within the inner problem are depending on the estimated parameter  $l_k$ , which is an optimization variable in the outer problem. Thereby, at each MPC iteration  $k$ ,  $l_k$  is determined from the outer optimization by minimizing the difference between the concentration of total biomass  $B_k = b^T p_k$  and/or extracellular species (substrates and products)  $z_k$  predicted by the deFBA model (the inner problem) and the ones obtained via measurements  $(\tilde{B}_k, \tilde{z}_k)$ , over previous sampling times  $T - ih$ ,  $i = 0, \dots, N_{mhe}$ , where  $N_{mhe}$  presents the number of time steps within the estimation horizon of the MHE. Note that the number of time steps  $N_{mhe}$  considered in MHE's prediction horizon is limited by the number of time points with available measurements. So, as the MPC iterations proceed one can increase  $N_{mhe}$  if it is needed for improved estimation. Upper and lower bounds  $l_k^{min}$  and  $l_k^{max}$  can be used to ensure biological plausibility of the estimated model parameters.

### 3.3 State estimation by resource balance analysis

During the MPC iterations, the states are updated by the current process measurements. To update the states, we need to have information on the concentration of biomass components  $p$  (enzymes, quota compounds and other macromolecules) individually, as an initial condition for the deFBA model. However, in practice it is only possible to measure the total biomass concentration  $B = b^T p$ . Therefore, an algorithm is needed to estimate the state of biomass components. While a rigorous state estimation for deFBA models can be conducted in principle through a mixed integer optimization [20], the required computational effort will preclude the application of that method online within an MPC scheme.

Instead of a full state estimation, we compute the optimal biomass compo-

sition by the resource balance analysis (RBA) [21] and use that as an estimate. Through the RBA problem it is assumed that the cell is at an optimal quasi steady state where for each given value of biomass, the metabolic fluxes and cellular components are allocated to maximize the cellular growth rate  $\mu$ . Note that this is consistent with the cellular objective optimization of the deFBA model in the inner problem, since maximization of the growth rate is often equivalent to the maximization of the biomass integral as in (6a).

The estimation algorithm is defined as:

$$\begin{aligned}
& \underset{\mu, V, p}{\text{maximize}} && \mu \\
& \text{subject to} && S_p V_p - \mu p = 0, \\
& && b^T p = \tilde{B}, \\
& && S_m V = 0, \\
& && V_{min}(\tilde{z}) \leq V \leq V_{max}(\tilde{z}), \\
& && \varphi_Q b^T p \leq p_Q, \quad Q \in \mathbb{Q} \\
& && \sum_{j \in \text{cat}(i)} |V_j / k_{cat,j}| \leq p_i, \quad i \in \mathbb{E} \\
& && V_{reg} = \tilde{u}.
\end{aligned} \tag{11}$$

which is a nonlinear optimization problem (due to the presence of the term  $\mu p$  in the mass balance equations of cellular components). For the measured value of the biomass  $\tilde{B}$  at each sampling time with corresponding measured value for the manipulated flux  $\tilde{u}$ , this problem gives an estimate for biomass components,  $p$  which is used to initialize the controller states in each iteration. Measurements on extracellular species ( $\tilde{z}$ ) are used to constrain fluxes by adjusting  $V_{min}$  and  $V_{max}$ ; as an example if a metabolite  $z_i$  is not present, the corresponding upper flux bound  $V_{max,i}(\tilde{z})$  is set to zero.

### 3.4 Adaptive MPC algorithm

The overall control algorithm is summarized as Algorithm 1 (in accordance to Figure 1). It should be noted that the first MPC iteration starts with the nominal value of uncertain parameters. Moreover, before the iterations of the MPC algorithm, the RBA algorithm is used to estimate the composition of the initial biomass  $p(0)$ , before applying any control, i.e., without the constraint on  $V_{reg}$  in (11).

### 3.5 Numerical solution

To solve the MPC and MHE optimization problems, we implement collocation methods to approximate the inner problem (the deFBA model) by discretization of dynamic variables in the time domain [22]. For the collocation, the time interval  $[0, t_f]$  is divided into  $n$  equally sized intervals, each containing  $K$  collocation points. The collocation points are determined by zeros of the Legendre

---

**Algorithm 1**

---

Estimate  $p(T)$  with RBA for  $T = 0$

For  $k = 1$  to  $n$

1. Given initial states  $\tilde{x}_k(T) = (\tilde{z}_k(T), p_k(T))$  and parameters  $l_k$ , solve the dynamic optimization problem (9) over  $[T, t_f]$ , which yields the optimal control sequence  $\{u_k(\cdot)\}_T^{t_f}$ .
  2. Apply the first move of control input  $\{u_k(\cdot)\}_T^{T+h}$  to the plant during the time  $[T, T+h]$ .
  3.  $T = T + h$
  4. Obtain new measurements: metabolites  $\tilde{z}_{k+1}(T)$ , total biomass  $\tilde{B}_{k+1}(T)$ , and control input  $\tilde{u}_{k+1}(T)$ .
  5. Solve the MHE problem (10) with inputs  $\{(\tilde{u}(\cdot), \tilde{x}(\cdot))\}_{T-N_{mhe}h}^T$  which yields the estimated parameters  $l_{k+1}$ .
  6. Solve the state estimation problem (11) with inputs  $\tilde{u}_{k+1}(T)$ ,  $l_{k+1}$  and measured total biomass  $\tilde{B}_{k+1}(T)$ , which gives estimated biomass components  $p_{k+1}(T)$
  7.  $k = k + 1$ , iterate.
- 

polynomials. The flux variable  $V(t)$  and the derivative of state variable  $\dot{x}(t)$  are discretized with a Lagrange interpolation scheme. The state variable  $x$  is discretized at the boundaries of the  $n$  time intervals and thus can be reset at each sampling step. By discretization of variables, the continuous optimization problem in (6) is approximated by a linear problem (LP). The detailed procedure for the discretization of variables can be found in [8].

The overall bilevel optimization problems are then solved in MATLAB. The outer optimization is performed using *fmincon*, and the inner LP optimization is solved using *linprog*. For the open-loop optimization, several initial conditions have been evaluated for the outer optimization to avoid obtaining a local minimum only. Also, for an efficient performance within the MPC, the optimal control inputs obtained from open-loop optimization are used as an initial guess for the first iteration of the closed-loop optimization. As the MPC proceeds, the optimal control values obtained from each iteration are used as the initial guess for the next iteration. This approach helps to avoid local optima during the MPC iterations.

## 4 Case study

As the case study, we consider batch growth of *E. coli* for converting glycerol into ethanol under microaerobic conditions [23]. This is a specifically interesting growth condition as the structure of the model changes in correspondence to the degree of oxygen limitation. In this case study, the proposed MPC algorithm regulates the metabolic modes in order to maximize ethanol production by adjusting the oxygen supply to the culture.

## 4.1 Network description

We constructed a reduced metabolic-genetic network model of *E. coli* including the core processes relating carbon uptake and growth. This is developed from the *in silico* model of the central *E. coli* metabolism [24]. The network includes glycerol dissimilation pathways and reactions for glycolysis, pentose phosphate pathway, anaerobic fermentation, and respiration together with appropriate production reactions for biomass components including catalytic enzymes, ribosomes and structural macromolecules. The main steps for generating such a reduced metabolic-genetic network are similar to what was presented in our other works [7, 12]. All reactions with corresponding enzymes and their catalytic constants are given in Tables 1 and 2, and for simplicity, gene IDs are used for the naming of enzymes.

