Bioprocess Optimization and Control Using Dynamic Constraint-Based Models

Dissertation
zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

von
Banafsheh Jabarivelisdeh
geboren am 19. 09 1987 in Amol-Iran

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik der
Otto-von-Guericke-Universität Magdeburg

Gutachter:
Prof. Dr.-Ing. Steffen Waldherr
Prof. Dr.-Ing. Rolf Findeisen
Prof. Dr.-Ing. Stefan Streif

eingereicht am 13. Januar 2021
Promotionskolloquium am 1. März 2021

Contents

List of figures VII

List of tables XI

List of symbols and abbreviations XIII

Abstract XVII

Deutsche Kurzfassung XIX

1 Introduction 1

1.1 Model-based optimization and control of bioprocesses 1
  1.1.1 Modeling of biological systems 1
  1.1.2 Application of model-based approaches in biotechnology 2
1.2 Contribution and outline of the thesis 3

2 Constraint-based models of metabolic networks 7

2.1 Constraint-based modeling and flux balance analysis 7
2.2 Dynamic flux balance analysis 10
2.3 Metabolic-genetic network models 12
  2.3.1 Dynamic enzyme-cost flux balance analysis 14
  2.3.2 Numerical solution 18
  2.3.3 Resource balance analysis 20
  2.3.4 General procedure for deriving a metabolic-genetic networks model 21

3 Bioprocess optimization and control by dynamic metabolic-genetic network models 25

3.1 Application of constraint-based modeling in bioprocess optimization and control 25
3.2 FBA-based approaches for metabolic engineering 26
3.3 Dynamic control for improved bioprocess productivity 29
  3.3.1 Dynamic metabolic engineering 31
  3.3.2 Dynamic optimization of fermentation condition 32

3.4 Dynamic control for improved bioprocess productivity based on deFBA model . . . 32
  3.4.1 Dynamic optimization problem formulation based on deFBA model . . . 33
  3.4.2 Solution procedure . . . 35
3.5 Application: improved ethanol productivity in *E. coli* . . . 36
  3.5.1 Network description . . . 36
  3.5.2 Control strategies for dynamic genetic- and process-level manipulations . . . 40
3.6 Results and discussion . . . 43
  3.6.1 Higher productivity by dynamic strategies . . . 43
  3.6.2 Different manipulation strategy and productivity by each variant of Strategy 1 . . . 43
  3.6.3 Robustness of a selected manipulation strategy . . . 46
  3.6.4 Optimal cellular resource allocation during the two-stage growth 48
  3.6.5 Computational limitations . . . 49
3.7 Conclusion . . . 53

4 Model predictive control of the bioprocess based on dynamic constraint-based models 55
4.1 Closed-loop control of the bioprocess . . . 55
4.2 Optimization and control of a fed-batch bioreactor . . . 56
  4.2.1 Fed-batch growth simulation . . . 57
  4.2.2 Control problem for productivity maximization . . . 58
4.3 Closed-loop control based on MPC . . . 58
  4.3.1 Model predictive control . . . 59
  4.3.2 State estimation by resource balance analysis . . . 60
  4.3.3 Linearized MPC . . . 61
  4.3.4 Overall MPC algorithm . . . 62
  4.3.5 Numerical solution . . . 62
4.4 Application: fed-batch growth of *E. coli* for improved ethanol productivity 63
4.5 Results and discussion . . . 64
  4.5.1 MPC performance with parametric errors . . . 65
4.6 Conclusion . . . 67

5 Adaptive predictive control of the bioprocess using constraint-based modeling and estimation 71
5.1 Adaptive control of the bioprocess . . . 71
5.2 MPC based on adapted metabolic-genetic model . . . 72
  5.2.1 Adaptive model predictive control . . . 73
  5.2.2 Moving horizon estimation . . . 74

5.2.3 State estimation by resource balance analysis 75
5.2.4 Adaptive MPC algorithm 76
5.3 Application: improved ethanol productivity in microaerobic growth of *E. coli* 76
5.3.1 Network description 77
5.3.2 Simulation cases for adaptive MPC 80
5.3.2.1 Case 1 82
5.3.2.2 Case 2 82
5.3.2.3 Case 3 85
5.4 Results and discussions 85
5.4.1 Case 1 85
5.4.2 Case 2 87
5.4.3 Case 3 92
5.4.4 Computational limitations 94
5.4.5 MPC with and without biomass state estimation 94
5.5 Conclusions 96

6 Conclusion and summary 101

A dFBA-based control problem for maximal productivity 105

B MATLAB implementation of RBA algorithm 107

Bibliography 111

List of Figures

2.1 Metabolic network for central metabolism in *E. coli*, taken from [1] . . 8

2.2 Graphical representation of an example metabolic network and the stoichiometric matrix 9

2.3 Conceptual base of FBA for constraint-based modeling of metabolic network, taken from [2]. 10

2.4 Discretization of variables in the time domain considering two collocation points at each time interval, $K = 2$. The collocation points are determined by zeros of the Legendre polynomials: $t_{i,1} = (i-1)h+(r_1+1)\frac{h}{2}$, $t_{i,2} = (i-1)h + (r_2+1)\frac{h}{2}$ (with $r_1 = -\frac{\sqrt{3}}{3}$, $r_2 = \frac{\sqrt{3}}{3}$). 20

3.1 Distribution of metabolic fluxes in a) wild-type and b) engineered organisms, taken from [3]. 27

3.2 General objectives of metabolic engineering strategies, taken from [4]. . 28

3.3 Schematic representation of the two-stage dynamic control approach. The approach involves a growth stage followed by a production stage, implemented by temporal manipulation of target pathways at genetic (DNA, RNA and enzyme) and/or process levels. 34

3.4 Aerobic growth of *E. coli* on glucose: substrate, products and biomass concentration profiles simulated by the deFBA model together with experimental data obtained by [5]. 40

3.5 Anaerobic growth of *E. coli* on glucose: substrate, products and biomass concentration profiles simulated by the deFBA model together with experimental data obtained by [5]. 41

3.6 Optimal concentration profiles of substrate, products and biomass resulting from the bilevel optimizations for several dynamic regulation strategies: (A) Strategy 1 with repressing the production pathway of ackA enzyme. (B) Strategy 1 with repression of the acetate flux. (Note that the optimal batch time, regulation time and concentration profiles are different for different variants of Strategy 1.) (C) Strategy 2 with a static *ackA* knockout and a switch from aerobic to anaerobic conditions. (D) Concentration profiles with a static *ackA* knockout with no dynamic manipulation. (GLC: glucose, LCT: lactose, ACT: acetate, ETH: ethanol). 44

3.7 Optimal time for repressing the production pathway of ackA enzyme as a function of its degradation constant, k_{deg}. The optimal regulation time increases with the degradation rate constant of the enzyme. 45
3.8 Percentage of total biomass (dry-weight) for key enzymes in the optimal solution for several regulation strategies: (A) Strategy 1 with repressing the production pathway of ackA enzyme. (B) Strategy 1 with repression of acetate flux. (C) Strategy 2 with a static *ackA* knockout and a switch from aerobic to anaerobic conditions. (D) Static *ackA* knockout with no dynamic manipulation. The distribution of enzymes is rearranged in different stages of each growth modes. 50
3.9 Approximation of $g(x)$ by lines $L_1(x)$ and $L_2(x)$. 52

4.1 Open-loop (A) vs. closed-loop (B) control of the bioprocess. 56
4.2 MPC control scheme. 59
4.3 Feeding rate from the closed-loop (solid line) and open-loop (dashed line) control considering no parametric errors. 65
4.4 Biomass, glucose and ethanol concentration profiles resulted from the closed-loop (solid line) and open-loop (dashed line) control considering no parametric errors. 66
4.5 Feeding rate from closed-loop (solid line) and open-loop (dashed line) control for -20% error in the k_{cat} of glucose uptake enzyme. 67
4.6 Biomass, glucose and ethanol concentration profiles resulted from closed-loop (solid line) and open-loop (dashed line) control for -20% error in the k_{cat} of glucose uptake enzyme. 68
4.7 Feeding rate from closed-loop (solid line) and open-loop (dashed line) control for +20% error in glucose feed concentration, GLC_f. 69
4.8 Biomass, glucose and ethanol concentration profiles resulted from closed-loop (solid line) and open-loop (dashed line) control for +20% error in glucose feed concentration, GLC_f. 70

5.1 Adaptive predictive control scheme. 72
5.2 Oxygen-limited growth of *E. coli*: rates of cellular growth, glycerol uptake and acetate formation simulated by the deFBA model together with the experimental data. 81
5.3 Rate constants selection and estimation procedure. 84
5.4 Optimal OUR pattern and concentration profiles of biomass, glycerol and ethanol, resulted from the open-loop optimization. 86
5.5 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. 87

5.6 OUR pattern from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line). 88
5.7 Biomass, glycerol and ethanol concentration profiles resulted from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line). . . 88
5.8 OUR pattern from adaptive MPC/Case 2 (dotted line), adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line). 89
5.9 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). 90
5.10 Enzymes with estimated constants at different OUR levels of the process corresponding to the MPC iteration ($k = 1, ..., N - 1$): *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). 92
5.11 Aerobic and anaerobic glycerol utilization fluxes during the considered microaerobic process (resulted from the open-loop optimization). 93
5.12 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. 95
5.13 Performance of adaptive MPC/Case 1 with state estimation (solid line) and with full state information (dashed line). 97
5.14 Biomass composition in adaptive MPC/Case 1 measured from the plant model (top) and estimated by the RBA (bottom). 98

List of Tables

3.1 FBA-based strain design algorithms. 29
3.2 Metabolic part of the deFBA model: Exchange and metabolic reactions with associated enzymes, and rate constants k_{cat} scaled by the factor f. 38
3.3 Genetic part of the deFBA model: Biomass reactions with values of weights, catalytic (k_{cat}) and degradation (k_{deg}) constants and initial conditions for biomass components. All biomass reactions are catalyzed by ribosome R. 39
3.4 Regulation constraints for the bilevel optimization problem. 42
3.5 Initial nutrient conditions and initial biomass $b^T p(0)$. 42
3.6 Bilevel optimization results for different strategies. 43
3.7 Comparison of biomass growth rates for changes in catalytic constants of metabolic reactions ($k_{cat,new} = k_{cat}/2$) 47

4.1 Initial conditions and model parameters. 64

5.1 Metabolic part of the deFBA model: Metabolic reactions with associated enzymes and rate constants k_{cat}, scaled by the scaling factor f. . . 78
5.2 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. . 79
5.3 Initial nutrient and initial biomass $b^T p(0)$. 80
5.4 Obtained scaling factor values for different oxygen-limited growth modes through the model validation. 82
5.5 Mismatch and adaptation parameters in each simulation case. 83
5.6 Final ethanol concentration resulted from different MPC schemes . . . 89
5.7 Estimated values of the identified parameters (k_{cat}) at each MPC iteration, with their nominal and plant values 91

List of symbols and abbreviations

Variables and parameters used in deFBA-based problems

B	total biomass (g), $B = b^T p$
b	molecular weights of macromolecules (g/mol)
$\mathbb{E}$	set of enzymes
F	feed flow rate (l/hr)
f	scaling factor of catalytic constants
k_{cat}	catalytic constant of enzymes (1/hr)
k_{deg}	degradation constant of macromolecules (1/hr)
K_m, K_I	saturation/inhibition constants
m	interacellular metabolites (mol)
p	macromolecules: enzymes, ribosome, etc. (mol)
p_Q	quota compounds (mol)
$\mathbb{Q}$	set of quota compounds
S	stoichiometric matrix
S_j^i	stoichiometric coefficients of species i for reaction fluxes V_j
t_f	final batch time (hr)
t_{reg}	regulation time (hr)
V	reaction rates (mol/hr)
V_z	exchange reaction rates (mol/hr)
V_m	metabolic reaction rates (mol/hr)
V_p	biomass production reaction rates (mol/hr)
V_d	biomass degradation reactions (mol/hr)
V_{min}	Lower bounds of flux rates
V_{max}	upper bounds of flux rates
V_{reg}	flux rate of the regulated pathway (mol/hr)
v	liguid volume (l)
z	exteracellular species (mol)
μ	overall growth rate (1/hr), $\mu = \dfrac{1}{B}\dfrac{dB}{dt}$
φ_Q	minimal biomass fraction for the quota compound p_Q

Acronyms used in metabolic-genetic networks of *E. coli*

AA	Amino acid
ACT	Acetate
AcCoA	Acetyl coenzyme A

ADP	Adenosine diphosphate
AKG	α-ketoglutarate
ATP	Adenosine triphosphate
ackA	Acetate kinase
adhE	Acetaldehyde dehydrogenase
acn	Aconitate hydratase
atpH	ATP synthase
acs	Acetyl-CoA synthetase
CoA	Coenzyme A
CO2	Carbon dioxide
cyo	Cytochrome bo3 ubiquinol oxidase
dhaK	Dihydroxyacetone kinase
ETH	Ethanol
E4P	Erythrose 4-phosphate
eno	Enolase
FOR	Formate
fba	Fructose-bisphosphate aldolase
fhl	Formate hydrogenlyase
G6P	Glucose-6-phosphate
GLY	Glycerol
GLC	Glucose
gln	Glutamine synthetase
glk	Glucokinase
glpK-D	Glycerol 3-phosphate dehydrogenase
gdhA	Glutamate dehydrogenase
gnd	6-phosphogluconate dehydrogenase
icd	Isocitrate dehydrogenase
LCT	Lactose
lacY	Lactose permease
mdh	Malate dehydrogenase
NADH	Nicotinamide adenine dinucleotide reduced
NAD$^+$	Nicotinamide adenine dinucleotide
nuo	NADH dehydrogenase
OAA	Oxaloacetate
O2	Oxygen
PEP	Phosphoenolpyruvate
PYR	Pyruvate
pdh	Pyruvate dehydrogenase
pfl	Pyruvate formate-lyase
pyk	Pyruvate kinase
pts	Transporter: enzyme II $^{\text{glc}}$

pgl	Phosphogluconolactonase
pgk	Phosphoglycetare kinase
ppc	Phosphoenolpyruvate carboxylase
Q	Quota compound
q	Ubiquinone
qH2	Ubiquinol
R	Ribosome
R5P	Ribose 5-phosphate
RU5P	Ribulose 5-phosphate
rpe	Ribulose-phosphate 3-epimerase
rpi	Ribose-5-phosphate isomerase
S7P	Sedoheptulose 7-phosphate
sucCD	Succinyl-CoA synthetase
T3P	Glyceraldehydes-3-phosphate
tkt	Transketolase
tal	Transaldolase
tpi	Triosphosphate isomerise
X5P	Xylulose 5-phosphate

Abstract

Determination of bioreactor operating policies for optimal performance of a bioprocess is a challenging task due to the highly variable nature of biological systems and our limited process knowledge. To address this challenge, model-based optimization and control approaches can be implemented for conducting *in silico* experiments to derive optimal control strategies for improved performance of bioprocesses.

For reliable performance of model-based optimization and control, it is crucial that the underlying model provides proper levels of detail to represent the real bioprocess and to address the full metabolic versatility. In this work, we consider bioprocess optimization and control by exploiting the capabilities of dynamic metabolic-genetic network models. In particular, we consider improving bioprocess productivity through temporal manipulations of metabolism using dynamic enzyme-cost FBA model (deFBA). The dynamic nature of this model and included details on gene level allow for direct temporal manipulation of gene expression, and through a proper formulation (a bilevel problem), one can identify optimal genetic and process level manipulation strategies according to the target performance criterion (productivity).

Moreover, advanced bioprocess control and optimization requires flexible and robust control strategy which guarantees the performance of the model-based approach in the presence of disturbances and existing uncertainties. To this aim, on-line adaptation schemes are integrated within our modeling approach which are suitable to control highly uncertain biological processes with fast reactions to disturbances. The adaptive approach could allow for online adaptation of the underlying model (deFBA) by estimating uncertain and variable model parameters in different stages of the process. In this direction, the developed deFBA-based approach is implemented inside a model predictive control (MPC) routine, combined with a moving horizon estimation (MHE) algorithm in order to adjust the underlying model online for different metabolic modes.

Considering the case study of ethanol formation in *E. coli* under different growth conditions, it is shown that the proposed approach is a suitable approach to optimize and control time-varying bioprocesses. Desired engineering objectives can be addressed by the proposed approach through temporal manipulations of the metabolism while process uncertainties can be handled efficiently using the adaptive nature of the implemented control scheme.

Banafsheh Jabarivelisdeh Magdeburg, den 5. März 2021

Deutsche Kurzfassung

Das Festlegen von Betriebsrichtlinien für Bioreaktoren für die optimale Durchführung eines Bioprozesses ist aufgrund der sehr unterschiedlichen Natur biologischer Systeme und unserer begrenzten Prozesskenntnisse eine herausfordernde Aufgabe. Um diese Herausforderungen zubewältigen, können modellbasierte Optimierungs- und Regelungsansätze implementiert werden, um in silico-Experimenten optimale Regelstrategien für eine verbesserte Leistung von Bioprozessen abzuleiten. Für eine zuverlässige Durchführung der modellbasierten Optimierung und Regelung ist es entscheidend, dass das zugrunde liegende Modell auf der richtigen Detailebenen liegt, um den tatsächlichen Bioprozess darzustellen und die volle Vielseitigkeit des Stoffwechsels zu berücksichtigen. In dieser Arbeit betrachten wir die Optimierung und Regelung von Bioprozessen, indem wir die Fähigkeiten dynamischer metabolisch-genetischer Netzwerkmodelle nutzen. Insbesondere erwägen wir, die Produktivität von Bioprozessen durch zeitliche Manipulationen des Stoffwechsels, unter Verwendung des dynamischen Enzymkosten-FBA-Modells (deFBA), zu verbessern. Die Dynamik dieses Modells und die darin enthaltenen Details auf Genebene ermöglichen eine direkte zeitliche Manipulation der Genexpression. Durch eine geeignete Formulierung (ein Problem auf zwei Ebenen) können optimale Manipulationsstrategien, auf Gen- und Prozessebene gemäÃŸ dem Zielleistungskriterium (Produktivität), identifiziert werden. Darüber hinaus erfordert eine fortschrittliche Regelung und Optimierung von Bioprozessen eine flexible und robuste Regelungsungsstrategie, die die Leistung des modellbasierten Ansatzes, bei Vorhandensein von Störungen und bestehenden Unsicherheiten, garantiert. Zu diesem Zweck sind Online-Anpassungsschemata in unseren Modellierungsansatz integriert, mit denen sich sehr unsichere biologische Prozesse mit schnellen Reaktionen auf Störungen regeln lassen. Der adaptive Ansatz könnte eine Online-Anpassung des zugrunde liegenden Modells (deFBA) ermöglichen, indem unsichere und variable Modellparameter in verschiedenen Phasen des Prozesses geschätzt werden. In dieser Richtung wird der entwickelte deFBA-basierte Ansatz in einer MPC-Routine (Model Predictive Control) implementiert, die mit einem MHE-Algorithmus (Moving Horizon Estimation) kombiniert wird, um das zugrunde liegende Modell online für verschiedene Stoffwechselmodi anzupassen. Betrachtet man die Fallstudie zur Ethanolbildung in *E. coli* unter verschiedenen Wachstumsbedingungen, wird gezeigt, dass der vorgeschlagene Ansatz ein zur Optimierung und Regelung zeitvariabler Bioprozesse geeignet ist. Gewünschte technische Ziele können durch den vorgeschlagenen Ansatz mithilfe zeitlicher Manipulationen des Metabolismus angegangen werden, während Prozessunsicherheiten unter Verwendung des adaptiven Charakters des implementierten Regelschemas effizient behandelt werden können.