The biomass components collected in the vector  $p$  and their weight vector  $b$  are listed in Table 2. The weight vector is calculated based on the stoichiometric coefficients of the respective subcomponents and their molar mass. The initial biomass composition is computed by the RBA algorithm to yield the maximum aerobic growth rate on glycerol. All biomass elements which are not catalysing reactions in this model are combined in a single quota compound  $Q$ . From the available data for *E. coli* [25], it is estimated that this quota makes up at least 55% of total biomass, which is formulated in the biomass composition constraint

$$0.55b^T p \leq Q. \quad (12)$$

The nominal values for the catalytic constants of enzymes are extracted from the enzyme database BRENDA [26] for *E.coli*, while the catalytic constants of the biomass reactions are based on the translation elongation rate of 12 amino acids/s in *E.coli* [27]. The complete deFBA model is then specified by the reactions in Tables 1 and 2 in combination with the biomass composition constraint (12). Initial nutrient conditions are summarized in Table 3.

## 4.2 Simulation cases for adaptive MPC

The proposed metabolic-genetic model is used to maximize the ethanol productivity by the MPC algorithm. A metabolic regulation is considered by manipulating the oxygen uptake rate *OUR* ( $V_{22}$  according to the numbering of metabolic pathways in Table 1).

We first describe the parameter variation that is used to generate artificial process data. To match the cellular growth rates to available experimental data, a scaling factor  $f$  on the catalytic constants of the enzymes is introduced as an adjustable parameter (Table 1). It has been experimentally shown that with *E. coli* growing microaerobically on glucose, decreases in the oxygen availability result in a decreased catabolic efficiency of the cell [28]. Our experimental studies indicate such a pattern for microaerobic *E. coli* growth on glycerol as well (data unpublished), suggesting lower metabolic activity of *E.coli* in oxygen limitation conditions compared to a fully aerobic growth. Due to this fact and through validating the metabolic-genetic network with the available experimental data,

Table 1: Metabolic part of the deFBA model: Metabolic reactions with associated enzymes and rate constants  $k_{cat}$ , scaled by the scaling factor  $f$ .

No.	Reaction	Gene/Enzyme	$k_{cat}/min^{-1}$
1	GLY + ATP + q $\rightarrow$ T3P + ADP + qH2	<i>glpK-D</i>	$1998 \cdot f$
2	GLY + PEP + NAD <sup>+</sup> $\rightarrow$ T3P + NADH + PYR	<i>dhaK</i>	$1447 \cdot f$
3	2T3P $\rightarrow$ G6P	<i>fba</i>	$630 \cdot f$
4	G6P + 2NAD <sup>+</sup> $\rightarrow$ RU5P + 2NADH + CO2	<i>gnd</i>	$1326 \cdot f$
5	RU5P $\rightarrow$ X5P	<i>rpe</i>	$78000 \cdot f$
6	RU5P $\rightarrow$ R5P	<i>rpi</i>	$3000 \cdot f$
7	X5P + R5P $\rightarrow$ S7P + T3P	<i>tkt</i>	$3402 \cdot f$
8	S7P + T3P $\rightarrow$ E4P + G6P	<i>tal</i>	$780 \cdot f$
9	E4P + X5P $\rightarrow$ T3P + G6P	<i>tkt</i>	$3402 \cdot f$
10	T3P + ADP + NAD <sup>+</sup> $\rightarrow$ PEP + NADH + ATP	<i>eno</i>	$3162 \cdot f$
11	PEP + ADP $\rightarrow$ PYR + ATP	<i>pyk</i>	$3960 \cdot f$
12	PYR + CoA $\rightarrow$ AcCoA + FOR	<i>pfl</i>	$768 \cdot f$
13	PYR + CoA + NAD <sup>+</sup> $\rightarrow$ AcCoA + NADH + CO2	<i>pdh</i>	$29160 \cdot f$
14	FOR $\rightarrow$ CO2	<i>fhl</i>	$169980 \cdot f$
15	PEP + CO2 $\rightarrow$ OAA	<i>ppc</i>	$32400 \cdot f$
16	OAA + AcCoA + NAD <sup>+</sup> $\rightarrow$ AKG + CoA + NADH + CO2	<i>acn</i>	$318 \cdot f$
17	AKG + ADP + 3NAD <sup>+</sup> $\rightarrow$ OAA + 3NADH + ATP + CO2	<i>sucCD</i>	$2684 \cdot f$
18	AcCoA + ADP $\rightarrow$ ATP + CoA + ACT	<i>ackA</i>	$800 \cdot f$
19	AcCoA + 2NADH $\rightarrow$ ETH + CoA + 2NAD <sup>+</sup>	<i>adhE</i>	$942 \cdot f$
20	ATP $\rightarrow$ ADP		
21	AKG + ATP + NADH $\rightarrow$ AA + ADP + NAD <sup>+</sup>	<i>gln</i>	$360 \cdot f$
22	O2 + 2qH2 $\rightarrow$ 2q + 8H <sup>+</sup>	<i>cyo</i>	$18000 \cdot f$
23	NADH + q $\rightarrow$ qH2 + 4H <sup>+</sup> + NAD <sup>+</sup>	<i>nuo</i>	$2220 \cdot f$
24	4H <sup>+</sup> + ADP $\rightarrow$ ATP	<i>atpH</i>	$3300 \cdot f$
25	ACT + ATP + CoA $\rightarrow$ ADP + AcCoA	<i>acs</i>	$3996 \cdot f$

Table 2: Genetic part of the deFBA model: Biomass reactions with values of weights, catalytic constants ( $k_{cat}$ ) and initial conditions for biomass components  $p(0)$ . All biomass reactions are catalyzed by ribosome R.

No.	Biomass production reactions	$b/$ ( $g\ mol^{-1}$ )	$k_{cat}/$ ( $min^{-1}$ )	$p(0)/$ ( $mM$ )
26	1657AA + 6628ATP $\rightarrow$ <i>glpK-D</i> + 6628ADP	180610	0.43	0.000044
27	3657AA + 14628ATP $\rightarrow$ <i>dhaK</i> + 14628ADP	398610	0.2	0
28	1696AA + 6784ATP $\rightarrow$ <i>fba</i> + 6784ADP	184860	0.42	0.0000061
29	1275AA + 5100ATP $\rightarrow$ <i>gnd</i> + 5100ADP	138980	0.56	0.0000032
30	225AA + 900ATP $\rightarrow$ <i>rpe</i> + 900ADP	24520	3.2	0.000000008
31	438AA + 1752ATP $\rightarrow$ <i>rpi</i> + 1752ADP	47740	1.64	0.00000091
32	1326AA + 5304ATP $\rightarrow$ <i>tkt</i> + 5304ADP	144530	0.54	0.00000041
33	634AA + 2536ATP $\rightarrow$ <i>tal</i> + 2536ADP	69110	1.14	0.0000018
34	1923AA + 7692ATP $\rightarrow$ <i>eno</i> + 7692ADP	209610	0.37	0.000025
35	1880AA + 7520ATP $\rightarrow$ <i>pyk</i> + 7520ADP	204920	0.38	0.000014
36	1766AA + 7064ATP $\rightarrow$ <i>pfl</i> + 7064ADP	192490	0.41	0
37	42096AA + 168384ATP $\rightarrow$ <i>pdh</i> + 168384ADP	4588460	0.02	0.0000015
38	2837AA + 11348ATP $\rightarrow$ <i>fhl</i> + 11348ADP	309230	0.25	0
39	3532AA + 14128ATP $\rightarrow$ <i>ppc</i> + 14128ADP	384990	0.2	0.0000006
40	1943AA + 7772ATP $\rightarrow$ <i>acn</i> + 7772ADP	211790	0.37	0.000108
41	4565AA + 18260ATP $\rightarrow$ <i>sucCD</i> + 18260ADP	497580	0.16	0.0000082
42	4291AA + 17164ATP $\rightarrow$ <i>ackA</i> + 17164ADP	467720	0.17	0
43	35640AA + 142560ATP $\rightarrow$ <i>adhE</i> + 142560ADP	3884760	0.02	0
44	5675AA + 22700ATP $\rightarrow$ <i>gln</i> + 22700ADP	618580	0.13	0.000034
45	1291AA + 5164ATP $\rightarrow$ <i>cyo</i> + 5164ADP	140720	0.56	0.0000071
46	4282AA + 17128ATP $\rightarrow$ <i>nuo</i> + 17128ADP	466740	0.17	0.000076
47	4895AA + 19580ATP $\rightarrow$ <i>atpH</i> + 19580ADP	533550	0.15	0.000116
48	652AA + 2608ATP $\rightarrow$ <i>acs</i> + 2608ADP	68130	1.1	0
49	7459AA + 9132R5P + 29836ATP $\rightarrow$ R + 29836ADP	2292420	0.1	0.000015
50	25AA + 27G6P + 10R5P + 36E4P + 12T3P + 100PEP + 283PYR + 274AcCoA + 178OAA + 3110ATP + 1361NADH $\rightarrow$ Q + 274CoA + 3110ADP + 1361NAD <sup>+</sup>	65000	28.8	0.0059

different scaling factors have been obtained for aerobic and oxygen-limited conditions. Therefore, we use variations in the scaling factor  $f$  (which addresses different rates of metabolism during the process) as a source for plant-model mismatch in our case study. We use a scaling factor of 0.9 for aerobic growth and 0.6 for all oxygen-limited growth modes to simulate the real process, while these values are not known in the control and estimation algorithms. Instead, the controller uses the metabolic-genetic model with a nominal value for the scaling factor as  $f = 1$ , assuming that there is no need to scale current  $k_{cat}$  values obtained from databases.

To implement the adaptive MPC and evaluate its performance, we consider different levels of mismatches and uncertainties by defining several cases which differ in parameters of plant-model mismatch as well as parameters considered for model adaptation. To do so, we consider three cases summarized in Table 4. Details of implementing each case are presented in following sections.

#### 4.2.1 Case 1

In the first case, we implement the adaptive MPC with adjusting the same parameter as is used to generate variations in the artificial process data. To do so, the scaling factor  $f$  is considered as the mismatch parameter which is estimated through the MHE in each MPC iteration and used to adapt the underlying model for the next MPC iteration. We call this the ideal case, as the MPC adjusts the exact parameter of the plant-model mismatch.

#### 4.2.2 Case 2

For the second case, we consider a different parameter set for the adaptation in the control scheme than what is used in the process simulations. Therefore, we leave the value of the scaling factor as its nominal value  $f = 1$  in the model used within the control scheme and instead try to adjust the individual catalytic constants  $k_{cat}$  directly in the MHE. In order to catch different rates of metabolism during the microaerobic process (resulted from different scaling factor for aerobic and oxygen-limited growth in the real plant simulation), one needs to adjust enzymatic constants of all metabolic pathways. To limit the computational complexity with respect to the number of parameters to be estimated by MHE, we follow a parameter selection procedure which identifies the most influential metabolic reactions in each microaerobic stage and then estimates enzymatic constants for these instead of adjusting the constants of all reactions. For the parameter selection, we perform a local sensitivity analysis in each MPC iteration which identifies pathways for which a manipulation of the catalytic constants has the highest impact on the cellular growth rate. To do so,

Table 3: Initial nutrient and initial biomass  $b^T p(0)$ .

Glycerol	$b^T p(0)$
255 mM	0.59 $gl^{-1}$

Table 4: Mismatch and adaptation parameters in each simulation case.

Simulation case	Case 1	Case 2	Case 3
Parameter of plant-model mismatch	$f$	$f$	$f$ and $k_{cat}$ 's
Parameter for model adjustment in MHE	$f$	$k_{cat}$ 's	$f$ or $k_{cat}$ 's

we perturb the value of each catalytic constant individually by  $k_{cat,new} = k_{cat}/A$  and  $k_{cat,new} = k_{cat}A$ , with a prespecified value of  $A > 0$ , and evaluate the effect of each perturbation on the growth rate. Parameters with higher sensitivity are then selected for model adjustment and their values are estimated through MHE. Such a sensitivity analysis is done over the whole parameter set in each MPC iteration based on the current state (measured concentrations and *OUR*) of the process.

For the parameter estimation through MHE, we implement a sequential optimization in which the catalytic constants are estimated based on their priority (resulting from the sensitivity analysis). To do so, a stopping criterion is defined by placing a tolerance on the objective function of the MHE (10) which determines a satisfactory level for matching model predictions to the measurements. Considering  $l$  as the vector containing ordered parameters, the procedure can be summarized as in Figure 2. The first sequence starts with the estimation of the first identified parameter  $l(1)$ . If the termination criterion is not met through the first optimization, the algorithm proceeds to the next sequence in which the first two identified parameters are estimated  $l(1 : 2)$ . In this way with an increased number of parameters in each sequence, MHE estimates as many parameters as needed to reach the termination criterion on the model-data difference, or until the list of parameters used for adaption is exhausted.

Therefore, in each MPC iteration a different set of catalytic constants  $k_{cat}$  is estimated by sequential MHE problems depending on their importance in each microaerobic stage. By the implemented sequential scheme, we are able to track the minimum number of parameters which are required to address process dynamics at each stage. However, one can also consider a simultaneous scheme for MHE in which a larger set of identified constants are estimated simultaneously through a single problem.

#### 4.2.3 Case 3

In this case, we consider a different parameter set for the adaptation in the control scheme than what is used in the plant simulations (such as Case 2), but with considering higher levels of uncertainties and modeling errors. To do so, we consider plant-model mismatch not only in  $f$  but also in all catalytic constants  $k_{cat}$ . This means different scaling factor  $f$  and catalytic constants  $k_{cat}$  for metabolic pathways are used for the real process simulation compared to those used in the control model (nominal values). Then the model is adapted by adjusting some selected catalytic constants (identified by the sensitivity analysis and estimated through the sequential MHE optimizations as explained in Case



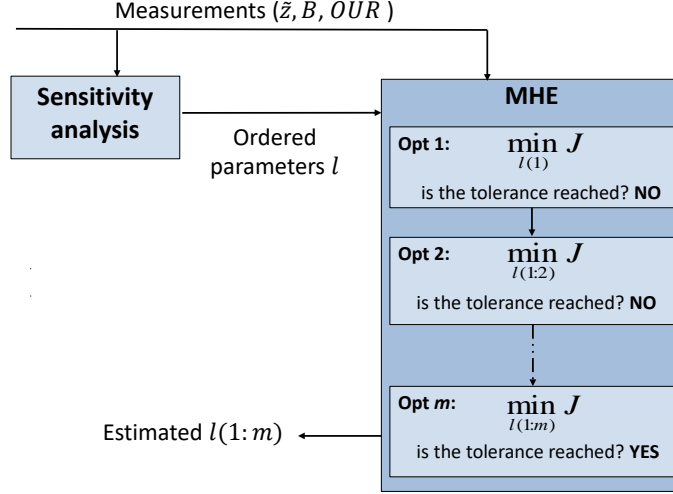


Figure 2: Rate constants selection and estimation procedure.

2), or only by adjusting one general parameter  $f$ .