Banafsheh Jabarivelisdeh Magdeburg, den 5. März 2021

1 Introduction

1.1 Model-based optimization and control of bioprocesses

Biological processes are inherently complex which makes their analysis, optimization and control challenging. Cell growth and bioproduct formation are the result of a very large number of cellular reactions and events (e.g. gene expression, translation and biochemical reactions), such that the cellular regulatory mechanisms, cells-medium interactions and the requirements for an optimal process performance are not easy to be analysed and estimated.

However, to keep a bioprocess economically viable, it is essential to operate it as good as possible. Thus, it is valuable to invest in process optimization and control. In industrial practice, improvements are often achieved by trial and error experiments and based on empirical methods in order to characterize the underlying biological system, which makes it time-consuming and expensive. Control of these processes is also usually based on recipes which has insufficient ability to handle possible process uncertainties, resulting in suboptimal production processes [6].

These challenges motivate the implementation of mathematical process models characterized by a priori knowledge of the process as a systematic approach for production process improvements.

1.1.1 Modeling of biological systems

Mathematical models describing the bioprocess as a function of cellular and process parameters enable prediction, optimization and control of the system. It allows for in silico runs of experiments, faster rate of improvements, a reduced number of required experiments and significant reduction of expenses.

The modeling framework for the bioprocess should address both the bioreactor system and the biological cell system as they have complex interactions and cannot be analyzed separately. The nutrients in the media are consumed by the living microorganism and metabolized into several products. Bioreactor models deal with mass transfer aspects and cell models deal with the kinetics of the microorganism. For that, kinetic expressions are used to quantify the rate of cellular processes (rate expressions for the cell growth as well as rate expressions for nutrient uptake and metabolite production) and the influence of other variables (relevant to the bioreactor operating condition) on them.

Bioprocess models in general can be categorized in four classes; structured/unstructured and segregated/non-segregated [7]. Unstructured, non-segregated models consider biomass as a whole (single variable) and cells within the population are assumed to have identical properties (only one morphological form). When subpopulations of cells show different behaviours, a segregated population model is used to give cells of the population different properties [8]. Structuring at the cell level means that the biomass is no longer described by a single variable, and additional variables are used including compartments, enzyme pools and intracellular metabolites [9]. In general, the application purpose of the modeling framework determines the complexity level and structure of the model used.

1.1.2 Application of model-based approaches in biotechnology

Model-based approaches are implemented for bioprocess optimization and control to derive optimal strategies for improved synthesis of desired bioproducts.

For the bioprocess optimization, models allow for determination of optimal process trajectories (commonly referred as the open-loop control or the optimal recipe). This means the model are used to identify optimal profiles of relevant manipulated variables; operating conditions, medium composition, and optimal process timing, respect to a target performance criterion [10]. The performance criterion is mainly defined based on yield or productivity of desired bioproducts which must be stated in the form of an objective function considering the relevant process constraints.

Besides that, process models are used to design control systems. To this aim, bioprocess models are implemented in order to derive control algorithms and to adjust the controller parameters to take care of properties such as stability, robustness and tracking control variables dynamics to their setpoints. As biochemical processes commonly have strongly variable dynamics, the operation condition must also be changed to cope with changing process characteristics. Therefore, model-based bioprocess control often involves online application of optimal control, where control actions are regularly recalculated based on the model and process information (online measurements) [11]. In this direction, bioprocess models are also implemented in developing software sensors. Software sensors are mainly state or parameter estimators which enable online estimation of unknown states and parameters which are important to support online bioprocess monitoring (such as monitoring reaction kinetics), control, and fault detection [12].

For bioprocess optimization and control, non-segregated unstructured models are classically the ones of practical interest for implementations, as they present the process in a very simplified form with a basic representation of cellular metabolism. Unstructured models generally involve lumped descriptions of intracellular metabolism in terms of a specific growth rate and constant yield coefficients, commonly referred to the substrate to biomass yield which is defined as the ratio of the amount of biomass produced to the amount of substrate consumed [13].

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. The regulatory complexity and adaptive responses to transient environmental conditions can not be considered within these models.

In order to improve model prediction capabilities, it is demanded to explore more insights from cellular metabolism. Therefore, precise models exploring intracellular metabolism are required while still low model complexity is desired.

Compared to unstructured models, constraint-based models [14, 15] allow a more detailed description of intracellular metabolism. Constraint-based models include metabolic network which consists of metabolites and metabolic fluxes connecting the metabolites. These models are founded on the stoichiometry of the metabolic network and give an estimate of metabolic fluxes distribution from an optimization principle based on a relevant biological objective. In fact these models are a suitable compromise between unstructured models and the detailed mechanistic descriptions of kinetic models (in which cellular regulations need to be accounted for by inclusion of detailed enzymatic kinetics as well as regulatory terms).

Constraint-based models can be implemented on many levels from characterizing and engineering of the host organism to manipulating and optimization of accompanying bioprocess conditions. They have been shown to be useful in characterizing the metabolic behavior of industrial microbes such as predicting growth and metabolites secretion patterns, determination of substrate utilization range and theoretical product yields, or predicting knockout phenotypes [4, 16]. However, the potential of these models for efficient bioprocess optimization is not sufficiently explored and their application in bioreactor operation and control is not yet well studied, which forms the core questions of this thesis.

1.2 Contribution and outline of the thesis

This thesis aims to contribute in developing advanced model-based approaches for improved performance of bioprocesses, such that desired engineering objectives are addressed by supplying suitable control strategies. It involves applications of dynamic constraint-based models for bioprocess optimization and control through dynamic manipulations of the cellular network for improved production of target biochemicals. In particular we consider increasing bioprocess productivity in which temporal regulation of metabolism is favored for keeping the cellular growth high enough during the process.

Temporal regulation of metabolism is mostly accompanied by genetic regulation of the enzymatic activity during the process, applied in transcriptional (DNA), translational (mRNA), or post-translational (protein) levels. For an efficient dynamic control of the metabolism, we implement a modeling approach based on dynamic metabolic-genetic network models in which the metabolic network is coupled to macromolecule production reactions using a dynamic stoichiometric balancing. Such a model allows to predict temporal variations of involved metabolic enzymes during the process and to address these different levels of manipulations. Through a proper formulation, the modeling approach allows for identification of optimal profiles for the relevant manipulated variables according to the target performance criterion (productivity).

Robustness and flexibility of derived control strategies are another important aspect considered within this work. In fact, for an advanced bioprocess control and optimization it is crucial to have a flexible control strategy which guarantees the performance of the model-based approach in the presence of disturbances and existing uncertainties. The common situation is that biological systems have strong time-varying dynamics. Moreover, small disturbances may lead to changes in the dynamics, and thus uncertainties are inevitable in modeling and simulation of bioprocesses. It means the underlying model may not be able to capture the actual cellular behaviour due to the existing plant-model mismatches. These all result in a poor performance of the control model in addressing desired objectives. To address these challenges, on-line adaptation schemes are integrated within our modeling approach for an adaptive and reliable control of the bioprocess, which allows indirect but fast reactions to disturbances.

As the applications of the developed control approach, within this thesis two variant metabolic-genetic networks of *Escherichia coli* are constructed which differ in available substrates and included pathways, and then implemented to simulate different types of growth modes.

Based on the mentioned objectives, this thesis is organized in six chapters including this introductory part. They are briefly summarized in the following.

In Chapter 2, constraint-based models from stationary approaches to dynamic ones and general principles for construction of these models are reviewed, along with their extension for considering coupled metabolic and genetic networks. In particular, a dynamic metabolic-genetic network model 'dynamic enzyme-cost Flux balance analysis' (deFBA) is presented in details, as the underlying model for bioprocess optimization and control in this thesis.

In Chapter 3, first we provide a review of constraint-based models applications for bioprocess optimization with a focus on bioprocess productivity improvement. Then we formulate a control problem as bilevel optimization framework to identify optimal strategies for dynamic genetic and process level manipulations to increase productivity. The developed problem is based on the deFBA model to capture the network dynamics and enable the analysis of temporal regulation in the metabolic-genetic network. For a simulation, a small-scale metabolic-genetic network of *E. coli* growing on glucose and lactose is first constructed. Then, our computational framework is implemented for an open-loop control of *E. coli* batch growth to maximize ethanol productivity. This example highlights the importance of integrating the genetic level and enzyme production and degradation processes for obtaining optimal dynamic gene and process manipulations. Parts of this chapter are published in [17, 18], in which Jabarivelisdeh and Waldherr contributed to derive the computational formulation, the metabolic-genetic network of the case study and to the analysis of the results. Manuscripts were written by Jabarivelisdeh.

In Chapter 4, we consider a closed-loop control of the bioprocess based on model predictive control which is particularly suited for fed-batch bioreactors. Through integrating measurement feedbacks into the open-loop deFBA-based bilevel problem, optimal fedbatch operating policies for maximal bioprocess productivity are determined while uncertainties in modeling parameters are handled efficiently. Advantages of considering feedback corrections for model-based control of an uncertain biological system are evaluated using the example of fed-batch fermentation of *E. coli* (based on the derived metabolic-genetic network in Chapter 3) for maximal ethanol productivity. Some parts of this chapter are published in [19]. Simulations and writing of the manuscript were performed by Jabarivelisdeh. Waldherr and Findeisen were involved in directing and supervising the work.

In Chapter 5, an adaptive control of the bioprocess is aimed. To this aim and for a reliable performance of the model-based approach, an adaptive model predictive control algorithm based on the deFBA-based bilevel problem is proposed. The adaptive approach considers online adjustment of the underlying model during the process and therefore addresses biological variability while compensating for possible uncertainties. As example application, a microaerobic batch growth of *E. coli* is considered for ethanol production (based on a metabolic-genetic network with glycerol as the substrate, derived in this chapter). The microaerobic growth involves different metabolic modes depending on the degree of oxygen limitations and is suitable to demonstrate the capabilities of the proposed adaptive approach for the bioprocess control. Parts of this chapter are published in [20]. Formulation, simulations and writing of the manuscript were performed by Jabarivelisdeh. Carius, Findeisen and Waldherr provided critical feedbacks and contributed directly in shaping the research and analysis of the results.

Chapter 6 summarizes the outcomes achieved from the implemented constraint-based modeling approach for bioprocess optimization and control, and presents general conclusions to be taken from this thesis.

2 Constraint-based models of metabolic networks

This chapter presents general principles of constraint-based models along with an overview of these models from stationary approaches to dynamic ones. Particularly, constraint-based modeling of metabolic-genetic networks is presented in details as the underlying model for the bioprocess simulations in this thesis.

2.1 Constraint-based modeling and flux balance analysis

Analysis of metabolic networks based on stoichiometric and constraint-based models have become one of the most common and successful modeling approaches in systems biology [4]. The interesting feature of the constraint-based model is that it is able to quantitatively predict the state of a metabolic network without any need for detailed kinetic information and only based on the stoichiometry of reactions. The metabolism of a cell is described by a metabolic network model consisting of metabolites, and metabolic fluxes connecting the metabolites. A typical metabolic network is shown in Figure 2.1, representing the central metabolism of the microorganism *E. coli*.

With n_m number of metabolites and n_r number of reactions, the metabolic network is summarized by the stoichiometric matrix $S \in \mathbb{R}^{n_m \times n_r}$ in which each row corresponds to one metabolite and each column to one reaction. Each matrix element $S_{i,j}$ presents the stoichiometric coefficient of metabolite i in reaction j, with a negative or positive sign. A positive value means that the metabolite is produced in that reaction, whereas a negative value indicates reactants. Then, one can write:

$$\dot{M}(t) = Sv(t). \tag{2.1}$$

where $M \in \mathbb{R}^{n_m}$ is the vector of metabolite concentrations and $v \in \mathbb{R}^{n_r}$ is the reaction flux vector of the network.

Figure 2.2 shows a simple metabolic network which comprises three internal metabolites (B, C, and D), three external species (A as the substrate and E, F as products) and six reactions connecting them ($v4$, $v5$, $v6$ as exchange reactions and $v1$, $v2$, $v3$ as internal reactions), with corresponding stoichiometric matrix for internal metabolites. The core assumption in constraint-based modeling of metabolic networks is that the intracellular metabolites are in steady state. The logic behind this assumption is that dynamics inside the cell are much faster than outside the cell, as metabolites of the central carbon metabolism have a relative high turnover [21], and the production rate of macromolecules is slow compared to metabolic reactions. Based on that, the intra-

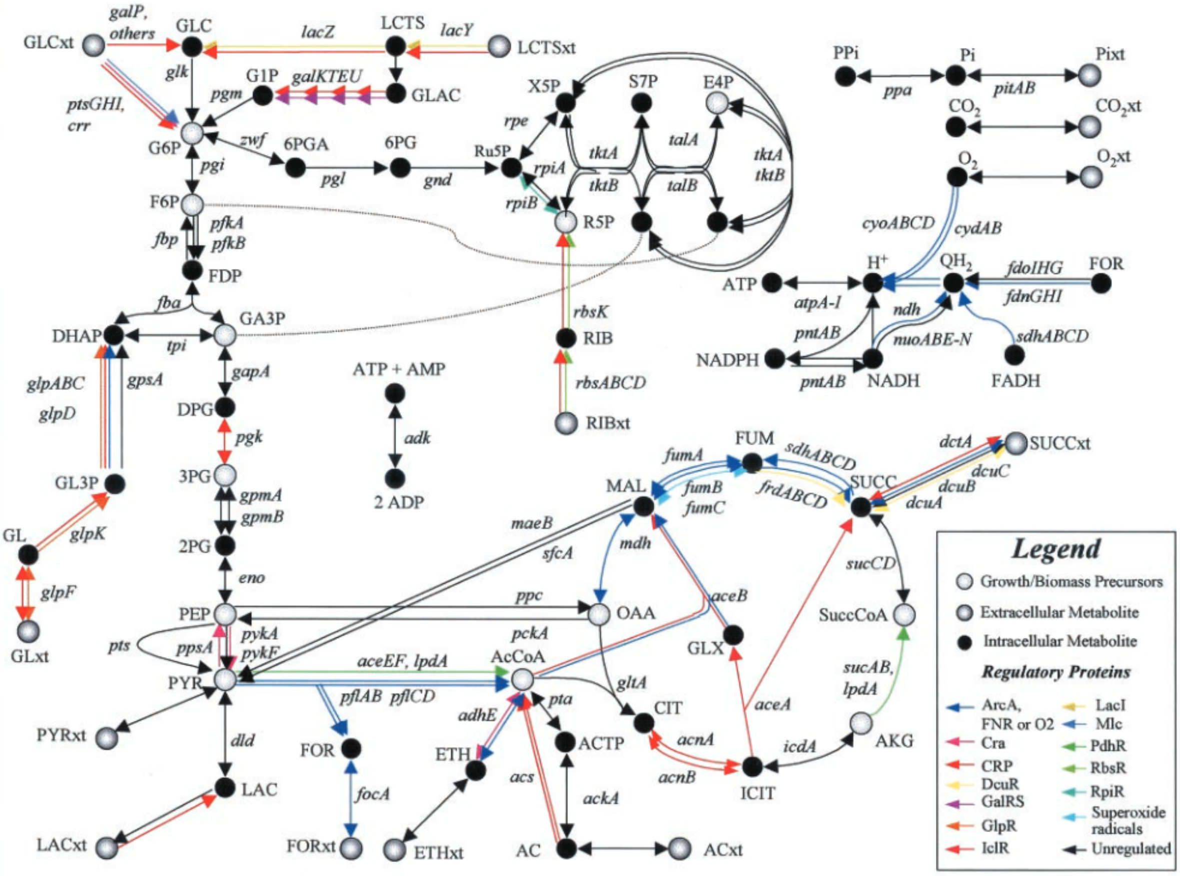

GLCxt
galP, others
GLC
lacZ
LCTS
lacY
LCTSxt
PPi
Pi
Pixt
ppa
pitAB
glk
G1P
galKTEU
pgm
GLAC
X5P
S7P
E4P
CO2
CO2xt
ptsGHI, crr
G6P
zwf
6PGA
6PG
rpe
Ru5P
tktA
talA
tktA
tktB
O2
O2xt
pgi
pgl
gnd
rpiA
tktB
talB
rpiB
R5P
F6P
cyoABCD
cydAB
fbp
pfkA
pfkB
FDP
ATP
H+
QH2
fdoIHG
FOR
fdnGHI
fba
atpA-I
ndh
GA3P
rbsK
DHAP
tpi
pntAB
nuoABE-N
sdhABCD
RIB
NADPH
glpABC
gpsA
gapA
ATP + AMP
pntAB
NADH
FADH
glpD
DPG
rbsABCD
pgk
adk
RIBxt
FUM
dctA
SUCCxt
3PG
fumA
sdhABCD
dcuC
GL3P
gpmA
gpmB
2 ADP
MAL
fumB
fumC
frdABCD
SUCC
dcuB
dcuA
maeB
GL
glpK
2PG
sfcA
mdh
sucCD
Legend
glpF
eno
aceB
SuccCoA
ppc
PEP
OAA
GLX
GLxt
pts
ppsA
pykA
pykF
pckA
AcCoA
sucAB, lpdA
aceEF, lpdA
PYR
pflAB pflCD
gltA
CIT
aceA
adhE
pta
PYRxt
dld
FOR
ETH
ACTP
acnA
icdA
AKG
acs
acnB
ICIT
LAC
focA
ackA
LACxt
FORxt
ETHxt
AC
ACxt
○ Growth/Biomass Precursors
○ Extracellular Metabolite
● Intracellular Metabolite
Regulatory Proteins
ArcA, FNR or O2
Cra
CRP
DcuR
GalRS
GlpR
IclR
LacI
Mlc
PdhR
RbsR
RpiR
Superoxide radicals
Unregulated

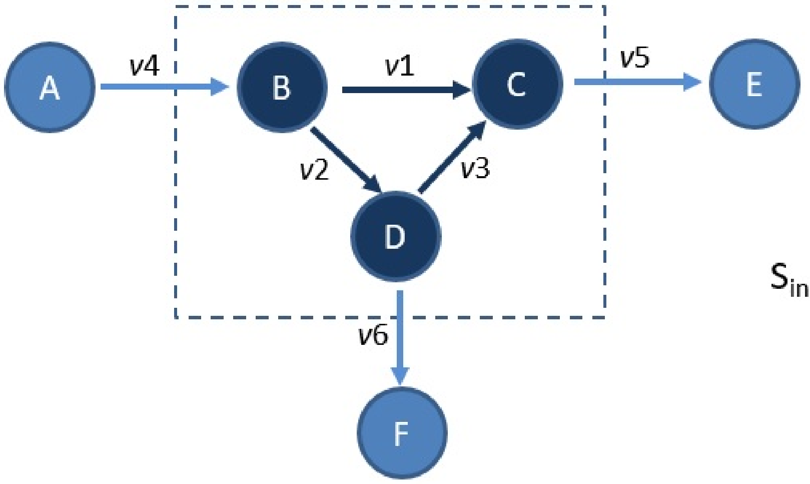
Figure 2.1: Metabolic network for central metabolism in *E. coli*, taken from [1]

cellular metabolites are considered to be in (pseudo) steady-state which means rates of production and consumption are equal for each metabolite. Based on this assumption, the mass balance for the intracellular metabolites is obtained as the general equation:

$$S_{int}.v = 0 \tag{2.2}$$

with S_{int} which includes the stoichiometric coefficients of intracellular metabolites in the reaction flux vector v (containing internal and exchange reactions). This is a homogeneous systems of linear equations. The degree of freedom of the system (i.e., the number of reactions minus the rank of S which shows the number of linearly independent rows) determines the number of unknown fluxes, for which measurements are required to calculate a unique flux solution. Since the number of reactions is typically much larger than the number of metabolites and not sufficient fluxes can be reached in practice, the system is usually underdetermined.

The classical approach to overcome this problem is to use constraint-based modelling in form of flux balance analysis (FBA) which solves this underdetermined system by defining an optimization problem [2, 14]. FBA considers typically the following set of constraints:



$$S_{int} = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{cccccc|c} v1 & v2 & v3 & v4 & v5 & v6 & \\ -1 & -1 & 0 & 1 & 0 & 0 & B \\ 1 & 0 & 1 & 0 & -1 & 0 & C \\ 0 & 1 & -1 & 0 & 0 & -1 & D \end{array}$$

Figure 2.2: Graphical representation of an example metabolic network and the stoichiometric matrix

- steady state mass balance constraints $S_{int}v = 0$,
- capacities constraints ($v_{min} \leq v \leq v_{max}$) which define upper or lower boundaries for some fluxes (which are often known for exchange reactions),
- irreversibility of reactions ($0 \leq v \leq v_{max}$),

which results in constraining the flux space to a convex cone of feasible fluxes, and then a flux solution is found by optimizing a linear objective function. The objective function is typically a linear combination of network fluxes which are specified by a n_r-dimensional vector w as $w^T v = w_1 v_1 + w_2 v_2 + ... + w_{n_r} v_{n_r}$. The most frequently used objective is maximizing the flux for biomass production (growth), which has been shown to be a biologically realistic cellular objective at least for some microorganisms (e.g. yeasts) under certain environmental conditions [22, 23]. To define biomass maximization as the objective, all elements of w are 0, except the element for the biomass production reaction which is set equal to 1. Overall, maximizing the linear objective together with other constraints results in a standard linear optimization problem - flux balance analysis:

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & w^T v \\ \text{subject to} \quad & Sv = 0, \\ & v_{min} \leq v \leq v_{max}. \end{aligned} \tag{2.3}$$

The process of constraining the solution space in FBA is illustrated in Figure 2.3, in which the most right figure shows the optimal solution which lies at a vertex of the solution space resulted from the linear programming problem.

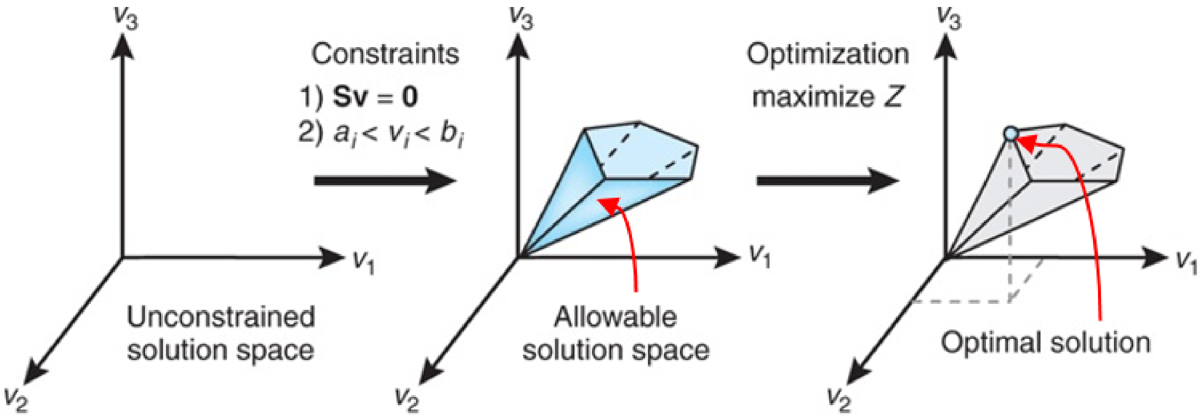


Figure 2.3: Conceptual base of FBA for constraint-based modeling of metabolic network, taken from [2].

Doing so, FBA predicts the particular steady state flux distribution by optimizing a "celluar" objective function. As FBA allows to study various functional properties of metabolic networks based on the network structure, its basic principles has been referred in modified versions in order to improve model predictions of cellular behaviour and to tackle specific qualities.

Extensions to FBA were made in several aspects including the integration of regulatory events [24, 25], considering thermodynamic constraints [26, 27], integration of gene expression data [28, 29], and capturing effects of genetic modifications [30] (for a comprehensive review, see [31]). One important driver for extending classical FBA is the desire to tackle transient changes of metabolic fluxes and species. FBA gives the optimal flux distribution assuming the cells are in steady state. To describe cellular processes in changing environments, several efforts have been made to simulate dynamic profiles of metabolite concentrations and metaboic fluxes [5, 15, 24]. We present next the dynamic flux balance analysis (dFBA) approach [15] as an established approach for dynamic optimization of metabolic networks.

2.2 Dynamic flux balance analysis

Dynamic flux balance analysis (dFBA) as a dynamic extension to FBA has been proposed to study transients in metabolism by computing optimal metabolic fluxes over a time range of interest.

In general, the model dynamics are given through a set of ordinary differential equations. These equations define the mass balances for species (total biomass, substrates and products),

$$\dot{z}(t) = S_e v(t) X(t),$$

$$
\begin{aligned}
\dot{X}(t) &= \mu X(t), \\
\mu &= \sum w_i v_i,
\end{aligned} \tag{2.4}
$$

where X is a scalar representing biomass concentration and z is the vector for extracellular species (substrates and products) concentrations with S_e as the corresponding stoichiometric matrix. v is the reaction flux vector per gram of the biomass (commonly as mmol.(g biomass)$^{-1}$.hr^{-1}), which its multiplication to the biomass concentration X (g biomass.l^{-1}) gives the absolute flux (mmol.l^{-1}.hr^{-1}). μ is the growth rate defined as a weighted sum of the reactions for growth precursors production. w_i are the amounts of the growth precursors required per gram of biomass. Along with the dynamic equations, several additional constraints can be considered including non-negative metabolite values and bounds on transport fluxes [15].

The dFBA appears in two different formulations: the dynamic optimization approach (DOA) and the static optimization approach (SOA).

The dynamic approach, DOA, considers a growth-related objective and allows computing optimal fluxes over the whole time range of interest. Considering maximizing the biomass amount at the end of the process as the objective function, the dynamics of cellular metabolism are described by the following dynamic optimization problem:

$$
\begin{aligned}
\underset{v(\cdot)}{\text{maximize}} \quad & X(t_f) \\
\text{subject to} \quad & \dot{z}(t) = S_e v(t) X(t), \\
& \dot{X}(t) = \mu(t) X(t), \qquad \mu = \sum w_i v_i \\
& z(t_0) = z_0, \quad X(t_0) = X_0, \\
& z(t) \geq 0, \quad X(t) \geq 0, \\
& S_{int} v(t) = 0, \\
& v_{min}(t) \leq v(t) \leq v_{max}(t).
\end{aligned} \tag{2.5}
$$

Through this problem, metabolic fluxes v as control variables are computed in order to maximize the biomass production $X(t_f)$ at the end of the process (final time t_f) based on the defined constraints including the steady state mass balance for intracellular metabolites, the initial condition for extracellular species and biomass concentration z_0, and X_0, positivity of species as well as the flux bounds (v_{min} and v_{max}). As upper bounds to metabolic fluxes, additional constraints can also be added to cover transport limitations. These might be the Michaelis-Menten kinetics for substrate uptake or inhibitory terms that reflect growth rate suppression by the presence of particular species.

Such a dynamic optimization problem can be solved in multiple ways, e.g. by discretizing the dynamic equations in the time domain using collocation methods [32, 33] resulting in a nonlinear programming problem. The details of the solution procedure using collocation approaches will be presented later in Section 2.3.2.

In contrast to DOA in which the optimization problem is solved once, the static approach SOA does not resolve the optimization problem over the complete timescale of interests. Instead it involves predicting dynamic changes in biomass and extracellular metabolites iteratively. For that, the time period is divided into N intervals. An FBA problem (as in problem (2.3)) is solved at the beginning of each time interval t with corresponding bounds on exchange fluxes, to obtain the flux distribution v and the growth rate μ (as the biomass production flux $\sum w_i v_i$) at that time instant. Then based on the obtained fluxes v and μ, and using an approximation of the differential equations (assuming the fluxes are constant over the intervals), the extracellular metabolites and biomass concentrations are updated for the following time interval $t + \Delta T$:

$$\begin{aligned} \Delta T &= \frac{t_f - t_0}{N}, \\ z(t + \Delta T) &= z(t) + S_e v \Delta T, \\ X(t + \Delta T) &= X(t) + \mu X(t) \Delta T. \end{aligned} \tag{2.6}$$

where ΔT is the length of the time interval. So for the SOA, dFBA solves N linear programming problems, one for each time step, in order to obtain the time profiles of extracellular metabolites and biomass concentrations. It can be seen as an approximate solution of the original problem.

These two approaches of dFBA have been evaluated and compared in [15] considering the analysis of the diauxic growth in *Escherichia coli*. In general, besides computing the time courses of biomass and external species, the dFBA approaches allow to formulate substrate uptake kinetics. Through the substrate uptake expressions, it is possible to represent maximum uptake rates respect to transport limitations, such that some known regulatory processes such as diauxic consumption of substrates and growth inhibition by extracellular metabolites can be addressed. Based on these features, dFBA has been implemented in several studies as a tool for phenotype predictions or process/strain optimizations in batch and fed-batch fermentation applications and co-culture simulations [34–39].

As dFBA models involve dynamic systems and their formulation is more complex than a simple FBA model, some efforts have been also made for more efficient implementations of these models and to make them more accessible as a modeling tool, such as the constraint-based reconstruction and analysis (COBRA) toolbox [40] and the DFBAlab simulator, both provided in the MATLAB environment [41].

2.3 Metabolic-genetic network models

So far, the biomass is captured only as one component (X) in FBA and dFBA models which is one point of critiques to these approaches. In fact, with a coarse definition of biomass, its different allocations to different cellular components and the associated metabolic processes cannot be represented in detail. Examples include the expression of different enzymes for catalyzing different metabolic reactions, or the allocation towards the ribosomal machinery for faster protein production vs. the allocation towards enzymes for a faster metabolic turnover.

Inclusion of biomass composition and enzyme costs in metabolic optimization can potentially improve the quality of predictions which has been studied in several works. As an extension to FBA in this direction for describing the metabolic network together with a variable biomass resource allocation, the resource balance analysis (RBA) approach has been proposed [42]. This approach considers the conversion of metabolites into enzymes and other cellular macromolecules and constrains metabolic fluxes based on the bioenergetic costs for producing required enzymes for metabolic pathways and their catalytic capacity. RBA yields a nonlinear optimization problem which can predict the cell composition of bacteria and flux distributions in a specific (constant) environment by maximizing the cellular growth. The optimization problem for RBA will be presented later within this chapter (Section 2.3.3). A conceptually equivalent approach has been proposed independently under the term of ME model (metabolism and macromolecular expression) as a means to explore the relationship between genotype and phenotype using biochemical representations of transcription and translation processes [43, 44].

Above mentioned approaches are however limited to stationary exponential growth, and extensions were needed for description of dynamic changes of metabolic-genetic fluxes. In this direction, dynamic enzyme-cost flux balance analysis (deFBA) has been proposed in order to understand how resources are distributed in response to the dynamically changing environment. Such a dynamic metabolic-genetic approach can be used to study the dynamic growth of microorganisms not only to get temporal profiles of species and fluxes but also to understand the transitions in cellular components distribution in response to changing operating conditions of a bioreactor.

In general, there are two subsystems which are important for metabolic-genetic network models:

- The metabolic network which represents catabolism, anabolism and other cellular functions. These reactions are catalyzed by enzymes.
- The translation apparatus which involves producing required enzymes (for metabolism) and other substantial cellular macromolecules.

Here, we present details for constructing these models with a focus on the deFBA formalism which is described in the following section.

2.3.1 Dynamic enzyme-cost flux balance analysis

The dynamic enzyme-cost flux balance analysis is a constraint-based modeling approach which includes metabolic reaction network coupled with gene expression. It can be considered as the extension of the established dFBA by considering a detailed description of biomass as well as the cost and catalytic capacity of enzymes [45, 46]. The deFBA in general is formulated as a dynamic optimization problem for maximizing a growth-related objective which 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.

The model consists of n_M biochemical species divided into different groups:

- extracellular species present in the environment (nutrients and products) with the molar amount vector $z \in \mathbb{R}^{n_z}_{\geq 0}$,
- intracellular metabolites with the molar amount vector $m \in \mathbb{R}^{n_m}_{\geq 0}$,
- macromolecules which are catalytic enzymes (gene products) and cellular building blocks with the molar amount vector $p \in \mathbb{R}^{n_p}_{\geq 0}$.

with $n_M = n_z + n_m + n_p$, $[z] = [m] = [p] =$ mmol.

To define the total biomass B, we use the molecular weights b_i of all macromolecules p, as $b = (b_1, b_2, ..., b_{n_p})$, $[b_i] =$ g/mol. The total biomass (dry weight in g) is obtained by:

$$B = b^T p \tag{2.7}$$

As with the species, we classify R network reactions into different groups:

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

where $R = R_z + R_m + R_p + R_d$. Corresponding to the unit used for molecular species, all reaction flux vectors use a unit of molar amount per time.

For a general network of this type, the differential equations describing the dynamics of the species are given using the stoichiometric matrix $S \in \mathbb{R}^{n_M \times R}$ as

$$\begin{aligned} \dot{z}(t) &= S^z_z V_z(t), \\ \dot{p}(t) &= S^p_p V_p(t) - S^p_d V_d(t), \end{aligned} \tag{2.8}$$

with

$$V_d = k_{deg} p. \tag{2.9}$$

Thereby, S^i_j ($j \in z, p, d$ and $i \in z, p$) are stoichiometric matrices with the stoichiometric coefficients of species i for reaction fluxes V_j. The biomass degradation flux V_d is modelled to be proportional to the amount of each biomass component p with corresponding degradation constants k_{deg}.

As explained earlier, the intracellular metabolites m can be considered to be in quasi-steady state which yields:

$$\dot{m} = S^m_z V_z(t) + S^m_m V_m(t) + S^m_p V_p(t) + S^m_d V_d(t) = 0, \tag{2.10}$$

where S^m_z, S^m_p and S^m_d are the stoichiometric matrices with the stoichiometric coefficients for the metabolic species in m in the exchange reactions, biomass production, and biomass degradation reactions flux vectors V_z, V_p, and V_d, respectively, and S^m_m is the stoichiometric matrix for the metabolic network. To simplify notation, reaction fluxes for exchange reactions, biomass production reactions and metabolic reactions are collected in the overall reaction flux vector

$$V = \begin{pmatrix} V_z \\ V_m \\ V_p \end{pmatrix} \in \mathbb{R}^{R_z + R_m + R_p},$$

which is considered as the free time-dependant variable to be used in the dynamic optimization. Note that degradation reactions are defined with the kinetic law (2.9) and not subject to optimization. The corresponding stoichiometric matrices are

$$S = \begin{pmatrix} S^z_z & 0 & 0 \\ 0 & 0 & S^p_p \end{pmatrix},$$

$$S_m = \begin{pmatrix} S^m_z & S^m_m & S^m_p \end{pmatrix}.$$

The state variables are also collected as

$$x = \begin{pmatrix} z \\ p \end{pmatrix} \in \mathbb{R}^{n_z + n_p}.$$

As there is no explicit flux for the biomass growth, and instead there are pathways for producing individual biomass components, the overall growth rate can be determined as

$$\mu(t) = \frac{1}{B(t)} \frac{dB(t)}{dt}, \tag{2.11}$$

with the total biomass $B(t)$.

The deFBA model also includes several biophysical constraints:

- Enzyme capacity constraints, in which the reaction fluxes are limited by the maximum enzymatic capacity. This maximum capacity is defined by the catalytic constant (k_{cat}) and the amount of associated enzymes. Individual enzymes may catalyze multiple reactions. Hence, the capacity constraint for a single enzyme p_i can be written as
$$\sum_{j\in cat(i)} |\frac{V_j}{k_{cat,\pm j}}| \leq p_i, \qquad i \in \mathbb{E} \tag{2.12}$$
where $cat(i)$ indexes the set of reactions catalysed by the enzyme p_i, $i \in \mathbb{E}$ in which $\mathbb{E}$ is the set of enzymes. $k_{cat,+j}$ ($k_{cat,-j}$) denotes the constant of forward (backward) reaction j with the flux of V_j. As an example, consider the case where enzyme p_1 catalyses a reversible reaction V_1 and an irreversible reaction V_2. This results in two constraints, each for one possible combination of reaction flux directions:
$$\begin{aligned}\frac{V_1}{k_{cat,+1}} + \frac{V_2}{k_{cat,+2}} &\leq p_1,\\ \frac{V_1}{k_{cat,-1}} + \frac{V_2}{k_{cat,+2}} &\leq p_1,\end{aligned} \tag{2.13}$$
- Biomass composition constraints: in addition to enzymes, the macromolecules p also include non-catalytic proteins and other molecules which do not contribute directly to metabolism and growth, but they are required by the cell to keep it working like DNA, RNA and cell walls. These species are denoted as "quota compounds" [46], and the constraints are expressed by enforcing a minimal fraction φ_Q ($0 \leq \varphi_Q \leq 1$) of the total dry weight $b^T p$ of the cell to be made of a certain quota compound p_Q
$$\varphi_Q b^T p \leq p_Q, \qquad Q \in \mathbb{Q} \tag{2.14}$$
where $\mathbb{Q}$ is the set of quota compounds.
- Enzyme-independent flux bounds which for example can be used to express the irreversibility of reactions
$$V_{min} \leq V \leq V_{max}. \tag{2.15}$$
- Positivity of molecular species
$$\begin{aligned} z &\geq 0,\\ p &\geq 0. \end{aligned} \tag{2.16}$$

Besides the above-mentioned constraints, an objective function is still required to define the complete optimization problem of the deFBA.