## 5 Results and discussions

To have maximal ethanol productivity through the introduced *E. coli* model, the open-loop optimization problem (7) is firstly applied to obtain the optimal batch time  $t_f$  and oxygen uptake rate  $OUR(t)$ . By implementing the open-loop problem (7) on the model discussed in Section 4, the final batch time is obtained as 14.9 hr for maximal ethanol productivity. The optimization results are shown in Figure 3. The bilevel optimization results in a decreasing  $OUR$  pattern in control of the microaerobic ethanol production. It suggests an initial aerobic phase of the process with gradual movement to the oxygen-limited conditions for improved process productivity, which in fact shows balancing between the aerobic condition favoring the cell growth and the anaerobic condition favoring ethanol formation.

The final batch time from the open-loop optimization (14.9 hr) is then used within the shrinking horizon MPC problem (9), but the  $OUR$  is adjusted online to yield the maximum final ethanol concentration while compensating for the plant-model mismatch. This is implemented on three defined cases in Section 4.2.

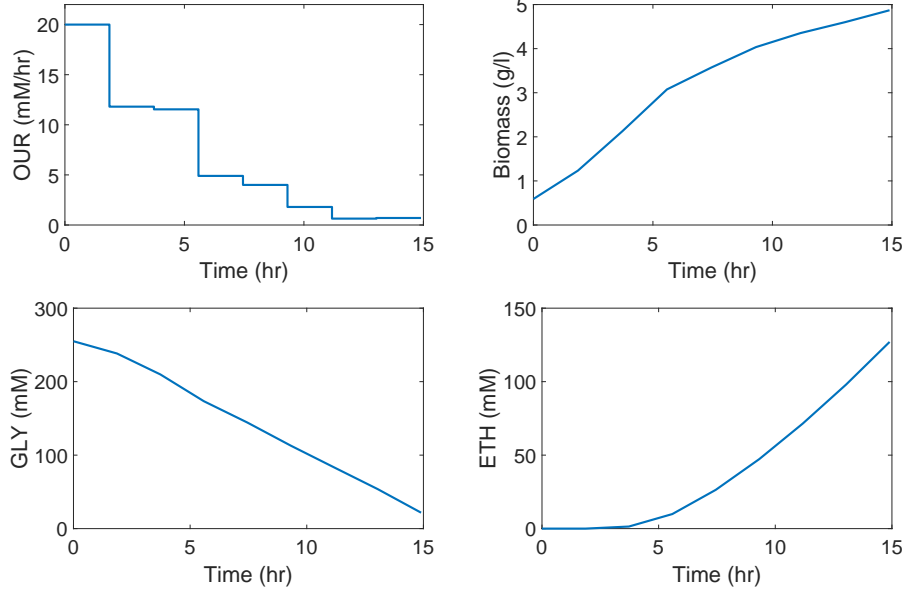


Figure 3: Optimal *OUR* pattern and concentration profiles of biomass, glycerol and ethanol, resulted from the open-loop optimization.

### 5.1 Case 1

For this case, we implement the MPC approach where the model is adapted online by adjusting the uncertain time-variable parameter  $f$ . Note that for the first iteration of the MPC, we use a nominal value for the scaling factor as  $f = 1$ . Figures 4 to 6 show the overall performance of the adaptive MPC (as in Algorithm 1) over the MPC without model adaptation in which  $f$  is considered constant all the time and equal to the nominal value of the parameter ( $f = 1$ ). Figure 4 shows the parameter estimation results; as the microaerobic growth proceeds,  $f$  decreases from 1 to around 0.6 as its value in the plant model. In order to compensate for the decreased  $f$  during the microaerobic growth (and therefore lower overall growth rate), the adaptive MPC proposes higher values of *OUR* during the process compared to the non-adaptive one. That means it allows a sufficient amount of biomass to be produced by keeping higher levels of aeration. Therefore, it predicts better consumption of the substrate, and as the result a higher amount of ethanol is produced (49 mM) compared to the non-adaptive controller (40.3 mM).

### 5.2 Case 2

Through the second case, we aim to improve the flexibility of the adaptive approach in addressing different growth modes and handling existing uncertainties.

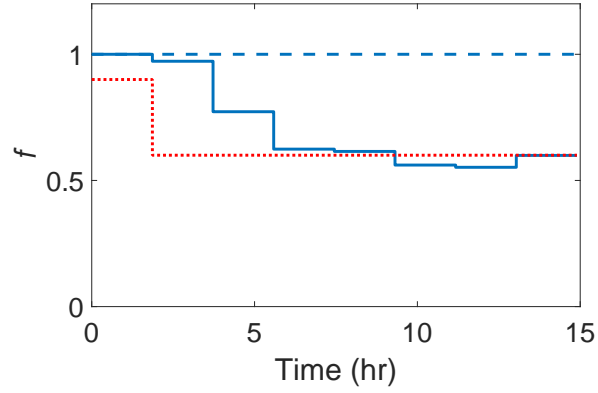


Figure 4: Scaling factor values estimated in adaptive MPC/Case 1 (solid line) and used in non-adaptive MPC (dashed line). The real system values are shown in red.

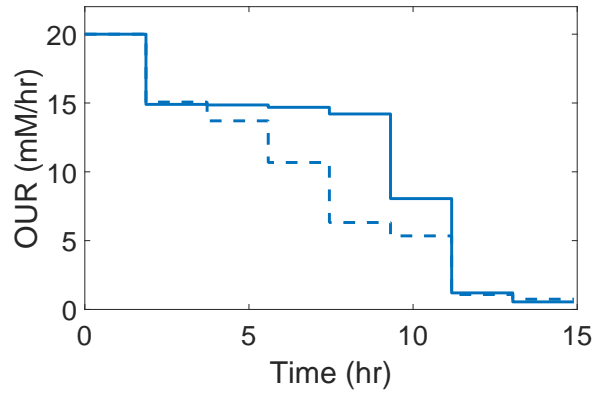


Figure 5: *OUR* pattern from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

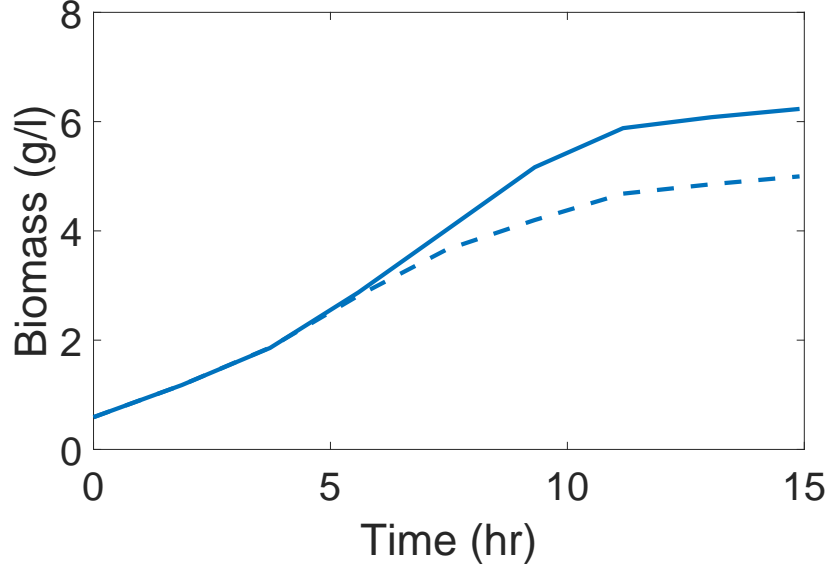


Figure 6: Biomass, glycerol and ethanol concentration profiles resulted from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

To address different rates of metabolism in the microaerobic growth, in this case we do not specify *a priori* which parameters are the exact source of the uncertainty and mismatch, and instead go for adjusting only important parameters. To do so, we adjust only the catalytic constants selected by a sensitivity analysis through the sequential MHE optimizations (as explained in Section 4.2.2) in each MPC iteration.

Here, we implement the MPC approach where the scaling factor  $f$  is the parameter of plant-model mismatch (as in Case 1) but the model is adapted by adjusting individual catalytic constants. To identify target parameters to be estimated through sequential MHE problems in each iteration the sensitivity analysis is performed; the values of all catalytic constants are individually perturbed to  $k_{i,new} = k_i/5$  and  $k_{i,new} = 5k_i$ , iterating over all metabolic reactions in order to evaluate the effect of each perturbation on the growth rate at the current state. Parameters with higher sensitivity are then estimated and used to adapt the model.

We evaluate the performance of the adaptive MPC in this case over Case 1 and also MPC without model adaptation (in which nominal values of catalytic constants are used), shown in Figures 7 and 8. As can be seen in Figure 7, the *OUR* values predicted by Case 2 are mostly slightly lower than those from Case 1 but still high enough for better consumption of glycerol and higher production of ethanol compared to the non-adaptive MPC.

Table 5 shows the final concentration of ethanol from adaptive MPC (Case

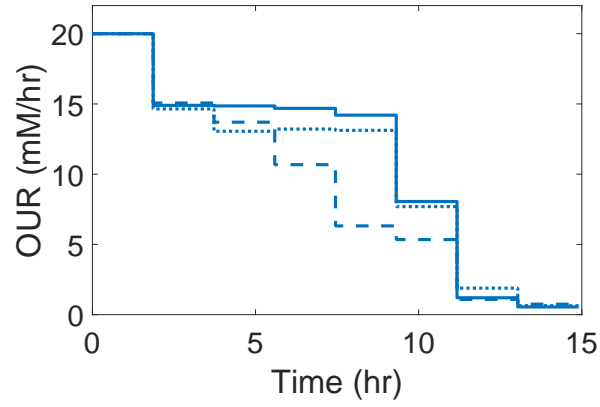


Figure 7: *OUR* pattern from adaptive MPC/Case 2 (dotted line), adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

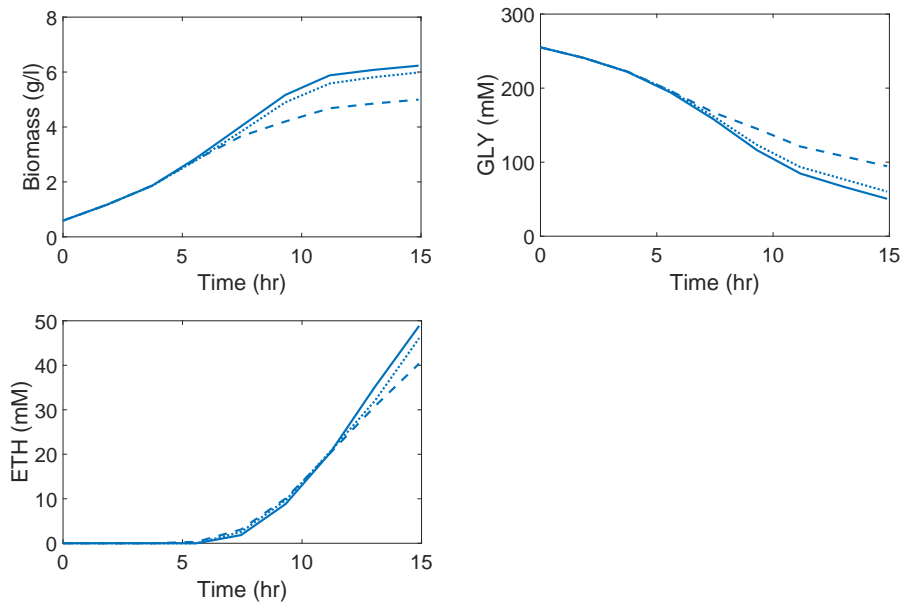


Figure 8: Biomass, glycerol and ethanol concentration profiles resulted from adaptive MPC/Case 2 (dotted line), adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

1 and Case 2) and non-adaptive MPC. We see the ethanol concentration from Case 2 is lower compared to Case 1 (as the ideal case which directly adjusts the scaling factor, that is the source of plant-model mismatch), but the value is still high compared to the non-adaptive one.

Figure 9 shows per MPC iteration the enzymes for which the catalytic constants were identified as sensitive parameters and were adjusted in the adaptation step. Each column is labelled by the corresponding aeration level (at each MPC iteration) and includes identified enzymes (by the sensitivity analysis), of which the catalytic constants are estimated through MHE and are used to adapt the deFBA model in the following iteration of MPC. Moreover, in each column enzymes are ordered from the bottom to the top according to the order of their selection from the sensitivity analysis (bottom with the most sensitive  $k_{cat}$  towards the top with less sensitive ones). As shown, in each iteration only few parameters need to be estimated and used for model adjustment and there is no need to adapt the catalytic constants of all metabolic pathways. Moreover, one can notice that as the oxygen-limited growth proceeds, there is a gradual movement from aerobically active enzymes to anaerobically active ones. Constants of enzymes *nuo* and *atpH* (responsible for respiration) are adjusted in the initial aerobic phase and earlier oxygen limited phases. The catalytic constant of enzyme *glpK-D* (responsible for glycerol utilization in aerobic condition) is relevant during initial stage of the oxygen-limited condition, while the constant of enzyme *dhaK* (responsible for glycerol utilization in anaerobic condition) is adjusted in later stages. Moreover, the constants of enzymes *pfl* (used for AcCoA production in anaerobic condition), *ackA* (responsible for acetate production), and *adhE* (responsible for ethanol production) are adjusted when a higher level of oxygen limitation is applied.

In fact switches between pathways in different metabolic modes show the importance of the parameter identification. For a better interpretation, we provide aerobic and anaerobic glycerol utilization fluxes during the process (resulted from the open-loop optimization) in Figure 10. While the process transitions to the microaerobic growth the contribution of the *glpK-D*-catalyzed pathway is decreased while the pathway catalyzed by *dhaK* contributes as the major pathway for glycerol utilization (compatible with experimental studies on microaerobic growth of *E.coli* on glycerol, [23], regarding the transition between respiratory and oxygen-limited utilization of glycerol). That is why the catalytic constant of enzyme *glpK-D* is the relevant parameter in the earlier stage of the process while the constant of enzyme *dhaK* is relevant in later stages.

Table 5: Final ethanol concentration resulted from different MPC schemes

MPC scheme	Ethanol concentration (mM)
Adaptive, Case 1	49
Adaptive, Case 2	46.1
Non-adaptive	40.3

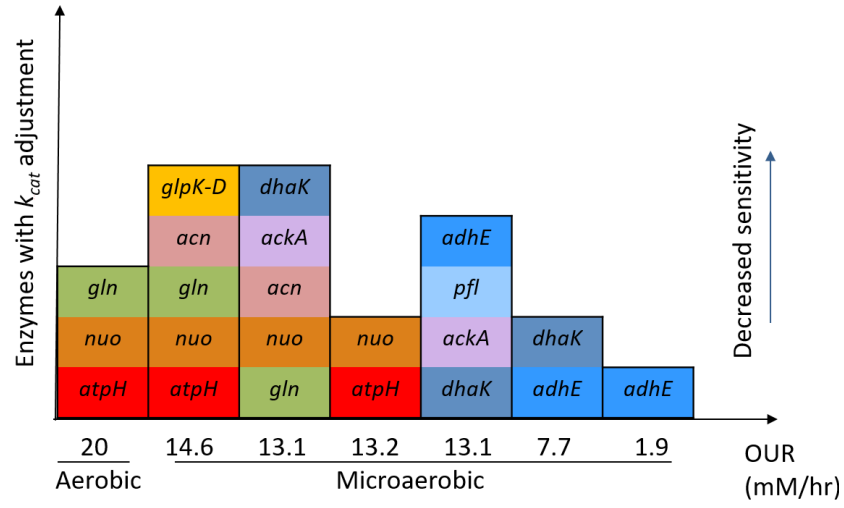


Figure 9: Enzymes with estimated constants at different *OUR* levels of the process corresponding to the MPC iteration ( $k = 1, \dots, n - 1$ ): *glpK-D* (reaction 1), *dhaK* (reaction 2), *pfl* (reaction 12), *acn* (reaction 16), *ackA* (reaction 18), *adhE* (reaction 19), *gln* (reaction 21), *nuo* (reaction 23), *atpH* (reaction 24).