Biomass maximization is a common objective used in constraint-based models and classical flux balance analysis. Similarly, the deFBA assumes that the cell evolves in a way to maximize its growth in the form of maximizing total biomass. In the original work [45], this objective is incorporated within the model in two alternative ways:

first, as biomass maximization at the end of the optimization horizon (at time t_f),

$$J_1 = B(t_f) = b^T p(t_f), \tag{2.17}$$

and second, as maximization of the biomass integrated over the considered time span $[t_0,\, t_f]$,

$$J_2 = \int_{t_0}^{t_f} B(t)dt = \int_{t_0}^{t_f} b^T p(t)dt. \tag{2.18}$$

The maximization of the terminal biomass (objective J_1) has been used in dFBA (as shown in Section 2.2), while the biomass integral (objective J_2) has been used as an evolutionary fitness measure in analysis of microbial metabolism [47].

These two objective functions can be compared by computing the variability of the optimal solutions of the deFBA and the biophysical meaning of the solution through considering an example of a minimal nutrient uptake network, taken from [48]. As discussed in [45], for the first objective function J_1, the results show non-unique solutions and a quite high flux variability as obtained by a Flux Variability Analysis [49]. Compared to that, for the second objective J_2 the uniqueness of the solutions is observed with zero flux variability for the considered example. The interpretation of the integral objective is almost identical to the growth rate maximization over the complete time scale, and therefore it allows for solutions which do not optimize the growth rate at any given time. Implementing the objective J_2 could also deliver results in agreement with the typical growth kinetics of bacterial cultures [47], as the growth at earlier time points would give a better objective functional value than the growth at a later time.

Therefore, we use the objective $J = J_2$ for the deFBA model implemented in this thesis.

In summary, the deFBA model is described by the following dynamic optimization problem:

$$
\begin{aligned}
\underset{V(\cdot)}{\text{maximize}} \quad & \int_{t_0}^{t_f} b^T p(t) dt \\
\text{subject to} \quad & \dot{x}(t) = \begin{pmatrix} \dot{z}(t) \\ \dot{p}(t) \end{pmatrix} = SV(t) - S_d^x V_d(t), \\
& S_m V(t) + S_d^m V_d(t) = 0, \\
& \sum_{j \in cat(i)} |V_j(t)/k_{cat,j}| \leq p_i(t), \qquad i \in \mathbb{E} \\
& V_{min}(t) \leq V(t) \leq V_{max}(t), \\
& \varphi_Q b^T p(t) \leq p_Q(t), \qquad Q \in \mathbb{Q}, \\
& x(t_0) = x_0 = (z_0, p_0), \\
& x(t) \geq 0.
\end{aligned}
\tag{2.19}
$$

where metabolic fluxes V as the control variables are computed in order to maximize the biomass production based on the given constraints and initial condition x_0.

Therefore, the key difference with the dFBA approach is that deFBA considers a detailed presentation of biomass and allocates available resources over the different enzymatic pathways in an optimal way in order to achieve maximum biomass growth.

In contrast to detailed kinetic models, where regulatory interactions from metabolites on enzyme synthesis are commonly used [50], the deFBA model as shown 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 [45]. However, the downside of this approach is the requirement for knowing the details of producing gene products and catalytic constants of enzymes in the model, as well as the increased numerical complexity due to the addition of biomass components and associated constraints.

The optimization problem (2.19) addresses the cellular growth in the batch mode. To be implemented for growth prediction in fed-batch or continuous modes, the model dynamics should be revised accordingly and other relevant constraints might be added. For example, additional constraints on the substrate uptake rate may be applied to address growth inhibitory effects which is a common situation for fed-batch growth modes.

2.3.2 Numerical solution

Solving problem (2.19) is important. So, we implement collocation methods to approximate this problem by discretization of dynamic variables in the time domain [32, 33]. For the collocation, the time interval $[t_0,\, t_f]$ is divided into N equally sized intervals, each of length

$$h = \frac{t_f - t_0}{N}. \tag{2.20}$$

Each time interval contains K collocation points. Overall, the collocation points are given by the sequence

$$t_{1,1},\ t_{1,2},\ \dots,\ t_{1,K},\ t_{2,1},\ \dots,\ t_{N,K}.$$

Within each interval, the q-th collocation point is at position r_q (relative to the interval $[-1, 1]$), where r_q is determined by the collocation scheme and order. The collocation points are thus computed as

$$t_{i,q} = (i-1)h + (r_q + 1)\frac{h}{2}. \tag{2.21}$$

The flux variable $V(t)$ and the derivative of the state variable $\dot{x}(t)$ are then discretized with a Lagrange interpolation scheme:

$$\begin{aligned} V(t) &= \sum_{q=1}^{K} V_{i,q} L_q\left(\frac{2t - 2t_{i-1} - h}{h}\right), \qquad t_{i-1} \le t \le t_i, \\ \dot{x}(t) &= \sum_{q=1}^{K} \dot{x}_{i,q} L_q\left(\frac{2t - 2t_{i-1} - h}{h}\right), \qquad t_{i-1} \le t \le t_i, \end{aligned} \tag{2.22}$$

where L_q, $q = 1, ..., K$ are the Lagrange polynomials

$$L_q(r) = \prod_{1 \le i \le K, i \ne q}^{K} \frac{r - r_i}{r_q - r_i}, \tag{2.23}$$

and $t_i = ih$, $i = 0, 1, ..., N$ are the boundaries of the time intervals used for the discretization. The state variable $x(t)$ is discretized at the boundaries of the N time intervals and its value within an interval is approximated by integrating over the time derivatives.

Here, we consider collocation points determined by zeros of the Legendre polynomials. The discretization of variables in the time domain is schematically shown in Figure 2.4, considering two collocation points at each time interval, $K = 2$.

In this way, the continuous optimization problem in (2.19) is approximated by a finite-dimensional problem in which the optimization is carried out over the vector $W \in \mathbb{R}^{NK(R_z+R_m+R_p+n_z+n_p)+N(n_z+n_p)}$ defined as

$$W = (V_{1,1}, V_{1,2}, ..., V_{N,K}, \dot{x}_{1,1}, \dot{x}_{1,2}, ..., \dot{x}_{N,K}, x_1, x_2, ..., x_N),$$

.

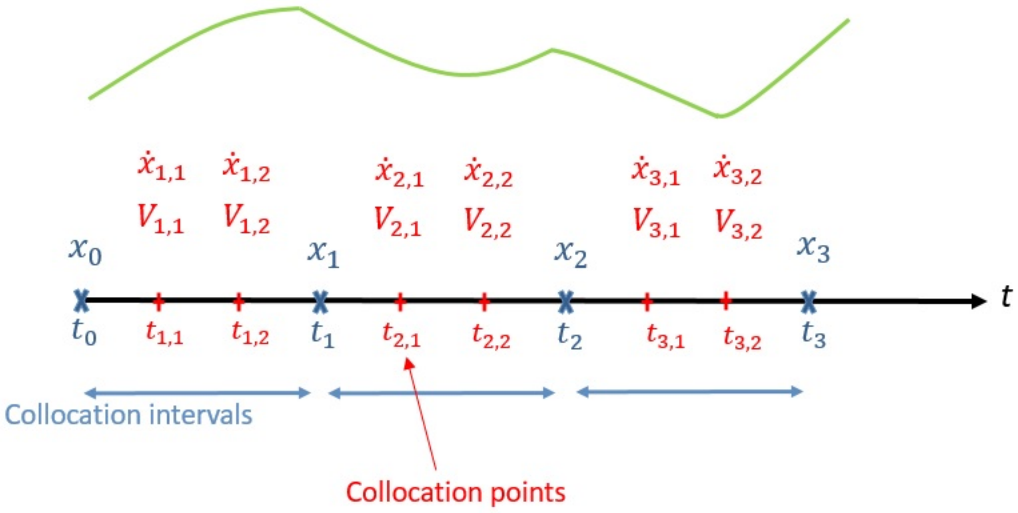


Figure 2.4: Discretization of variables in the time domain considering two collocation points at each time interval, $K = 2$. The collocation points are determined by zeros of the Legendre polynomials: $t_{i,1} = (i-1)h + (r_1+1)\frac{h}{2}$, $t_{i,2} = (i-1)h + (r_2+1)\frac{h}{2}$ (with $r_1 = -\frac{\sqrt{3}}{3}$, $r_2 = \frac{\sqrt{3}}{3}$).

The finite dimensional optimization is then derived as:

$$\begin{aligned} \underset{W}{\text{maximize}} \quad & c^T W + d \\ \text{subject to} \quad & \mathcal{M}_e W = \mathcal{V}_e, \\ & \mathcal{M}_i W \leq \mathcal{V}_i. \end{aligned} \tag{2.24}$$

where the vector c and the number d stem from the discretization of the objective functional. The equality constraint matrix $\mathcal{M}_e$ and the vector $\mathcal{V}_e$ stem from the collocation of differential equations, the initial condition, and the quasi steady state constraint. The inequality constraint matrix $\mathcal{M}_i$ and the vector $\mathcal{V}_i$ stem from the collocation of inequality constraints in problem (2.19). The resulting LP can then be solved numerically using standard commercial solvers such as CPLEX or Gurobi. More details for discretization of variables can be found in [45].

2.3.3 Resource balance analysis

Based on the derived formulation for deFBA approach, it is straightforward to formulate the resource balance analysis (RBA) problem which addresses cell simulation in steady state and constant environmental conditions. The main assumption of RBA is the stationary growth of the cell. Based on this assumption, the cellular concentrations of macromolecules need to remain constant, which means the amount of macromolecules p_i increases at the same rate as the biomass. This assumption is defined as the below constraint in RBA:

$$S_p^p V_p = \mu p. \tag{2.25}$$

All other constraints from deFBA including enzyme capacity and biomass composition are implemented in RBA as well, but in their time-independent form. Moreover, a constraint for the biomass amount in the system ($b^T p = B_0$) is added to get a well defined optimization problem. The RBA problem is then defined as a nonlinear optimization problem (due to the presence of the term μp in mass balance equations of cellular components) which maximizes the growth rate:

$$\begin{aligned} \underset{\mu,V,p}{\text{maximize}} \quad & \mu \\ \text{subject to} \quad & S_p^p V_p - \mu p = 0, \\ & b^T p = B_0, \\ & S_m V = 0, \\ & \sum_{j \in cat(i)} |V_j / k_{cat,j}| \leq p_i, \qquad i \in \mathbb{E} \\ & V_{min} \leq V \leq V_{max}, \\ & \varphi_Q b^T p \leq p_Q, \qquad Q \in \mathbb{Q}. \end{aligned} \tag{2.26}$$

Within the RBA problem formulation, it is assumed that the cell is at a quasi steady state and for each given value of the biomass B_0, the metabolic fluxes and cellular components are allocated to maximize the cellular growth rate μ. In other words, through the problem (2.26), optimal metabolic fluxes and biomass components are computed to have a maximal steady state cellular growth based on the given constraints and total biomass amount B_0.

2.3.4 General procedure for deriving a metabolic-genetic networks model

In this section, we provide the main requirements for generating metabolic-genetic network models. The overall protocol for deriving such a network can be summarized as the following steps. A comprehensive procedure for generating and exchanging metabolic-genetic network models can be however found in [46].

- The most important part of deriving a metabolic-genetic network model is a (genome-scale) metabolic reconstruction of the target organism. The metabolic network is constructed based on an annotated genome sequence of the organism and includes describing the reactions of central carbon metabolism, the reactions of the amino acid synthesis pathways and the pathways for the biosynthesis of precursors of structural cell components. So far, around 2600 draft reconstructions for a multitude of organisms are freely available via online databases like

BioModels [51] and BiGG Models [52], from which one can obtain the metabolic reconstruction for the organism of interest.

It is important to note that it is not in general possible to simulate a full genome-scale deFBA model due to the size of the resulting linear program. While simulation of steady-state resource allocation in genome-scale networks can be feasible by RBA, dynamic approaches such as deFBA are still constrained to smaller sizes. Till now, networks with up to 500 metabolic reactions have been successfully simulated with deFBA as shown in [53]. However, if the starting metabolic network is too large, it is required to reduce the size of the network by relevant network reduction tools such as the minimal network finder in [54] or NetworkReducer [55], which allows to preserve desired network functionalities.

- Next, the genetic network is constructed by describing pathways for catalytic proteins, ribosomes and quota compounds.

  Catalytic protein (enzyme) production reactions are constructed based on the gene-reaction mapping (which describes which genes are involved in the catalyzing of each reaction) and their corresponding amino acid sequences. This information can be obtained from online databases such as Genbank [56] and UniProt [57]. It should be highlighted that via UniProt it is not only possible to get data on a the gene sequence, but also to access to the Enzyme Commission (E.C.) number and the subunit stoichiometry for enzyme complexes.

  The protein production reaction then can be specified based on the computed amino acid count of each protein and other additional reactants like energy cofactors (e.g. in the form of ATP) required in producing each protein. Note that for enzyme complexes, one needs to calculate the amino acids cost of each subunit and then sum those up respect to the stoichiometry of each subunit. The wrong enzyme compositions obtained from this step can change the contribution of the involved enzymes for the growth simulation.

  To set up ribosomes synthesis, we consider it as a large enzyme complex with a single production reaction which describes the full translation and assembly process. To construct the production reaction of ribosomes, the required ribosomal proteins, ribosomal RNA and energy cofactors need to be taken into account. This information on compositions of ribosomes can be found in the Kyoto Encyclopedia of Genes and Genomes (KEGG) resource [58], or the Ribosomal Protein Gene Database [59].

  Then, we derive the production reaction for quota species such as DNA, membranes, cell wall and RNA. As explained earlier these components play no catalytic roles in the RBA/deFBA models, but we must enforce their production via the biomass composition constraint to address that part of the resource cost required for proper cellular functionality. To reduce the number of quota compounds in the model, we lump these together and build a single reaction that produces the merged quota compounds.

To construct the production pathway for quota compounds, the appropriate biomass fraction for these compounds should be first set up. For that we need information on the percentage of biomass subcomponents, which can be obtained from literature. For example, from the database BioNumbers [60], the composition of dry weight of the bacteria *E. coli* is reported as 55% protein (catalytic and non-catalytic), 20% RNA, 10% Lipids, and 15% others.

After obtaining data on fractions of biomass components included within the model, the biomass reaction of the metabolic network reconstruction can be useful to proceed. Based on this overall biomass production reaction and considering the portion of metabolites already used in enzymes and ribosomes synthesis reactions, an average pathway for producing the quota compounds of the cell is constructed. In cases which the ratio of biomass components is not directly available in the literature, a detailed biomass production pathway itself can be a good place to look, as it represents the typical biomass composition of the target organism. In fact the stoichiometric coefficients of the involved metabolites of the biomass reaction can describe the average composition of the cell. This is explained in detail in [46].

The enzymes, ribosomes, and the structural component and their synthesis reactions are then added to the stoichiometric matrix.

- Then, the reaction catalytic constants (turnover rates) should be extracted from databases and integrated into the model. Catalytic constants are very important parameters in metabolic-genetic network models involved in enzyme capacity constraints, which highly impact the contribution of different metabolic pathways during the growth.

  The two main databases for obtaining enzymatic constants are BRENDA [61] and SABIO-RK [62]. These databases often enable the user to filter catalytic constant values for different organisms, mutant strain, pH value, etc. As a simple rule of thumb, one should look for wild-type data obtained at a physiological pH for the organism of interest, to take care of the model generality [46].

  Beside extracting from databases, it is also possible to estimate the catalytic constant of enzymes based on their properties; for example in [63], the catalytic constant of enzymes has been considered to be proportional to the enzyme solvent accessible surface area (SASA) and estimated from the enzyme molecular weight.

- Once the model is constructed, the important remaining step is tuning the model to match experimental growth rates or measured reaction fluxes obtained in laboratory or retrieved from literature.

This step may involve adjusting some model parameters if there is a noticeable mismatch between the growth rate μ predicted by the model and the one from literature. For example if the simulated growth is too small, one should check the k_{cat} values again as this is most likely caused by very limiting values of some core pathways. On the other hand, a very large simulated growth value probably means the lack of maintenance or quota requirements, and one should check the construction of any biomass composition constraints or consider maintenance fluxes.

3 Bioprocess optimization and control by dynamic metabolic-genetic network models

In this chapter, we consider the applications of constraint-based models for optimization and control of the bioprocess. First, we review the FBA-based computational algorithms implemented for optimization of biotechnological processes. In particular we focus on their applications for improving bioprocess productivity which are mainly conducted by dynamic implementation of metabolic engineering strategies along with dynamic control of fermentation condition.

As the next step, we implement dynamic metabolic-genetic network models for improving bioprocess productivity through dynamic regulation of metabolism. To this aim, within this chapter we focus on formulating the control problem for our target objective (productivity), defining proper gene-level and process-level strategies and implementing the proposed approach on a well-established case study using an experimentally validated metabolic-genetic model.

3.1 Application of constraint-based modeling in bioprocess optimization and control

Constraint-based models which represent detailed descriptions of cellular metabolism compared to unstructured models, is an effective simulation approach to form the basis for model-driven bioprocess optimization and control. Desired quantities to be usually improved through the bioprocess optimization are product yield and productivity, and FBA-based approaches can be implemented to address these metrics by finding optimal production strategies [64].

FBA-based models allow to explore potential production capabilities of metabolic networks. In this direction, a main and important application of constraint-based models is in the area of metabolic engineering, for strain design. To this aim, several algorithms have been proposed to identify targets of genetic interventions in order to improve the production process (reviewed in [65]). One example is Optknock algorithm [66], which is a FBA-based approach proposed for identifying candidates of gene deletions to increase the product yield. For genetic-level applications, these models can be also useful in determining the optimal time for applying genetic interventions (as implemented in [37]). Besides that, in the process-level, FBA-based models can be implemented to determine the optimal process and bioreactor condition, such as the substrate feeding profile or the aeration level for an optimal cell growth. These implementations of FBA-based models in determining genetic and process level manipulation strategies for improved production processes are addressed in details within the following sections.

3.2 FBA-based approaches for metabolic engineering

Increasing demands for the production of industrially relevant bioproducts, necessitate the design of microbial strains that can produce valuable biochemicals which are optimized not only for economical but also sustainability aspects. Metabolic engineering as an enabling technology for this process involves developing new experimental and theoretical methodologies for targeted improvements of microbial hosts. It aims to redirect metabolic fluxes in order to amplify the production of the compound that wild type organisms produce in low concentration which is not cost effective for industrial-scale production [4, 67]. Figure 3.1 shows the basic principle for metabolic engineering; the host organism is engineered such that the metabolic fluxes are redistributed to achieve increased production of the desired product. In this direction, FBA-based approaches have been used to search for suitable interventions that redirect reaction fluxes towards the product, resulted in several successfully engineered production strains [4].