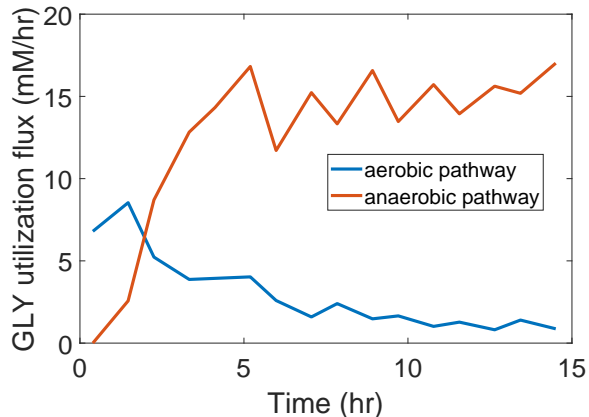


Figure 10: Aerobic and anaerobic glycerol utilization fluxes during the considered microaerobic process (resulted from the open-loop optimization).

So, the most sensitive parameters in each growth mode are identified in order to represent the process dynamics correctly.

The results obtained from Case 2 show the flexibility of the proposed adaptive approach to address different metabolic modes by adjusting a minimum number of relevant parameters.

### 5.3 Case 3

For further evaluation of the performance of adaptive MPC in handling existing uncertainties, in Case 3 higher levels of modeling error in rate-relevant parameters are considered than Cases 1 and 2. To do so, we perturb all catalytic constant by adding 20% uniform white noise to each  $k_{cat}$  value in the plant simulation. This means we use variations in  $f$  and each individual  $k_{cat}$  to simulate the real process while nominal values of these parameters are used in the controller.

For the model adaptation in this case with modeling errors in all rate-relevant parameters, we do not adjust all those uncertain parameters and instead go for adjusting only important parameters. To do so, we attempt to adapt the underlying model by (A) adjustment of influential catalytic constants (as done in Case 2 based the procedure explained in Section 4.2.2), and (B) adjustment of  $f$  as an overall rate-relevant parameter.

We evaluate the performance of the adaptive MPC over the non-adaptive one by implementation of several random noise sets on  $k_{cat}$  values. Through Case 3(A) with adjustment of the influential catalytic constants, as shown in Figure 11 A, we see the adaptive MPC mostly results in higher production of ethanol compared to the non-adaptive MPC. Among those we also notice cases



in which there is negligible improvement in final ethanol concentration from the adaptive MPC, but with higher product yield compared to the non-adaptive one (as can be seen in Noise set 3). This observation is most likely because in some cases implementing perturbations in enzymatic constants result in a limited capacity for ethanol production of the cell, such that by implementing the MPC with the model adaptation the final ethanol concentration can not be improved compared to the non-adaptive one. Although, we see that the adaptive MPC adjusts the *OUR* level for higher yield by avoiding unnecessary biomass production, such an improvement in the yield by the adaptive MPC can not be guaranteed, as within our control problem improving product yield is not directly considered as an objective.

For the next step, again we apply several random noise sets on  $k_{cat}$  values, but now the adaptive MPC scheme estimates only the overall scaling factor  $f$  through Case 3(B). As shown in Figure 11 B, we see that the adaptive MPC does not necessarily improve the ethanol production, as for some of the noise sets it results in a lower value of the final ethanol concentration compared to the non-adaptive MPC. Such an observation is not surprising, as with such a highly uncertain system and adaptation of only one model parameter, the controller can not capture the process dynamics correctly. Therefore, one should not expect the adaptive MPC to improve the overall control performance while proper parameters are not chosen for model adjustment.

Based on the results from the higher levels of plant-model mismatch considered in this case, we argue that even with a very uncertain system the adaptive MPC allows for improved performance, however it still depends on a good selection of the uncertain parameters to be adjusted.

The average computation time necessary for each iteration of the adaptive MPC in the cases of one general parameter adjustment (e.g. Case 1 with adaptation of  $f$ ) is 74 s on a standard desktop computer, while this time increases to 561 s for the case of selected  $k_{cat}$  adjustment (e.g. Case 2). The more computation time in Case 2 is mainly due to the larger number of parameters to be estimated through the sequential MHE problems in each iteration compared to Case 1 in which one parameter is estimated through a single MHE problem. In this work  $n = 8$  iterations are considered for the MPC optimization. Increasing the number of MPC iterations (decreased sampling time) would result in improved performance as it speeds up the parameter correction procedure, but it results in higher computational time.

#### 5.4 MPC with and without biomass state estimation

Next, the effect of using the RBA method as an approximate steady state estimation is evaluated. To this end, we compare the performance of the adaptive MPC for the case of estimating the state of biomass components against the case where full state information would be available. In previous sections, the MPC used the RBA algorithm through the problem (11) in order to estimate the biomass composition. Here, we compare the control performance with state estimation to an ideal situation where measurements for all biomass components would be available for Case 1 with adjustment of  $f$ . As it is shown in Figure 12,

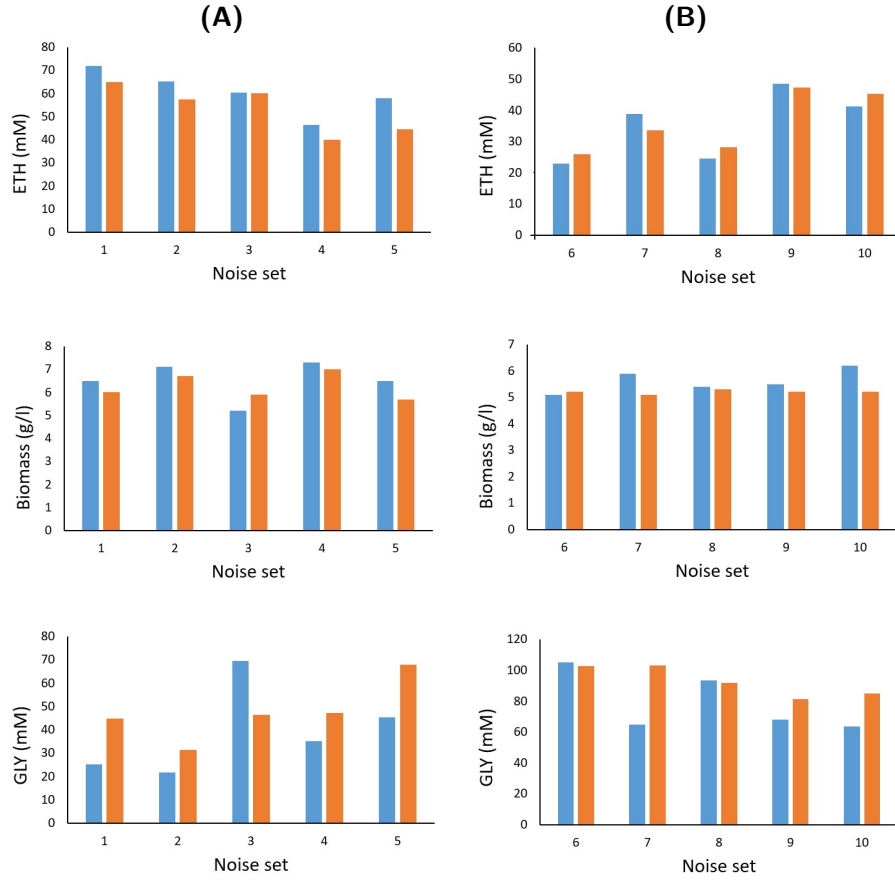


Figure 11: Ethanol, biomass and glycerol concentration terminal values resulting from adaptive MPC (blue bars) and non-adaptive MPC (orange bars) for different noise sets applied on  $k_{cat}$  values with the adjustment of (A) influential  $k_{cat}$  values and (B) the scaling factor  $f$  value.

there are some differences in the estimated  $f$ , predicted  $OUR$  and concentration trajectories of species. However, it can be seen that despite these differences the control actions in both cases yield a very similar final value of produced ethanol.