The general objectives of metabolic engineering strategies are depicted in Figure 3.2. The Figure presents the phenotypic phase plane with respect to two key quantities: the biomass yield and the product yield. The yellow area shows all combinations of these yields in the feasible steady state flux vectors of the network. This allows to identify two extreme points: one point for maximum biomass yield (corresponding to the wild type strain, as the basic assumption of FBA) and another point for maximum product yield, where all the substrate is completely converted to the product and no biomass is produced. The practically desired phenotype is shown by the blue area in which the flux vectors is distributed in a way that the product yield is high, while still allowing a reasonable biomass yield for a higher production process rate. Therefore, interventions which cut away undesired regions by knockouts and flux redistributions are followed.

A simple approach to do so is deleting pathways of undesired products formation and possibly overexpression of pathways leading to target product formation. An important point is that pathways for building biomass precursors should not be removed. However, it is still not ensured that the mutant uses pathways to product formation as its objectives and it may adjust its metabolism such that only pathways for biomass synthesis are active while no product is produced. An strategy to overcome this problem is the coupling of cellular growth and product formation. This means one follows gene deletions which make the secretion of the product an essential pathway required for the growth, and thus the target product is produced whenever the organism synthesize biomass. In this way, the product formation is coupled to biomass synthesis.

Coupling the product synthesis with the growth (biological and economical objectives) has some advantages. One is that a minimum yield is guaranteed whenever the cells grow. Moreover, due to adaptive evolution the cells move towards higher growth rates over time [68]. As the formation of the desired product is in the route of the cell growth, the product yield is also increased [4].

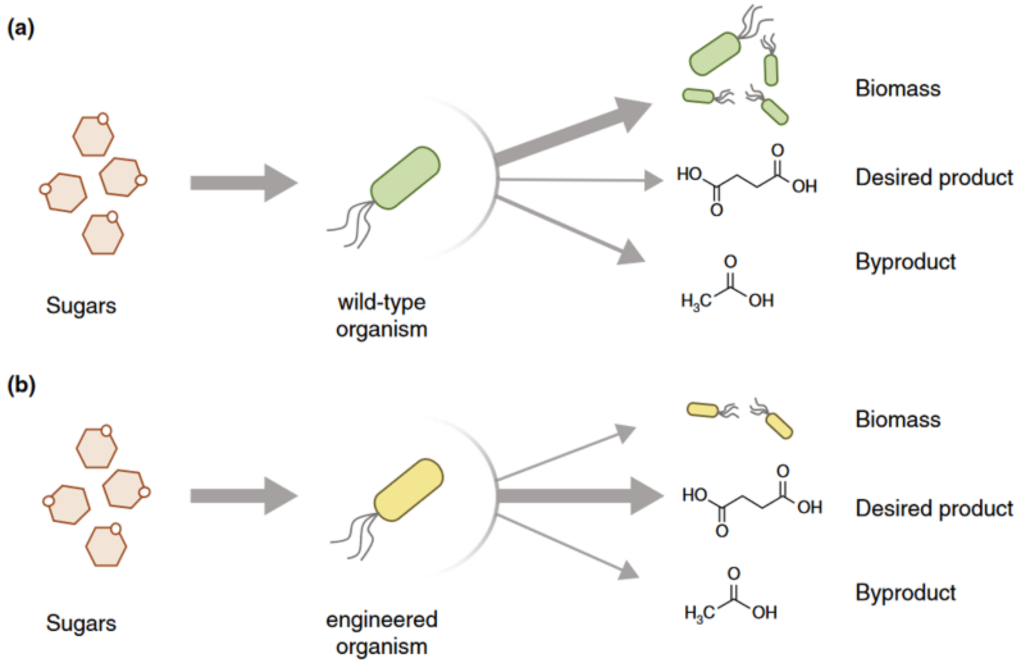


Figure 3.1: Distribution of metabolic fluxes in a) wild-type and b) engineered organisms, taken from [3].

Many of the FBA-based strain design algorithms seek to derive strategies which lead to coupled biomass and product formation. For this purpose, most of these strain design algorithms implement bilevel optimization formulations that allow to address two competing objective functions at the same time (e.g. the biological objective vs. the bioengineering objective).

Available approaches aim in general to identify target reaction (gene) knockouts, non-native additional pathways, a combination of knockouts, down-regulations and over-expressions leading to the overproduction of a desired chemical. These approaches mainly address improving product yield as the desired quality for the engineered strain.

In this direction, Optknock was the first FBA-based optimization method proposed to identify suitable reaction knockouts in metabolic networks for achieving the highest production yield [66], leading to a bilevel optimization problem. The inner optimization problem maximizes the biological objective (growth) as the FBA model, while the outer optimization maximizes product synthesis by identification of optimal gene/reaction knockouts under the inner problem constraint. The obligatory growth-coupled product synthesis is thus a direct aim of this approach by implementing a nested optimization. To model gene/reaction deletions, Optknock incorporates binary control

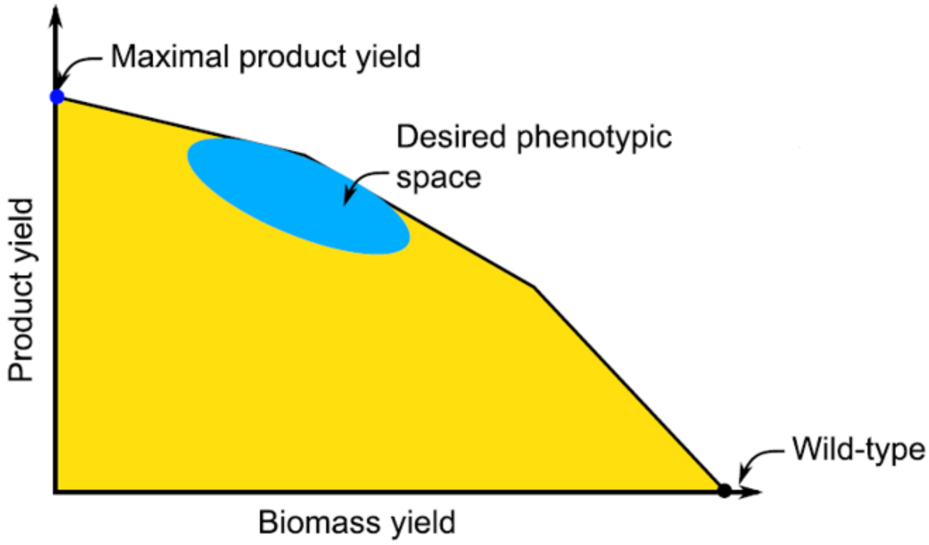


Figure 3.2: General objectives of metabolic engineering strategies, taken from [4].

variables. Optknock is expressed as the following bilevel mixed integer optimization problem:

$$\begin{aligned} \underset{y_j}{\text{maximize}} \quad & v_{product} \\ \text{s.t.} \quad & \sum_j (1-y_j) \leq k, \\ & \underset{v_j}{\text{maximize}} \quad v_{biomass} \\ & \text{s.t.} \quad Sv_j = 0, \\ & \qquad v_{min}.y_j \leq v_j \leq v_{max}.y_j, \end{aligned} \tag{3.1}$$

where y_j are binary variables selected from the outer optimization (for maximizing the production flux) and they are defined to have a value of one if reaction j is active and a value of zero if it is inactive. v_j represents the flux of reaction j and k is the number of allowable knockouts. Depending on y_j values determined from the outer optimization (0 or 1), the constraint

$$v_{min}.y_j \leq v_j \leq v_{max}.y_j,$$

defines the state of reactions through the inner FBA problem; A reaction flux v_j is set to zero if y_j is equal to zero, otherwise v_j is free to have any value between a lower v_{min} and an upper v_{max} bound. To solve the problem (3.1), the bilevel optimization is transformed into a single level one in the original work [66] based on linear programming duality theory, resulting an overall mixed integer linear problem (MILP).

Predictions of Optknock have been verified by successful application to real problems for improved product yields [69, 70]. This approach was further modified to algorithms

Table 3.1: FBA-based strain design algorithms.

Name	Type of optimization problem	Type of intervention	Reference
Optknock	Bilevel, MILP	Knockouts	[66]
RobustKnock	Multi-level, MILP	Knockouts	[71]
OptGene	Evolutionary	Knockouts	[72]
OptStrain	Bilevel, MILP	Addition of non-native reactions	[73]
OptReg	Bilevel, MILP	Knockouts, upregulations and downregulations	[74]
OptORF	Bilevel, MILP	Knockouts and overexpressions (of metabolic and regulatory genes)	[75]
OptForce	Bilevel, MILP	Knockouts, upregulations and downregulations	[76]
GDLS	Heuristic	Knockouts, upregulations and downregulations	[77]

which allow for other types of network manipulations e.g. inclusion of regulatory constraints or heterologous reactions. A brief summary of these methods is presented in Table 3.1. For a comprehensive review see [65]. Compared to Optknock, among these modified approaches, in RobustKnock the objective of the outer optimization problem (engineering objective) was adjusted to maximize the minimal production of the target product. This reformulation in fact makes the product secretion really obligatorily coupled to biomass formation, which is an effective improvement to the original Optknock formulation. It should be kept in mind that coupling in these approaches is considered based on the assumption of growth optimal behavior of the cell (with growth maximization as the cellular objective). Therefore, in cases that the organism does act optimally respect to its growth, one should expect to achieve lower yield and suboptimal production of the target product consequently.

Solving MILPs from Optknock and similar methods is more complicated than the LP of the classical FBA and is even more challenging in case of multiple knockouts. To speed up the calculation, other algorithms have been implemented; OptGene [72] for example applies an evolutionary algorithm and GDLS [77] uses a heuristic search algorithm.

3.3 Dynamic control for improved bioprocess productivity

Yield and productivity are the main variables which must be optimized for an economical and viable bioprocess. Classically, yield has high priority and is particularly considered as the main objective in developing strain design algorithms.

As shown in the previous section, developed computational design algorithms mainly aim to maximize product yield using static approaches (gene additions or/and deletions and up- or down-regulation of gene expression). Through these genetic manipulations, the product formation flux is maximized which in fact can drain metabolites needed for biomass synthesis. This negatively affects the cellular growth rate. Even though the growth rate in a genetically modified strain can be also improved by adaptive evolution, this strategy can not be applicable in every case, as the growth rate may not increase to an acceptable level or the initial rates can be very low for successful adaptive evolution [3]. Therefore, most of the strategies which are applied for yield improvement, result in strains with low volumetric productivity (the target metabolite produced per unit time and volume) due to the impaired growth rate.

In industrial applications, optimizing the productivity of an engineered strain by reducing the processing time is usually the actual performance parameter in reducing operation costs, and increasing the product yield does not necessarily result in sufficient production rates [78]. This suggests that the trade-off between growth and production is crucial in the design of a bioprocess and determination of optimal operating conditions for maximizing these conflicting metrics is essential to ensure commercial viability.

Recently, algorithms have been developed based on FBA approaches which consider not only product yield but also productivity. For example, in [79], hypothetical strains are generated by establishing many operating points with a defined growth rate and product flux. These strains are then evaluated for productivity, yield and titer to identify the optimal strain maximizing an objective considering all these three metrics. However, similar to other available approaches, this algorithm assumes static manipulations of the metabolism in which the enzymes will not be dynamically controlled.

When the productivity is the bottleneck of a bioprocess, dynamic control of the metabolism proved to be more advantageous, as they permit to regain the bacterial growth which is severely affected by static strategies such as gene deletions [80]. Here the dynamic control means that manipulations of genetic and metabolic networks are not implemented statically at the beginning of the process and instead the metabolism can be modified temporally during the process. However, determination of the optimal manipulation strategy is an essential task for a proper balance between the growth and production. To achieve this aim, dynamic constraint-based models can be a suitable modeling tool which allows for temporal control of cellular networks and to capture the trade-off between yield and productivity. They are implemented to derive dynamic control strategies, mainly in two directions: dynamic metabolic engineering and dynamic optimization of fermentation conditions, explained in following sections. Both of these strategies are implemented for improving bioprocess productivity in considered case studies within this thesis.

3.3.1 Dynamic metabolic engineering

As mentioned, most of strain design algorithms placed emphasis on yield improvement using static gene modifications without the possibility for a temporal control of enzyme expression (static metabolic engineering), resulting in reduced bacterial growth.

One efficient approach to overcome the reduced growth imposed by genetic modifications is the use of two-stage dynamic optimization, which includes a growth stage followed by a production stage [3]. In the growth stage, the flux distribution is as in the wild-type. Through this stage, a high growth rate is achieved by the wild-type level expression of genes and the biomass is generated quickly. Following the growth stage, the production stage is considered in which the genetic manipulations is implemented in order to maximize the flux through the product pathway. Therefore, instead of applying the gene alterations at the start of the process (static metabolic engineering), the target gene expression is regulated at an optimal time during the process to balance between cell growth and product formation (dynamic metabolic engineering), resulting in improved productivity.

For dynamic metabolic engineering purposes, there are diverse mechanisms to implement temporal regulations of gene expression practically. For instance, the gene expression can be regulated in response to an environmental signal, such as temperature. This has been used in [81], in which glycolysis in *Escherichia coli* has been temporally controlled by regulating the expression of a key glycolytic enzyme through changes in temperature. A genetic control module based on quorum sensing is another practical mechanism to this purpose. Quorum sensing acts to control cell-density dependent processes in bacteria and can be a basis to design a circuit to dynamically regulate the expression of target genes at desired times and cell densities. In [82], a quorum sensing-based circuit has been implemented for dynamic control of endogenous essential genes in glycolysis to redirect fluxes into a heterologous engineered pathway in *Escherichia coli*. Besides, temporal genetic regulations can be also conducted based on the Clustered regularly interspersed short palindromic repeats (CRISPR) system of a bacteria cell. Based on that, the technology (CRISPR) interference (CRISPRi) has been introduced as an RNA-guided platform for control of gene expression [83, 84]. This system allows to hinder transcription of the targeted DNA and to cause gene repression, by utilizing a single protein and a designed guide RNA.

For an efficient implementation of dynamic metabolic engineering strategies for improved process productivity, it is however essential to find the optimal timing for implementing the genetic regulation and the temporal profile of the corresponding flux. In this direction, constraint-based models have been implemented to assess the applicability of the two-stage control and to determine the optimal profiles of target metabolic pathways. Such efforts started with some works by Gadkar et al. [37, 85]. In [37] using constraint-based metabolic models of *Escherichia coli*, it was demonstrated that a dynamic manipulation of gene expression achieves a higher productivity compared to applying static gene knockouts. To maximize productivity for a target metabolite, the authors implemented a bilevel optimization framework that computes the optimal temporal profile of a manipulated metabolic flux. This bilevel optimization included FBA with a growth objective to determine the metabolic fluxes, as the inner problem. The optimal profile of the manipulated flux and the optimal regulation strategy were determined using the static optimization approach (SOA) of the established dynamic flux balance analysis (dFBA) approach [15] through an outer nonlinear problem. The strategy proposed in [37] was further integrated to an algorithm based on genetic toggle switch for its practical implementation [86, 87]. Anesiadis and co-workers in [86] proposed an integrated computational model to manipulate the key metabolic fluxes. The system is based on quorum sensing and its feasibility for dynamic control of gene expression was demonstrated *in vivo*.

3.3.2 Dynamic optimization of fermentation condition

Besides engineering of the cellular metabolism, temporal control of the fermentation process itself (bioreactor operating condition) represents another aspect of bioprocess optimization for improved productivity. To increase the process productivity, genetic manipulations can be coupled with dynamic regulations at the process level. In process level control, the cellular growth (impaired by gene deletions, species-specific inhibitory effects, etc) can be improved by modifying environmental conditions, including dissolved oxygen concentration, substrate feeding rate, inducer concentration and pH. In fact, implementing suitable operating policies for the process allows for higher rates of product formation and improved productivity.

To this aim, dynamic constraint-based models can be used to design key operating conditions such as optimal feeding policy, scheduling of batches, aeration pattern, and media composition [16]. One common implementation of process level dynamic control is considering two growth stages by performing a switch between aerobic and anaerobic conditions, where the switching time is chosen in an optimal way. In several studies [38, 88, 89], the productivity has been considered as a function of this operating strategy and the dFBA model has been used to determine the optimal switching time by adjusting the aeration level.

3.4 Dynamic control for improved bioprocess productivity based on deFBA model

Despite of the advantages associated with dynamic manipulation of cellular metabolism for improved process productivity, so far it has been addressed in few studies and the modeling tools for this purpose are still limited [3]. In general, temporal genetic manipulation can be implemented at transcriptional (DNA), translational (mRNA), and post-translational (enzyme) levels. Each of these levels requires unique engineering tools to address the dynamic control at different time-scales. This makes dynamic implementation of metabolic engineering strategies more challenging rather than the static strategies [90].

For an efficient dynamic control of the metabolism, advanced model-based approaches are required to address the different levels of manipulations and to identify proper strategies which are optimal regarding the target engineering objective. To regulate the enzymatic activity during the process, the underlying model should account for the genetic level of cellular networks and the temporal variations of involved metabolic enzymes. To achieve this aim, during this project we consider the dynamic enzyme-cost FBA (deFBA) model (as explained in Chapter 2) as the underlying model for temporal bioprocess optimization. The dynamic nature of the deFBA model and its ability to define biomass composition including metabolic enzymes allow to explicitly include the temporal control of enzyme expression in the optimization problem.

As shown earlier, constraint-based models are founded on the cells maximizing a reasonable biological objective such as biomass, which is often opposed to the production of a target metabolite from a bioprocess engineering point of view. A bilevel optimization is a suitable approach to describe the bioprocess optimization and to address conflicting objectives, as implemented in several strain design algorithms. Through a bilevel formulation, in the outer optimization problem, the engineering objective is maximized, while the inner problem represents the cellular model where the cellular objective is maximized.

3.4.1 Dynamic optimization problem formulation based on deFBA model

Here, we formulate a bilevel optimization to obtain the optimal strategies to temporally manipulate cellular metabolism for improved bioprocess productivity. To take care of process productivity, we consider dynamic regulation of metabolism in a two-phase process which includes a growth stage followed by a production stage [3]. It is implemented by a dynamic manipulation of one or several specific reaction fluxes. The target flux for temporal manipulation may be selected by a static bilevel optimization method such as Optknock [66] (explained in Section 3.2). In the first phase, the manipulated flux is still active to favour the cell growth. Then at an optimal time the manipulated flux is deactivated (by genetic or process level repressions) to switch the metabolism to the second phase. In that phase, the metabolic fluxes mainly contribute to product formation, while growth may be reduced. This is shown in Figure 3.3, in which in the first stage growth enzymes are mainly active while in the second stage contribution of production enzymes is promoted.