To track the biomass composition during the growth in adaptive MPC (Case 1) with feedbacks obtained from direct measurement of biomass components and from the RBA estimation, Figure 13 represents the percentage of some individual key enzymes. As can be seen, while there are some quantitative mismatches, the patterns of estimated enzyme levels are qualitatively following the true model values; the percentage of enzymes *nuo* and *atpH* (used for aerobic respiration) is high at the initial aerobic phase of the growth and then starts to decrease in following microaerobic phases. On the other hand, as microaerobic growth proceeds the percentage of enzyme *adhE* is increased in order to promote producing ethanol. While the process switches from aerobic to microaerobic growth, the percentage of enzyme *glpK-D* (aerobic utilization of glycerol) decreases while the percentage of enzyme *dhaK* (anaerobic glycerol utilization) increases (as observed in Section 5.2 as well). Similarly, there is a down-regulation of *pdh* (responsible for aerobic AcCoA production) and up-regulation of *pfl* (used for anaerobic AcCoA production) when transferring from aerobic to oxygen-limited conditions. It should be noted that these patterns for contribution of cellular components in different stages of the growth are directly related to both metabolic and genetic part of the deFBA model including enzyme production costs and their constants.

Despite the qualitative similarities, the RBA method does not provide a quantitatively reliable estimation of the biomass components in this case. Nevertheless, this estimation error has only a negligible effect on the controller performance in this study: in fact, the optimal control action and the controller predictions of the objective value (final ethanol concentration) are consistent between the two cases. We conclude that, although RBA gives only a qualitative estimation of the enzyme levels in each microaerobic stage (as it is a steady state approach), it could still be used as an estimator of biomass components within the proposed control scheme.

## 6 Conclusions

Constraint-based approaches are useful for bioprocess optimization and control as they address the full metabolic versatility compared to unstructured models. However, despite the additional metabolic details integrated in these models, there remains a high level of model uncertainties which limits their application in bioreactor control. Moreover, biological variability additionally limits the validity range of these models. In this work, we implement a combination of model predictive control and moving horizon estimation for an adaptive and flexible control of the bioprocess, based on dynamic resource allocation models. In particular, improving bioprocess productivity in microaerobic growth regimes is considered. By considering several simulation cases, we demonstrate the use-

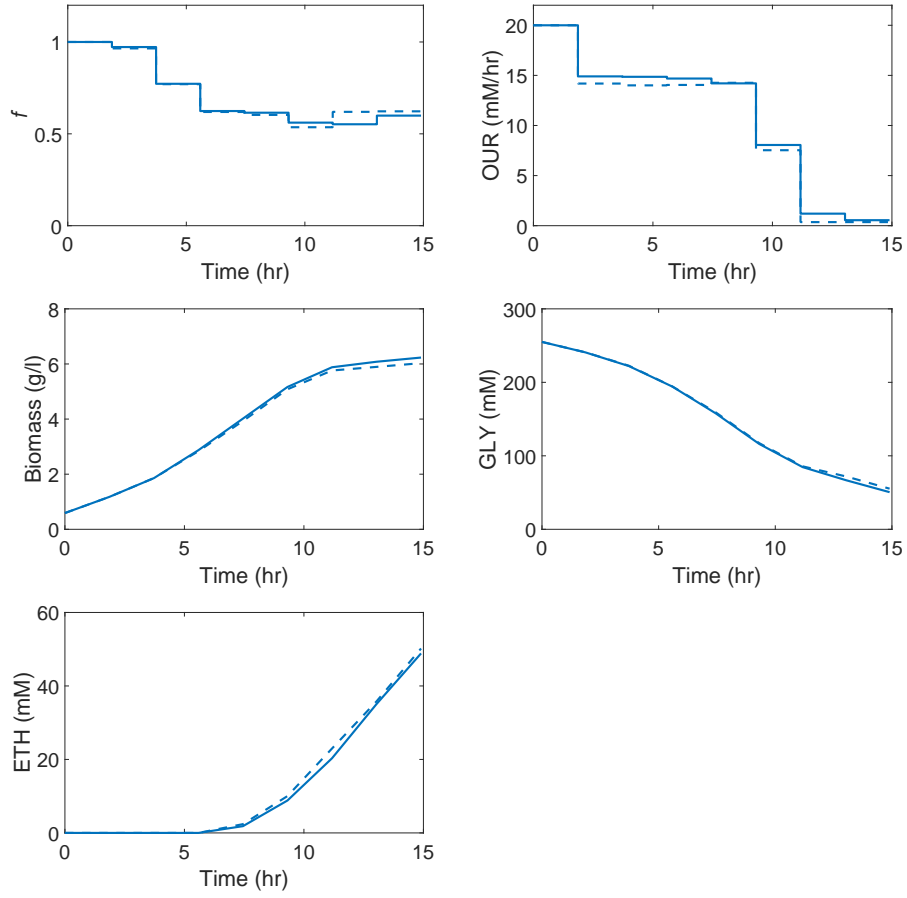


Figure 12: Performance of adaptive MPC/Case 1 with state estimation (solid line) and with full state information (dashed line).

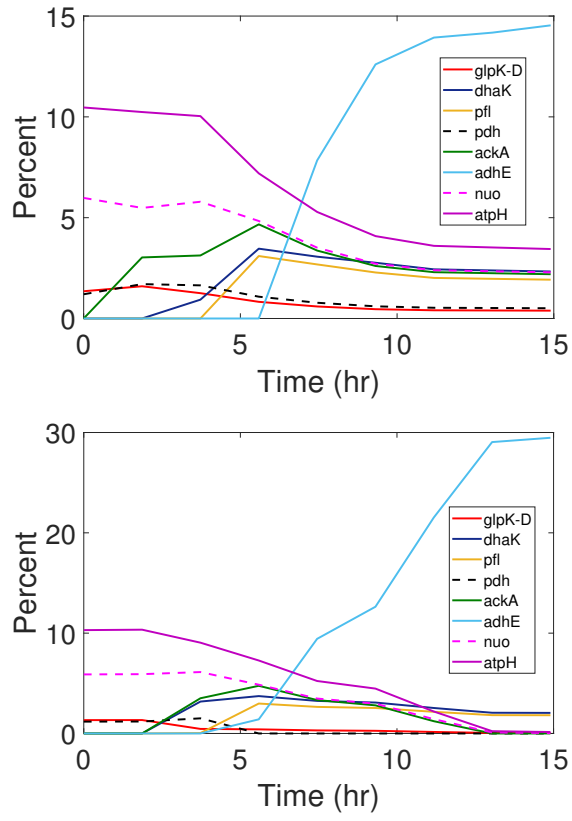


Figure 13: Biomass composition in adaptive MPC/Case 1 measured from the plant model (top) and estimated by the RBA (bottom).

fulness of the proposed adaptive control approach to address different biological states during the process and to handle associated uncertainties. Moreover, the proposed approach could capture the process dynamics by adjusting a few parameters through integrating a parameter identifiability analysis. Based on the results, the adaptive MPC adjusts the control inputs to appropriately balance between the biomass growth and the target product formation for an optimal product formation compared to a non-adaptive approach.