In the bilevel optimization problem, the productivity is used as the objective. It is defined as the target product concentration $x_t(t_f)$ at the end of batch divided by the

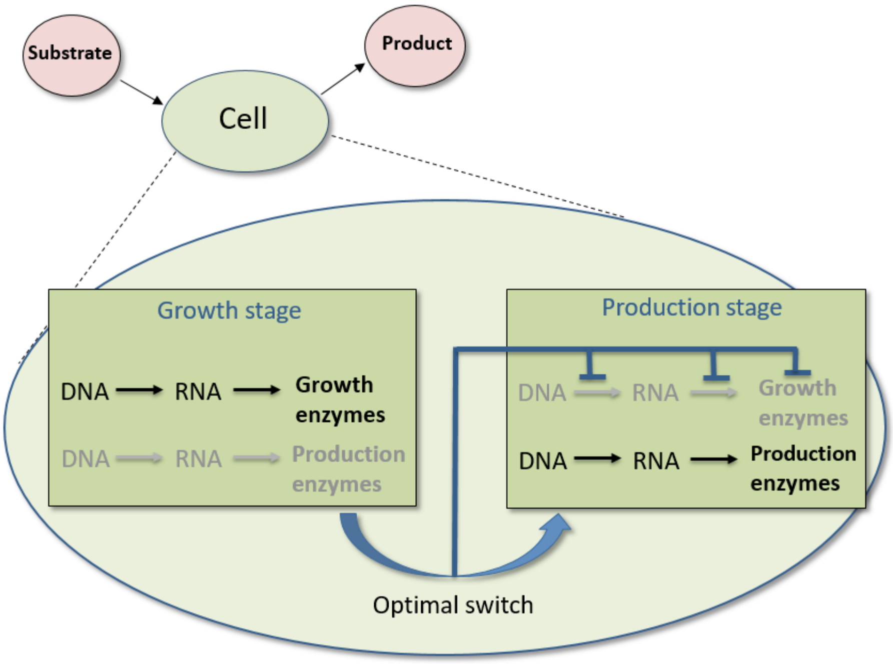

Substrate
Product
Cell
Growth stage
DNA
RNA
Growth enzymes
DNA
RNA
Production enzymes
Production stage
DNA
RNA
Growth enzymes
DNA
RNA
Production enzymes
Optimal switch

Figure 3.3: Schematic representation of the two-stage dynamic control approach. The approach involves a growth stage followed by a production stage, implemented by temporal manipulation of target pathways at genetic (DNA, RNA and enzyme) and/or process levels.

batch time [86]:

$$pr = \frac{x_t(t_f)}{t_f}. \tag{3.2}$$

By applying a bilevel optimization approach, we are searching for the optimal profile of the manipulated fluxes that produces the target product with maximal productivity. In the inner problem, deFBA model as presented in (2.19) is used to determine the optimal metabolic flux distribution for cellular objective optimization.

The formulated dynamic bilevel problem using the deFBA model is as follows:

$$
\begin{aligned}
\underset{t_f, t_{reg}}{\text{maximize}} \quad & \frac{x_t(t_f)}{t_f} \\
\text{s.t.} \quad & t_0 \leq t_{reg} \leq t_f, \quad t_{min} \leq t_f \leq t_{max}, \\
& \underset{V(\cdot)}{\text{maximize}} \int\limits_{t_0}^{t_f} b^T p(t) dt \\
& \text{s.t.} \quad \dot{x}(t) = \begin{pmatrix} \dot{z}(t) \\ \dot{p}(t) \end{pmatrix} = SV(t) - S_d^x V_d(t), \\
& \qquad S_m V(t) + S_d^m V_d(t) = 0, \\
& \sum_{j \in cat(i)} |V_j(t)/k_{cat,j}| \leq p_i(t), \qquad i \in \mathbb{E} \\
& \qquad V_{min}(t) \leq V(t) \leq V_{max}(t), \\
& \qquad \varphi_Q b^T p(t) \leq p_Q(t), \qquad Q \in \mathbb{Q} \\
& \qquad x(t_0) = x_0 = (z_0, p_0), \\
& \qquad x(t) \geq 0, \\
& \qquad V_{reg}(t) = 0 \quad \text{for} \quad t \geq t_{reg}.
\end{aligned}
\tag{3.3}
$$

In the outer optimization, the optimal batch time t_f and the optimal time of regulation for the manipulated flux t_{reg}, are determined to achieve maximum productivity of the target metabolite x_t. With the metabolic fluxes defined as $V = (V_{reg}, V_{unreg})$, the regulation time is the time in which the target flux for manipulation V_{reg}, is down-regulated. Here, we define the down-regulation by repressing the manipulated flux to zero. The outer optimization is subjected to an inner optimization (the deFBA model) which addresses maximizing biomass. Since the deFBA model is used as a constraint in (3.3), the resulting reaction fluxes $V(t)$ will at the same time be optimal for the biomass objective J (equation 2.18) in the inner problem. However, this optimal solution will typically be different based on the choice of regulation time t_{reg} and batch time t_f imposed by the outer problem.

Using this defined bilevel problem with the deFBA model, one can directly represent a manipulation of gene expression pathways through adapting the biomass production fluxes V_p. However, in previous approaches with the FBA and dFBA models, it is not possible to directly represent a manipulation of the enzyme production pathway or gene expression in the optimization problem, as within these models the biomass production is defined as a single reaction and there are no separate pathways for the biomass components including enzymes.

3.4.2 Solution procedure

We solve the bilevel optimization problem (3.3) in MATLAB. The outer optimization is performed using *fmincon*. The inner deFBA model is solved as described in Section

2.3.2. The resulting linear optimization, e.g. (2.24), for the inner problem is solved using *linprog* in each iteration of the outer problem's solution. Note that problem (3.3) is typically not convex, so appropriate care has to be taken to ensure that a global optimum is found. For that, several initial conditions have been evaluated for the outer optimization to avoid obtaining a local minimum for the problem's variables.

3.5 Application: improved ethanol productivity in *E. coli*

To validate and try the approach, we selected an anaerobic fermentation using *E. coli* for ethanol production, a well-established case from the metabolic engineering perspective. To this aim, first we develop a small-scale network model of *E. coli* including substantial metabolic and genetic pathways which is capable to predict *E. coli* growth. Then this network model is used within the developed bilevel approach for the control of target pathways, resulting in maximal ethanol productivity.

3.5.1 Network description

We consider a small-scale metabolic-genetic network model of *E. coli* including the core processes relating carbon uptake and growth, which is developed from an earlier *in silico* model of the central *E. coli* metabolism [1]. Deriving this metabolic-genetic network is guided by [46] and based on the general procedure outlined in Section 2.3.4.

This network includes uptake reactions for carbon sources 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 (including non-catalytic proteins, lipids and other cellular components). All reactions with corresponding enzymes and their catalytic constants are given in Tables 3.2 and 3.3 and for simplicity, gene IDs are used for the naming of enzymes. The main simplifications and reduction steps for obtaining this network are as below:

- As carbon sources, glucose and lactose are considered. Glucose uptake in the model is implemented through the PTS system.
- All reactions from pentose phosphate pathways are described by reaction 5 [91].
- All reactions from the TCA cycle are lumped into reactions 11 and 12.
- Only acetate and ethanol are considered as fermentation products, while production pathways for other co-products like succinate and lactate are deleted.
- Series reactions are lumped into one reaction which uses the smallest catalytic constant of the involved enzymes. For the lumped reactions, we consider *Lumped enzymes* which are obtained based on the composition of individual enzymes, p_i

and their catalytic constants $k_{cat,i}$, $i \in E_l$, in which E_l is the set of involved enzymes in a lumped reaction. In scaling to the enzyme with the smallest catalytic constant, we define

$$k_{min} = \min_{i \in E_l} \quad k_{cat,i}, \tag{3.4}$$

and the composition of the lumped pathway enzyme p_l is then obtained as

$$p_l = \sum_{i \in E_l} \frac{k_{min}}{k_{cat,i}} p_i. \tag{3.5}$$

- The production reaction for the structural component is based on the composition of the dry weight of an *E. coli* cell [60]. From the overall biomass production reaction of *E. coli*, the components of the ribosome and enzymes that are modelled individually in the deFBA model are subtracted, and the remainder (which are not catalysing reactions) are combined in a single quota compound Q. From the available data for *E. coli* [60], it is estimated that this quota makes up at least 55 % of the total biomass. This is formulated in the biomass composition constraint as
$$0.55 b^T p \leq Q. \tag{3.6}$$
- During the degradation reactions, we assume the macromolecules break down to their individual components with no energy usage or synthesis. For example, each catalytic enzyme decomposes only to its amino acid content with no ATP production and consumption (as in Table 3.3). The degradation constants for enzymes are approximated from the bulk protein half-lives in *E. coli* [60]. The ribosome and the structural component are considered to have a much longer half-lives (and be degraded much slower) than metabolic enzymes.

The biomass components collected in vector p and their weight vector b are listed in Table 3.3. The weight vector are calculated based on the stoichiometric coefficients of the respective subcomponents and their molar mass. Considering the initial biomass concentration of $B_0 = b^T p(0) = 0.005$ gl^{-1}, the initial biomass composition $p(0)$ is computed by the resource balance analysis (RBA) [42], using (2.26), to yield the maximum growth rate for the considered metabolic-genetic network growing on glucose and lactose.

The catalytic constant of the enzymes is set to be proportional to the enzyme solvent accessible surface area (SASA) and estimated from the enzyme molecular weight [63]. It is assumed that all biomass reactions are catalyzed by the ribosome R and their catalytic constants are obtained based on the translation elongation rate of 12 amino acids/s in *E. coli* [92]. It may be also needed to scale the obtained catalytic constants by an overall scaling factor (termed as f within this thesis) in order to match the cellular growth rate to available experimental data.

Table 3.2: Metabolic part of the deFBA model: Exchange and metabolic reactions with associated enzymes, and rate constants k_{cat} scaled by the factor f.

No.	Reaction	Gene	Enzyme	k_{cat}/ (min^{-1})
1	GLC + PEP → G6P + PYR	*pts*	pts	$8750 \cdot f$
2	LCT → GLC(in)	*lacY*	lacY	$1908 \cdot f$
3	GLC(in) + ATP → G6P + ADP	*glk*	glk	$1533 \cdot f$
4	G6P + ATP → 2T3P + ADP	*tpi*	tpi	$2133 \cdot f$
5	G6P + 6NAD$^+$ → T3P + 6NADH	*pgl*	pgl	$1585 \cdot f$
6	T3P + NAD$^+$ + ADP → PEP + NADH + ATP	*pgk*	pgk	$1740 \cdot f$
7	PEP + ADP → PYR + ATP	*pyk*	pyk	$5762 \cdot f$
8	PYR + CoA → AcCoA + FOR	*pfl*	pfl	$5676 \cdot f$
9	PYR + CoA + NAD$^+$ → AcCoA + NADH	*pdh*	pdh	$59728 \cdot f$
10	PEP → OAA	*ppc*	ppc	$9518 \cdot f$
11	OAA + AcCoA + NAD$^+$ → AKG + CoA + NADH	*icd*	icd	$3171 \cdot f$
12	AKG + 2NAD$^+$ + ADP → OAA + 2NADH + ATP	*mdh*	mdh	$2444 \cdot f$
13	AcCoA + ADP → ATP + CoA + ACT	*ackA*	ackA	$1809 \cdot f$
14	AcCoA + 2NADH → ETH + CoA + 2NAD$^+$	*adhE*	adhE	$52329 \cdot f$
15	4AKG +4NADH +8 ATP → 5AA + 4NAD$^+$ + 8ADP	*gdhA*	gdhA	$7560 \cdot f$
16	2NADH + O2 + 2.5 ADP → 2NAD$^+$ + 2.5ATP	*nuo*	nuo	$1935 \cdot f$

The complete deFBA model is then specified by the reactions in Tables 3.2 and 3.3 in combination with the biomass composition constraint (3.6).

We validate the deFBA model predictions using the available data for substrates, product, and biomass in aerobic and anaerobic batch growth of *E. coli* on glucose [5]. From the considered metabolic-genetic network of *E. coli*, the deFBA model can predict the cellular growth and metabolic dynamics under aerobic and anaerobic conditions. For the model validation, a scaling factor is implemented for all catalytic constants of metabolic reactions, in order to match the time for complete glucose substrate metabolization in aerobic and anaerobic conditions to the experimental data. For this network, a scaling factor of $f = 0.6$ for aerobic growth and $f = 0.65$ for anaerobic growth are required to meet the experimental batch times. As shown in Figures 3.4 and 3.5, despite of a slight overestimation in final products, predictions obtained from the deFBA model are in good agreement with the experimental data. One reason for the overestimation in final products by the model is that we only considered pathways for acetate and ethanol as fermentation products, while production pathways for other co-products like succinate are overlooked.

Next step, we use the validated model as a basis for *in silico* experiments and to derive dynamic manipulation strategies in order to enhance ethanol productivity.

Table 3.3: Genetic part of the deFBA model: Biomass reactions with values of weights, catalytic (k_{cat}) and degradation (k_{deg}) constants and initial conditions for biomass components. All biomass reactions are catalyzed by ribosome R.

No.	Biomass Reaction	b/ (g mol^{-1})	k_{cat}/ (min^{-1})	k_{deg}/ (hr^{-1})	$p(0)$/ (μM)
	Production reactions				
17	2358AA + 9432ATP → pts + 9432ADP	257022	0.31		0.00011
18	2346AA + 9384ATP → lacY + 9384ADP	255714	0.31		0.00014
19	321AA + 1284ATP → glk + 1284ADP	34989	2.24		0.00017
20	2272AA + 9088ATP → tpi + 9088ADP	247648	0.32		0.000087
21	2304AA + 9216ATP → pgl + 9216ADP	251136	0.31		0.0006
22	1775AA + 7100ATP → pgk + 7100ADP	193475	0.41		0.00075
23	1880AA + 7520ATP → pyk + 7520ADP	204920	0.38		0
24	1766AA + 7064ATP → pfl + 7064ADP	192494	0.41		0.00017
25	42096AA + 168384ATP → pdh + 16838ADP	4588464	0.02		0
26	3532AA + 14128ATP → ppc + 14128ADP	384988	0.2		0.000038
27	2928AA + 11712ATP → icd + 11712ADP	319152	0.25		0.00011
28	4186AA + 16744ATP → mdh + 16744ADP	456274	0.17		0
29	1124AA + 4496ATP → ackA + 4496ADP	122516	0.64		0.00032
30	35640AA + 142560ATP → adhE + 142560ADP	3884760	0.02		0
31	5866AA + 23464ATP → gdhA + 23464ADP	639394	0.12		0.000012
32	1795AA + 7180ATP → nuo + 7180ADP	195655	0.4		0.00146
33	7459AA + 7420G6P + 38968ATP → R + 38968ADP	2289611	0.1		0.00063
34	466AA + 11G6P + 20T3P + 7144ATP + 1527NADH → Q + 7144ADP + 1527NAD$^+$	54980	1.55		0.05
	Degradation reactions				
35-50	Enzyme$_i$ → n_i AA (e.g. pts → 2358AA)			0.014	
51	R → 7459AA + 7420G6P			0.007	
52	Q → 466AA + 11G6P + 20T3P			0.007	

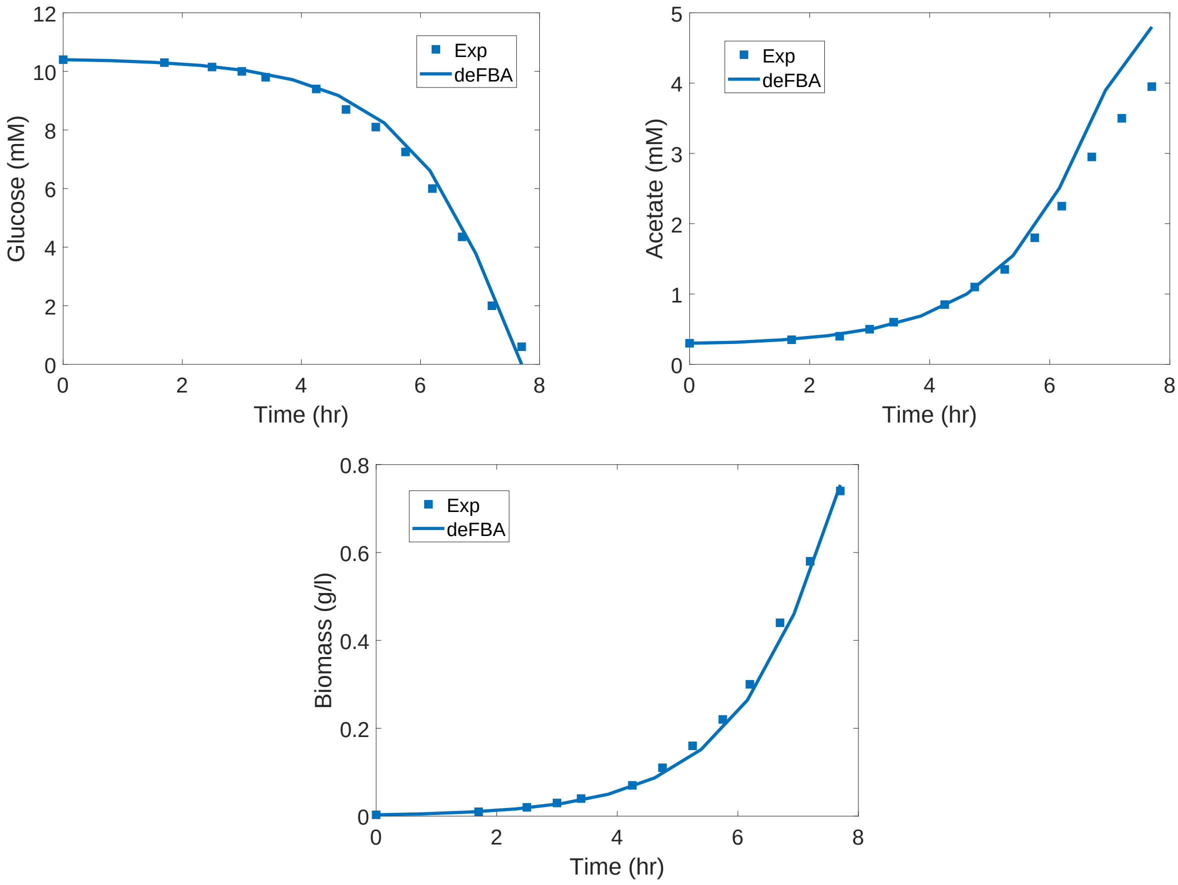


Figure 3.4: Aerobic growth of *E. coli* on glucose: substrate, products and biomass concentration profiles simulated by the deFBA model together with experimental data obtained by [5].