The results of our simulation study show that adaptive MPC with a selection of relevant parameters, as in the simulation cases 2 and 3(A), is a promising approach in bioprocess control. As a comparison, we have seen in Case 3(B) that adjusting a single parameter will not be sufficient to ensure performance in the case of plant-model mismatches for a complex bioprocesses. Here, constraint-based models are useful because they offer a high flexibility and sufficient degrees of freedom to represent a wide range of biological behaviours, in contrast to for example unstructured models which generally use much fewer parameters. However, due to the initially large number of parameters, and not all of them being relevant for the process dynamics, a selection of important parameters to be used in the adaptation needs to be done. Here, we have implemented a sensitivity analysis which is done over the whole parameter set in each MPC iteration to identify sensible parameters at that specific microaerobic level. However, it should be noted that for larger networks with high-dimensional parameter sets it can not be computationally efficient to do the sensitivity analysis over the whole parameter set in each iteration and instead one should consider performing the parameter selection offline in order to identify a set of most important parameters. In that case, even more precise parameter selection methods (e.g. the global sensitivity analysis approach proposed in [29]) for the pre-selection of parameters under various conditions and inputs can be implemented. This subset of parameters then can be considered for estimation and model adjustment during MPC optimizations, resulting in a reduced total computation cost of the MPC.

Overall, we argue that the proposed deFBA-based control scheme is a suitable approach to control not only microaerobic but any bioprocesses with switches in the dynamics of the system. Desired engineering objectives can be addressed by the proposed deFBA-based bilevel approach through temporal manipulations of the metabolism while process uncertainties can be handled efficiently using the adaptive nature of the implemented control scheme. Further work includes the experimental implementation of the proposed approach for control of the microaerobic growth of *E. coli* in a lab-scale bioreactor.

**Funding:** This work was supported by the EU-Programme ERDF (European Regional Development Fund) within the Research Center of Dynamic Systems (CDS).

## References

- [1] C. Brosilow and B. Joseph. *Techniques of Model-based Control*. Prentice-Hall International Series in the Physical and Chemical Engineering Sciences. Prentice Hall, 2002.
- [2] Amit Varma and Bernhard O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12:994–998, 1994.
- [3] Radhakrishnan Mahadevan, Jeremy S. Edwards, and Francis J. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, 83 3:1331–1340, 2002.
- [4] Steffen Klamt, Oliver Haedicke, and Axel von Kamp. *Stoichiometric and Constraint-Based Analysis of Biochemical Reaction Networks*, pages 263–316. Springer International Publishing, Cham, 2014.
- [5] Anthony P. Burgard, Priti Pharkya, and Costas D. Maranas. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.
- [6] Kapil G. Gadkar, Francis J. Doyle III, Jeremy S. Edwards, and Radhakrishnan Mahadevan. Estimating optimal profiles of genetic alterations using constraint-based models. *Biotechnology and Bioengineering*, 89(2):243–251, 2005.
- [7] Banafsheh Jabarivelisdeh and Steffen Waldherr. Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming. *Biotechnology and Bioengineering*, 115(7):1829–1841, 2018.
- [8] Steffen Waldherr, Diego A. Oyarzún, and Alexander Bockmayr. Dynamic optimization of metabolic networks coupled with gene expression. *Journal of Theoretical Biology*, 365:469–485, 2015.
- [9] Liang Chang, Xinggao Liu, and Michael A. Henson. Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. *Journal of Process Control*, 42:137–149, 2016.
- [10] Banafsheh Jabarivelisdeh, Rolf Findeisen, and Steffen Waldherr. Model predictive control of a fed-batch bioreactor based on dynamic metabolic-genetic network models. *IFAC-PapersOnLine*, 51(19):34–37, 2018. 7th Conference on Foundation of Systems Biology in Engineering FOSBE 2018.
- [11] James B. Rawlings, David Q. Mayne, and Moritz M. Diehl. Model predictive control: Theory, computation, and design. *Nob Hill Pub*, 2009.

- [12] Alexandra-M. Reimers, Henning Lindhorst, and Steffen Waldherr. A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites*, 7(3), 2017.
- [13] Madalena Chaves and Diego A Oyarzún. Dynamics of complex feedback architectures in metabolic pathways. *Automatica*, 99:323–332, 2019.
- [14] Ali R. Zomorodi, Patrick F. Suthers, Sridhar Ranganathan, and Costas D. Maranas. Mathematical optimization applications in metabolic networks. *Metabolic Engineering*, 14(6):672 – 686, 2012.
- [15] Stephen S. Fong, Anthony P. Burgard, Christopher D. Herring, Eric M. Knight, Frederick R. Blattner, Costas D. Maranas, and Bernhard O. Palsson. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnology and Bioengineering*, 91(5):643–648, 2005.
- [16] Nikolaos Anesiadis, William R. Cluett, and Radhakrishnan Mahadevan. Dynamic metabolic engineering for increasing bioprocess productivity. *Metabolic Engineering*, 10(5):255–266, 2008.
- [17] Veronica Adetola, Darryl DeHaan, and Martin Guay. Adaptive model predictive control for constrained nonlinear systems. *Systems and Control Letters*, 58(5):320–326, 2009.
- [18] James B. Rawlings. *Moving Horizon Estimation*, pages 1–7. Springer London, London, 2013.
- [19] Peter Klaus Findeisen. *Moving Horizon State Estimation of Discrete Time Systems*. University of Wisconsin–Madison, 1997.
- [20] Steffen Waldherr. State estimation in constraint based models of metabolic-genetic networks. In *2016 American Control Conference (ACC)*, pages 6683–6688. IEEE, 2016.
- [21] Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218, 2011.
- [22] Lorenz T. Biegler. An overview of simultaneous strategies for dynamic optimization. *Chemical Engineering and Processing: Process Intensification*, 46(11):1043–1053, 2007.
- [23] Guyton Durnin, James Clomburg, Zeno Yeates, Pedro J.J. Alvarez, Kyríacos Zygorakis, Paul Campbell, and Ramon Gonzalez. Understanding and harnessing the microaerobic metabolism of glycerol in *Escherichia coli*. *Biotechnology and Bioengineering*, 103(1):148–161, 2009.
- [24] Oliver Haedicke and Steffen Klamt. Ecolicore2: A reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Scientific Reports*, 7:39647, 2017.



- [25] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38:D750–D753, 2009.
- [26] Ida Schomburg, Antje Chang, Sandra Placzek, Carola Söhngen, Michael Rother, Maren Lang, Cornelia Munaretto, Susanne Ulas, Michael Stelzer, Andreas Grote, Maurice Scheer, and Dietmar Schomburg. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Research*, 41:D764–D772, 2013.
- [27] Xiongfeng Dai, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Pat-salo, Hiroyuki Okano, James R. Williamson, Kurt Fredrick, Yiping Wang, and Terence Hwa. Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nature Microbiology*, 2:16231, 2016.
- [28] Svetlana Alexeeva, Klaas J. Hellingwerf, and M. Joost Teixeira de Mat-tos. Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *Journal of Bacteri-ology*, 185(1):204–209, 2003.
- [29] Yunfei Chu and Juergen Hahn. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE Journal*, 53(11):2858–2870, 2007.