3.5.2 Control strategies for dynamic genetic- and process-level manipulations

Based on the derived metabolic-genetic network model, we aim to implement suitable strategies to temporally optimize a batch fermentation of *E. coli*, resulting in maximized ethanol productivity.

With glucose and lactose as substrates under anaerobic conditions, *E. coli* produces a mixture of products, including succinate, acetate, lactate and ethanol, in order to maintain the redox balance [93]. For increased ethanol production in the cell, one possible strategy is to delete pathways for the competing co-products, which leads to an increased carbon flux toward the ethanol pathway. However, deleting the gene responsible for acetate formation ($ackA$) significantly reduces the growth rate under anaerobic conditions, since this pathway also generates ATP and its deletion reduces the ATP production. Due to the associated growth rate reduction, this reduces productivity [94]. Therefore, we consider deleted pathways for succinate and lactate production, but follow a two-stage dynamic optimization strategy with a manipulation of the acetate pathway for improving the productivity.

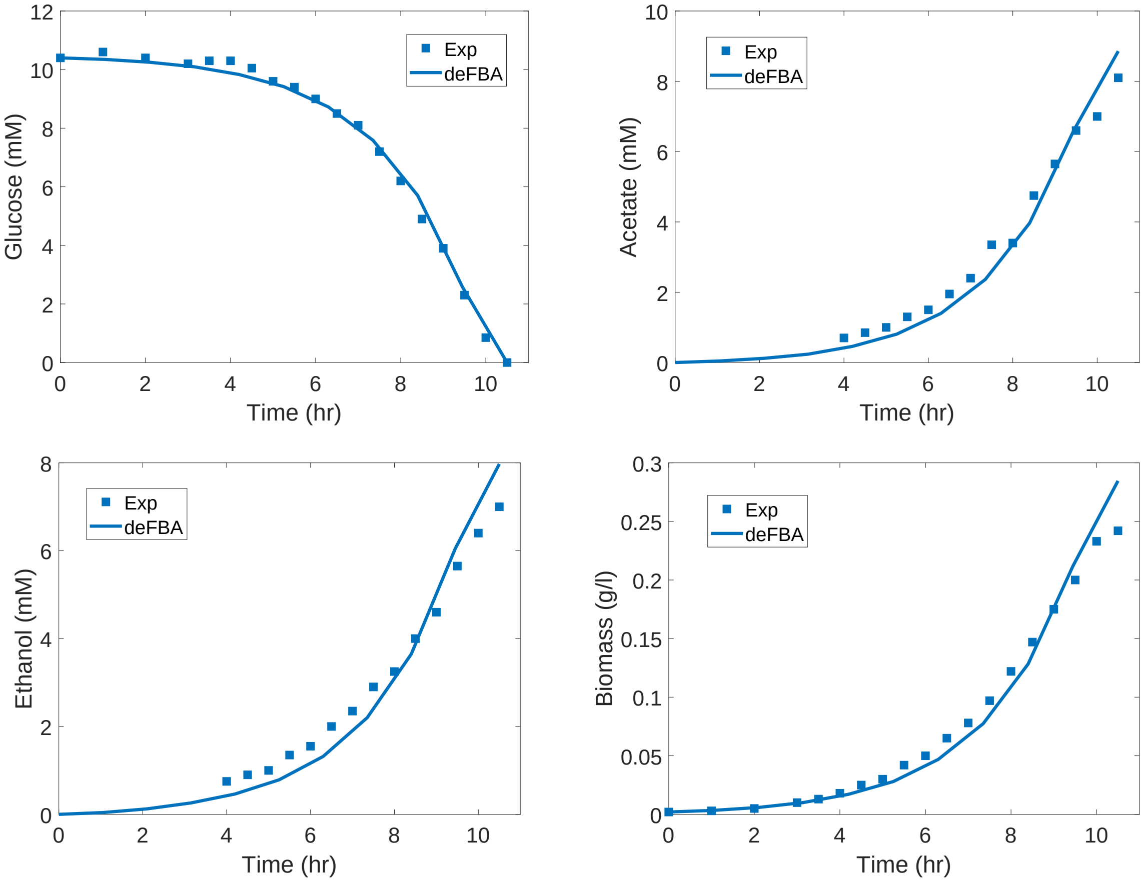


Figure 3.5: Anaerobic growth of *E. coli* on glucose: substrate, products and biomass concentration profiles simulated by the deFBA model together with experimental data obtained by [5].

Table 3.4: Regulation constraints for the bilevel optimization problem.

Strategy 1: Acetate pathway regulation		Strategy 2: Static *ackA* knockout + aerobic-anaerobic switch
Repression of gene expression	Repression of metabolic flux	
$V_{29}(t) = 0$ for $t \geq t_{reg}$	$V_{13}(t) = 0$ for $t \geq t_{reg}$	$V_{29}(t) = 0$ for $t \geq 0$, $V_{16}(t) = 0$ for $t \geq t_{reg}$

Table 3.5: Initial nutrient conditions and initial biomass $b^T p(0)$.

Glucose (GLC)	Lactose (LCT)	Biomass, $b^T p(0)$
5 mM	30 mM	0.005 gl^{-1}

Here, we implement the dynamic bilevel optimization problem for maximizing ethanol productivity with two dynamic manipulation strategies: temporal genetic regulation, and a switch from aerobic to anaerobic conditions.

In order to come up with high ethanol productivity, we consider first a dynamic manipulation of the acetate formation pathway. This maintains a high cell growth rate in anaerobic conditions during the initial growth stage. For the switch to the production stage, the flux through the acetate pathway is repressed. Through the bilevel optimization problem, we seek the optimal time point for switching from the growth stage (with an active acetate pathway) to the production stage (with a repressed acetate pathway and increased ethanol formation).

As the second strategy, we consider a combined aerobic-anaerobic batch growth of *E.coli*. For enhanced production of ethanol, we propose static knockouts of competing pathways, including the acetate pathway by deletion of the corresponding gene *ackA*. Instead, a high growth rate during the growth stage is achieved through maintaining aerobic conditions. At the switch to the production stage, the oxygen supply to the culture is removed and the culture then grows anaerobically to produce ethanol. In this case, the productivity of ethanol depends on the switch time from aerobic to anaerobic conditions.

The specific regulation constraints for each of the considered strategies are presented in Table 3.4. For the dynamic repression of the acetate pathway (Strategy 1), we include two types of regulation: a repression of *ackA* gene expression and a direct repression of the corresponding metabolic flux.

Initial nutrient conditions are summarized in Table 3.5.

Table 3.6: Bilevel optimization results for different strategies.

Parameter	Static *ackA* knockout	Strategy 1: Acetate pathway regulation		Strategy 2: Static *ackA* knockout + aerobic-anaerobic switch
		Repression of gene expression	Repression of metabolic flux	
Batch time, t_f (hr)	14.6	12.6	12.2	10.7
Regulation time, t_{reg} (hr)	0	1	6.1	5.3
Produced ethanol (mM)	52.7	48.5	50	42.3
Productivity (mM hr^{-1})	3.61	3.86	4.1	3.95

3.6 Results and discussion

3.6.1 Higher productivity by dynamic strategies

The bilevel optimization framework defined as problem (3.3) is applied to obtain the optimal time points of the considered manipulations (outlined in Section 3.5.2) for maximal ethanol productivity. For the two implemented dynamic strategies (dynamic acetate pathway regulation and aerobic-anaerobic switch), the optimal results are presented in Table 3.6, and concentration profiles of substrates, products and produced biomass are shown in Figure 3.6 A-C. Table 3.6 also contains results for the anaerobic condition in which no dynamic regulation is considered and instead the acetate pathway is inactive throughout the batch (by knocking out the gene *ackA*). A higher productivity is obtained for all of the dynamic strategies compared to the static manipulation. In fact, deleting the acetate pathway at the start of the batch leads to a biomass growth reduction. This, in turn, means that a longer batch time is needed to consume all the substrates, and thus ethanol is produced with lower productivity but with higher yield in comparison to the dynamic strategies (Figure 3.6 D).

3.6.2 Different manipulation strategy and productivity by each variant of Strategy 1

For Strategy 1, our results in modulating expression of target genes instead of static gene deletions are qualitatively in agreement with previous studies on dynamic genetic alterations for higher productivity [37, 86]. However, in these works, genetic manipulations which involve the down-regulation of gene expression have been described in the model by directly repressing the corresponding metabolic fluxes, because the implemented models (FBA and dFBA) did not allow describing a direct manipulation of the gene expression. With the deFBA model, it is possible to distinguish genetic manipulations explicitly from a manipulation of the metabolic flux, and our results indicate that repressing the production pathway of the target enzyme (ackA) results in a different productivity compared to repressing the corresponding metabolic flux (acetate flux). Even more importantly, the optimal regulation time for the maximal

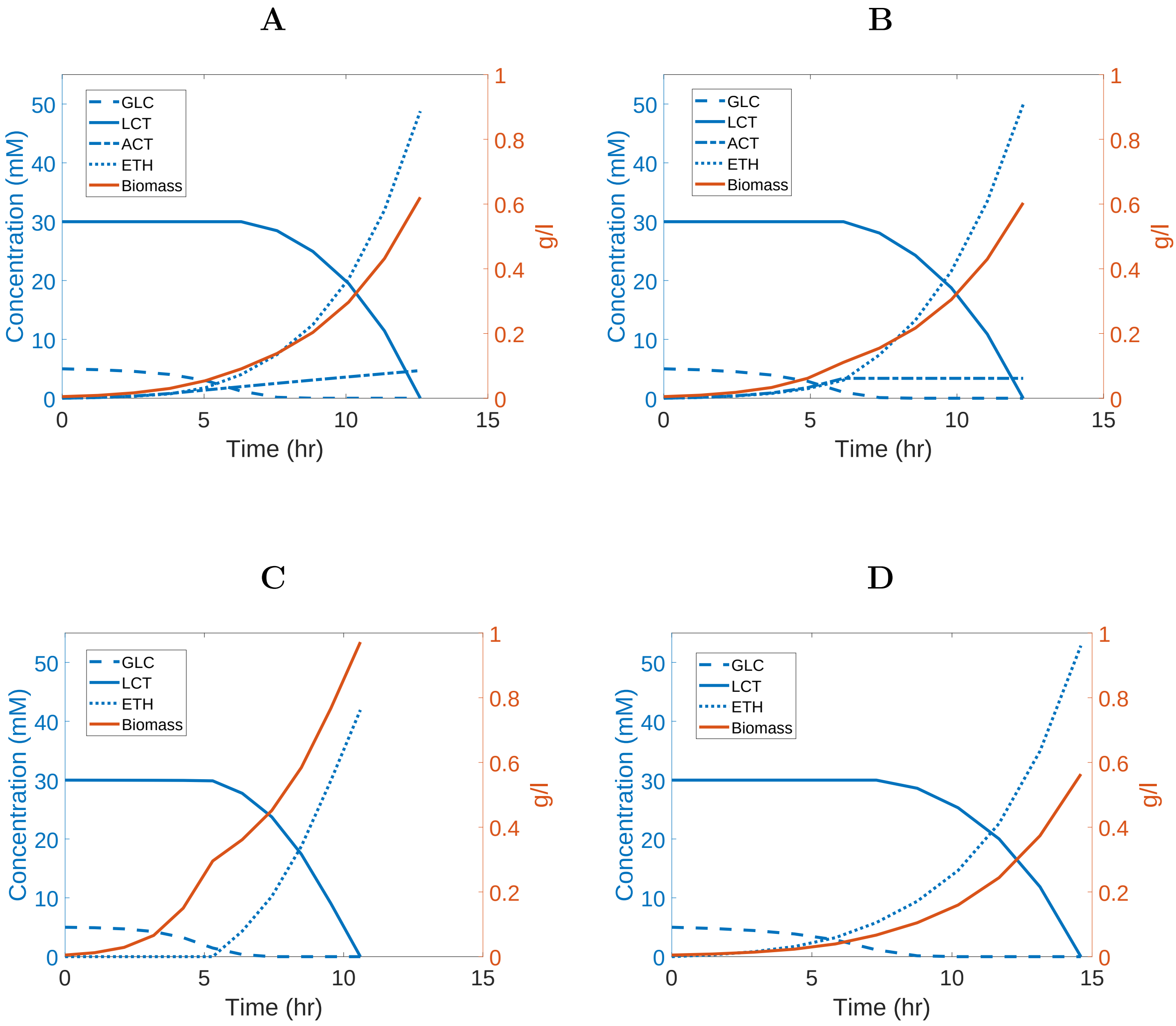


Figure 3.6: Optimal concentration profiles of substrate, products and biomass resulting from the bilevel optimizations for several dynamic regulation strategies: (A) Strategy 1 with repressing the production pathway of ackA enzyme. (B) Strategy 1 with repression of the acetate flux. (Note that the optimal batch time, regulation time and concentration profiles are different for different variants of Strategy 1.) (C) Strategy 2 with a static *ackA* knockout and a switch from aerobic to anaerobic conditions. (D) Concentration profiles with a static *ackA* knockout with no dynamic manipulation. (GLC: glucose, LCT: lactose, ACT: acetate, ETH: ethanol).

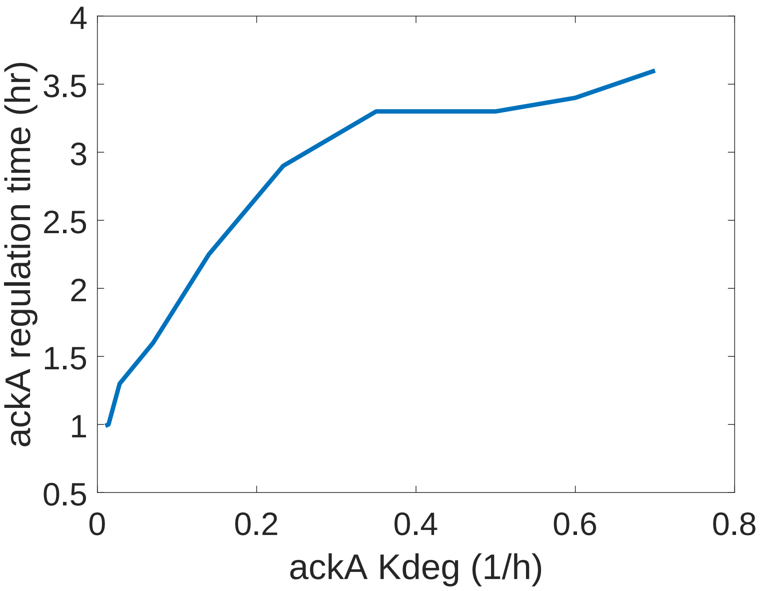


Figure 3.7: Optimal time for repressing the production pathway of ackA enzyme as a function of its degradation constant, k_{deg}. The optimal regulation time increases with the degradation rate constant of the enzyme.

productivity is also quite different in these two variants of Strategy 1 (Table 3.6). The bilevel optimization suggests a relatively short time (1 h) to repress the production of the ackA enzyme. However, the ackA enzyme already produced up to that point continues to catalyze acetate production until the end of the batch (Figure 3.6 A). This continuous production of acetate with low rate results in lower productivity for ethanol in comparison to the case in which the regulation is applied directly to the metabolic flux of the acetate pathway (Figure 3.6 B). Thereby the rate of on-going metabolic flux depends on the target enzyme degradation rate compared to the process batch time.

For this network, an average half-life of 50 hours is considered for ackA based on the protein half-lives in *E. coli* [60], which means it degrades slowly relative to the process time. However, different degradation rates for the target enzyme lead to different optimal regulation times for repressing the production pathway of that enzyme. For ackA, the effect of the degradation rate on the predicted regulation time for maximal productivity is shown in Figure 3.7. From that, it can be noticed that for a higher degradation constant, k_{deg} of ackA (faster degradation rate), the optimal time for repression of the ackA production pathway (reaction 29) is predicted to be later in contrast to a situation where the target enzyme degrades slowly (lower degradation constant).

As discussed above, the deFBA model allows to distinguish between a dynamic repression of just the gene expression vs. a dynamic inactivation of the enzyme itself and the associated repression of the metabolic flux. Our bilevel optimization results suggest that with dynamic regulation, repressing the competing metabolic flux leads to a higher productivity (+ 0.24 mM/h) compared to just the repression of the associated gene expression. So, for a practical implementation of this strategy, besides designing methods to manipulate the expression of a target gene by inducible and repressible promoters [81, 82, 84, 86], other techniques for the dynamic manipulation of the enzymatic activity such as implementing inducible enzyme degradation (by degradation tags) should be taken into account [95], as the enzyme already synthesized may persist without active degradation. In contrast, the models previously used do not permit a distinction between genetic and metabolic flux regulation, and in principle only consider the repression of the metabolic flux.

In a recent study about increasing the productivity of itaconic acid production by *E. coli* using a two-stage process strategy [96], it has been noticed that there is a considerable delay when switching from the growth stage to the production stage through a dynamic genetic repression in the TCA cycle. The delay they noticed between the time of implementing the genetic repression, and the time in which the corresponding flux is turned off, is due to the activity of pre-existing enzymes, which matches well to our observations in this work. These results imply that it is highly important to take enzyme expression pathways into account during the model optimization, especially when planning for dynamic manipulations on the gene expression level only.

3.6.3 Robustness of a selected manipulation strategy

From the results in Table 3.6, we see the deFBA-based bilevel optimization problem predicts that among implemented dynamic strategies, the second variant of Strategy 1 (dynamic repression of metabolic flux) provides the highest productivity for this particular network as well as the highest yield. The results also suggest instead of just a dynamic gene knockout (first variant of Strategy 1), the aerobic-anaerobic switch seems to be more promising as besides giving higher productivity, it should also be less challenging to be implemented technically.

As the enzyme capacity constraint is a very important constraint within the deFBA model, we perform a sensitivity analysis on the catalytic constants of metabolic reactions. Besides comparing the relative importance of the different enzymes, this sensitivity analysis also allows to evaluate the robustness of the above comparison between strategies obtained from the bilevel optimization.

For the sensitivity analysis, the values of the catalytic constants are individually perturbed to $k_{cat,new} = k_{cat}/2$, iterating over all metabolic reactions. For each perturbation, the deFBA model is solved with the network from Section 3.5.1. From the perturbed solutions, we calculate the biomass growth rates in both aerobic and anaerobic conditions, as shown in Table 3.7. It can be noticed that changing the k_{cat} value of reaction 6 (enzyme pgk) has the biggest impact on the anaerobic growth rate, and it is also among the most influential reactions in aerobic conditions. Thus, for the robustness analysis of the bilevel optimization results, we specifically consider an uncertainty in this value.

Performing the bilevel optimization with a range of different k_{cat} values for reaction 6, we find the same ordering of strategies with respect to productivity as in the nominal case for -30% up to +35% changes in its original value. For k_{cat} values less than -30% of the original value, the Strategy 2 (aerobic-anaerobic switch) becomes the strategy with the highest productivity, instead of the second variant of Strategy 1 (dynamic repression of metabolic flux). On the other hand, for values higher than +35%, the second variant of Strategy 1 has still the highest productivity, but the first variant of Strategy 1 (dynamic repression of gene expression) surpasses Strategy 2. Therefore, we find the range of parameter uncertainties that the ordering of optimal strategies is robust against, which sounds to be wide enough. However it should be noted that large parameter perturbation can of course affect the predictions from the optimization, and a good estimate of these constants is still desirable.

In this way, the proposed model and bilevel optimization approach can be helpful for decision making between different control strategies for metabolic engineering and bioprocess design purposes.

Table 3.7: Comparison of biomass growth rates for changes in catalytic constants of metabolic reactions ($k_{cat,new} = k_{cat}/2$)

Reaction No.	anaerobic growth rate ratio μ_{new}/μ	aerobic growth rate ratio μ_{new}/μ
1	0.9775	0.989
2	0.9449	0.9634
3	0.9943	0.9996
4	0.8955	0.9572
5	0.9958	0.9189
6	0.8268	0.9279
7	0.9602	0.9779
8	0.9377	1
9	1	0.9785
10	0.9864	0.9778
11	0.9775	0.9671
12	1	0.9778
13	0.9379	1
14	0.9366	1
15	0.9948	0.9967
16	1	0.8579

3.6.4 Optimal cellular resource allocation during the two-stage growth

From the optimization prediction by the deFBA model shown in Figure 3.6, one can notice that the cells first take up glucose and after its depletion, they switch to lactose according to the glucose-lactose diauxie experimentally observed in *E. coli*. This uptake pattern can be explained as the optimal solution of the deFBA model with the metabolic-genetic network of *E. coli*. One reason is that the uptake reaction for glucose has a higher enzymatic efficiency than the uptake reaction for lactose. Moreover, glucose uptake and transformation to glucose-6-phosphate (reaction 1) require less investment in terms of enzyme production cost compared to the lactose uptake and transformation to glucose-6-phosphate (reactions 2 and 3).

It has been already shown that the sequential uptake of substrates limits the ability of the microbial culture to efficiently produce bio-products and results in lower productivities [36]. Therefore, when a microorganism shows a sequential uptake pattern for available substrates, it is highly demanded to apply models which include the effect of this mechanism on the productivity of the target product. In other words, implementing models which cannot capture the catabolite repression by substrates (like FBA and dFBA models previously used in the bilevel optimization for maximal productivity) may not result in accurate predictions of an optimal strategy for metabolic engineering purposes with mixed substrates and in practice we probably achieve a lower productivity than expected based on the model prediction.

For further evaluation on the optimal resource distribution within the cell during the two-stage growth, we provide the biomass composition for the implemented strategies. Figure 3.8 represents the percentage of some key individual enzymes in case of dynamic regulations (A-C) and static ackA knockout(D). As it can be seen, for both variants of Strategy 1, the percentage of ackA is high in the growth stage, and then in the production stage (after the time which regulation is applied) its percentage is reduced. On the other hand, the percentage of enzyme adhE (responsible for ethanol production) is mainly upward leading to ethanol production with higher production rate in the production stage. The enzyme pfl (responsible for producing AcCoA from pyruvate under anaerobic condition) is active during the whole batch time. It should be mentioned that as expected no contribution of pdh (used for AcCoA production from pyruvate in aerobic condition) is predicted by the model in Strategy 1 with the anaerobic growth.

For the Strategy 2, the percentage of enzyme nuo (responsible for ATP production in aerobic growth) is increased at the first phase of the growth (aerobic growth) and then starts to decrease around the regulation time until the end of the batch. On the other hand, at the second phase of the growth (switch to anaerobic condition) the percentage of enzyme adhE (for ethanol production) is increased in order to promote producing ethanol. Interestingly, in Strategy 2 it is noticed that when switching from aerobic to anaerobic condition, the model predicts down-regulation of pdh (aerobic

AcCoA production) and up-regulation of pfl (anaerobic AcCoA production) without considering any regulatory information. Using the genetic part of the deFBA model, we infer that is probably because pdh is a very expensive enzyme and the cell prefers to switch it down as soon as there is no more need to produce additional NADH required for respiration through the AcCoA production pathway (metabolic pathway 9).

The preference of the cell for glucose uptake can be again noticed in Figure 3.8, from the profile of pts (responsible for glucose uptake) which is active from the beginning, and profiles for lacY and glk (responsible for lactose uptake and transformation) which their concentrations increase in the later stages of the growth.

3.6.5 Computational limitations

For the implemented deFBA-based bilevel optimization, keeping the inner model linear allows for an efficient numerical solution. For the simulations, we consider $N = 10$ time intervals and $K = 2$ collocation points within each interval. Based on that, the resulted LP of the deFBA model consists of 1400 decision variables and 1260 algebraic constraints. Each solution of this LP requires 1.2s on a standard desktop computer, and on average 30 iterations on the inner model are needed to perform the outer optimization. The required CPU time for the inner problem depends on the number of collocation steps, which defines the number of variables within the inner problem (as explained in Section 2.3.2). Although in this work we implement a small-scale network, deFBA models with larger networks (up to around 500 reactions) have already been efficiently simulated, e.g. [53, 97]. We think that a bilevel optimization is still feasible with large-scale models.

Although implementing a linear model for the inner problem is an advantage from the computations required, it can be a potential limitation of the deFBA model that flux bounds are approximated to change linearly with respect to the enzyme concentration, and do not depend on substrate metabolite concentrations. This is an approximation from the commonly used nonlinear reaction kinetics, such as the Michaelis-Menten kinetics for substrate uptake, or inhibitory terms that reflect growth rate suppression by the presence of particular species. It has been argued by [97] that due to metabolic enzymes working mostly at saturation and also typically low substrate k_m values (e.g. 0.015 mM for glucose [15]), this approximation does not induce a significant error. However, it should be noted that for applications with substrate-relevant growth inhibition, such an approximation works for low substrate concentrations and in case of higher levels of inhibition it can impose significant prediction errors.

It is possible to formulate the deFBA model using nonlinear constraints, but with a much higher computational cost for the associated nonlinear optimization. While this will probably be prohibitive for the bilevel approach, one could still try to validate the results for the optimal strategy using a more elaborate model. Besides, for the deFBA model with nonlinear constraints one can also consider approaches for approximating

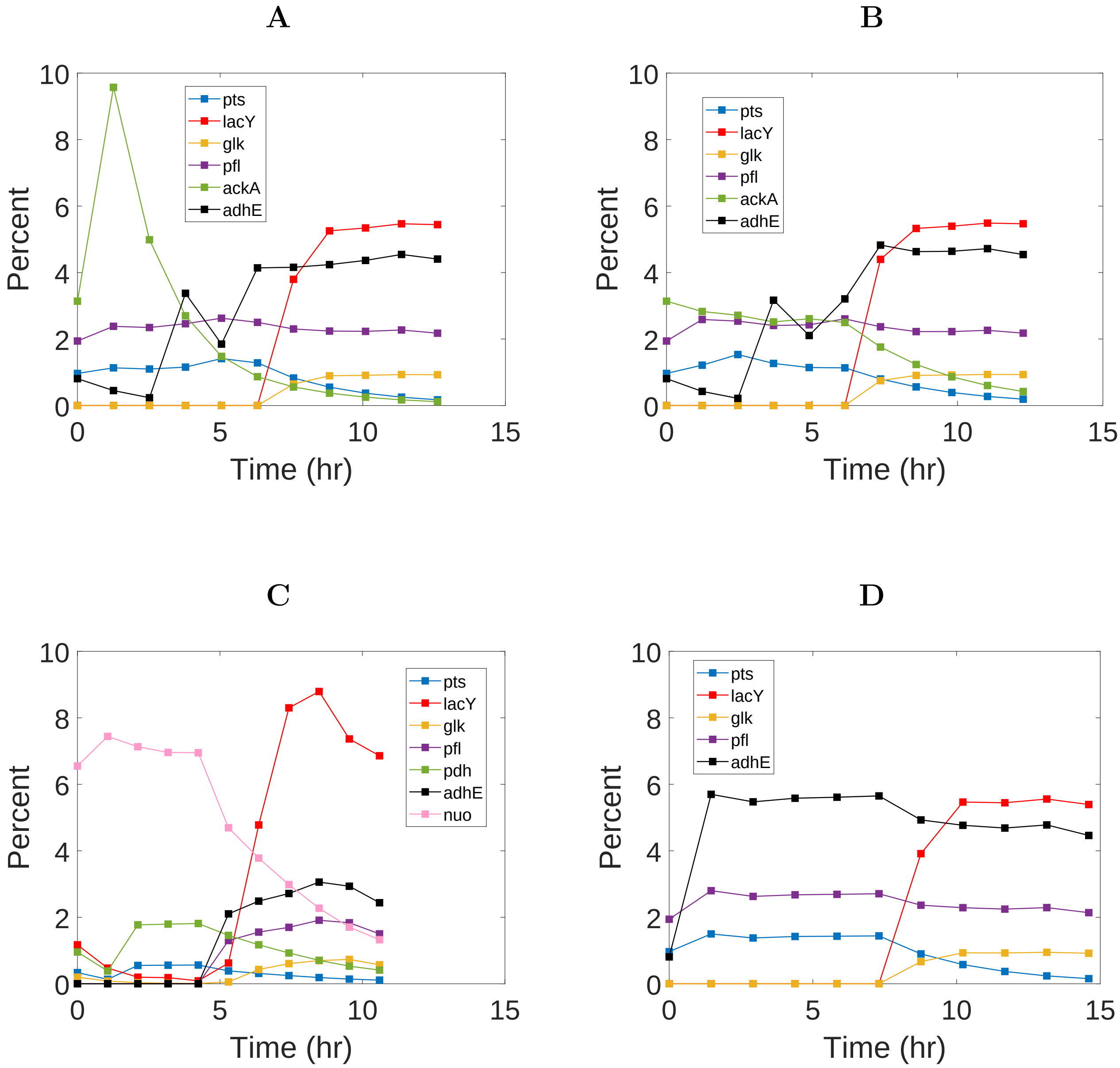


Figure 3.8: Percentage of total biomass (dry-weight) for key enzymes in the optimal solution for several regulation strategies: (A) Strategy 1 with repressing the production pathway of ackA enzyme. (B) Strategy 1 with repression of acetate flux. (C) Strategy 2 with a static *ackA* knockout and a switch from aerobic to anaerobic conditions. (D) Static *ackA* knockout with no dynamic manipulation. The distribution of enzymes is rearranged in different stages of each growth modes.

the nonlinearity with one or multiple linear constraints in order to maintain efficient computation times while still keeping a sufficient model accuracy [17, 19].

As a relevant evaluation, we briefly outline such an approximation we have implemented in [17] for a nonlinear dFBA model. The model is similar to the DOA formulation of dFBA presented in Section 2.2, with an additional inequality constraint on the substrate uptake to address growth inhibitions. To deal with the nonlinear dynamics of dFBA, we use the absolute metabolic flux which is not defined per gram of biomass. To do so, we define the absolute metabolic flux V, as a multiplication of the metabolic flux per gram of biomass v, and the biomass concentration X,

$$V(t) = v(t)X(t). \tag{3.7}$$

This transformation results in linear dynamics of the dFBA model, while the growth inhibition constraint still remains nonlinear with a general form of

$$V^{uptake}(t) \leq \frac{v_{max}^{uptake} X(t)}{1 + z(t)/K_I}, \tag{3.8}$$

in which the uptake flux of the substrate V^{uptake} is reduced as the concentration of a metabolite z increases (v_{max}^{uptake} and K_I are the predefined maximum uptake flux and the metabolite inhibition constant, respectively). We approximate this constraint based on a tangent plane approximation of the function

$$g(X, z) = \frac{v_{max}^{uptake} X(t)}{1 + z(t)/K_I}. \tag{3.9}$$

Based on the shape of this function's graph, two linear approximations are performed at two points of this surface $(X_1,\ z_1,\ g(X_1, z_1))$ and $(X_2,\ z_2,\ g(X_2, z_2))$ by defining the corresponding tangent planes as $L_1(X, z)$ and $L_2(X, z)$. Conceptually, each of the planes can be considered as an approximation to the original surface, and the constraint in (3.8) can be approximated by the mixed logical/linear constraint

$$\begin{aligned} V^{uptake} &\leq L_1(X, z), \quad \text{or} \\ V^{uptake} &\leq L_2(X, z). \end{aligned} \tag{3.10}$$

This is conceptually shown in Figure 3.9 in a two dimensional space, with the function $g(x)$ to be approximated by lines $L_1(x)$ and $L_2(x)$.

The constraint (3.10) is numerically implemented by a time-dependant binary variable λ, such that:

$$\begin{aligned} V^{uptake}(t) - L_1(X(t), z(t)) &\leq m\lambda(t), \\ V^{uptake}(t) - L_2(X(t), z(t)) &\leq m(1 - \lambda(t)), \end{aligned} \tag{3.11}$$

with m as a positive large constant ($m >> 0$). Each value of $\lambda \in (0, 1)$ determines

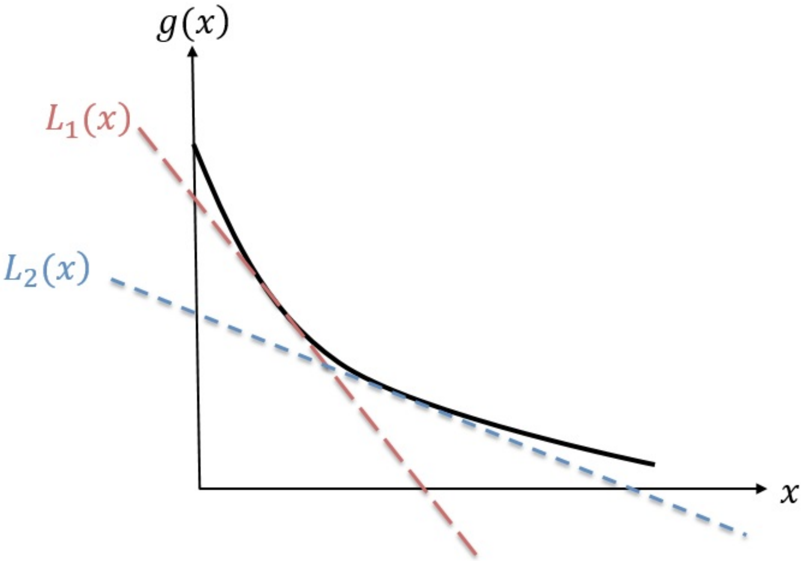


Figure 3.9: Approximation of $g(x)$ by lines $L_1(x)$ and $L_2(x)$.

which plane (L_1, or L_2) should be used for the approximation and in fact which of these inequality constraints in (3.11) is active during the optimization.

After replacing the nonlinear constraint (3.8) by the two linear constraints in (3.11), the dynamic dFBA optimization can be solved using the collocation method (with the procedure explained in Section 2.3.2). Due to the presence of binary variables λ, discretization of variables results in a mixed-integer linear problem (MILP) which can then be solved numerically with a suitable solver, such as *intlinprog* in MATLAB.

In [17], the dFBA model has been implemented within a bilevel formulation for improving bioprocess productivity by dynamic manipulation of the metabolic network. Through implementing the dFBA-based bilevel problem for a case study, it was shown that the bilevel optimization based on MILP (for dFBA with constraint approximations) gives optimal predictions for maximal productivity very close to those from the bilevel optimization with NLP (corresponding to the original nonlinear dFBA without approximating the nonlinear constraint), but with much lower computational cost. For additional details relevant to this part, the bilevel optimization problems for maximal productivity based on nonlinear and linearized dFBA models are presented in Appendix A.

To reduce the approximation errors, selection of appropriate points for defining planes is however important for surface approximation, such that the resulting planes cover as much of the area below the surface as possible. In case of nonlinear constraints with high degree of nonlinearity, the possible treatment would be considering more approximation points to increase the accuracy of approximation, which in turn increases the number of binary variables.

Therefore, linear approximations can be a possible remedy to reduce the computational cost associated with nonlinear deFBA optimizations, while the linearized model is still enough accurate for the growth simulation. Within the following chapter, a linear approximation of a nonlinear deFBA model (implemented for closed-loop control of fed-batch bioreactors) is considered, which provides more details on the tangent plane approximation.

3.7 Conclusion

In this chapter, we have implemented a bilevel optimization framework in order to improve biochemical productivity via a dynamic regulation of metabolic-genetic networks using genetic and process level manipulations. The constraint-based deFBA model has been applied in order to find the optimal regulation times for maximal productivity of product formation with a bilevel dynamic optimization approach.

The method has been employed to design a regulation strategy for improved ethanol production by *E. coli*, by finding the optimal manipulation time for (1) repressing the acetate pathway and (2) switching from aerobic to anaerobic conditions. Generally, the dynamic manipulation approaches result in a higher productivity, but lower yield compared to the static gene knockout. This offers a degree of flexibility to choose between maximum productivity of the target product and maximum yield of the target product at the expense of a longer batch time.

As discussed, the manipulation of the acetate pathway can involve different levels such as regulation of the metabolic flux vs. regulation of the enzyme expression. The difference in results for the manipulations on different levels captured by deFBA indicates that the construction and utilization of models that can describe the intended manipulation strategy is essential in achieving reliable predictions.

Overall, the deFBA-based bilevel problem can be a good basis for open-loop optimization of the process for deriving process- and gene- level control strategies, when the derived metabolic-genetic networks model represents the real process with high accuracy. However, possible uncertainties and modeling errors may exist resulting in an overall performance far from the optimum. This is an important aspect of model-based optimization which is taken into account in following chapters employing model-based control strategies, taking care of the process uncertainties.

4 Model predictive control of the bioprocess based on dynamic constraint-based models

In this chapter a closed-loop control of the bioprocess is considered by introducing measurement feedback into the open-loop strategies, in order to handle uncertainties in modelling parameters. The bilevel problem developed in the previous chapter is implemented within a model predictive control routine for online adjustment of controller states based on measurements. The control problem is specifically developed for a fed-batch culture and through the online control it is aimed to obtain fed-batch operating policies including the substrate feeding and process-level regulation of metabolism for optimizing the productivity of a target product. An example of fed-batch fermentation of *E. coli* for maximal ethanol productivity is considered to evaluate the advantages of feedback corrections.

4.1 Closed-loop control of the bioprocess

In the previous chapter, we have implemented the dynamic enzyme-cost FBA model within a bilevel framework to improve bioprocess productivity. The deFBA-based bilevel approach as the open-loop control problem could improve bioprocess productivity by determining optimal strategies for dynamic manipulation of cellular metabolism in both genetic and process levels.

Through the open-loop control, operating strategies are determined once through the offline optimization and there is no consideration of measurement feedbacks. Such a offline control relies completely on the accuracy of the underlying process model for reproducing responses close to the real process based on the derived profiles of the manipulated variables.

As an open-loop control of the bioprocess in our work the deFBA-based bilevel optimization is solved based on the defined parameters and a given initial condition and then the determined control sequences are applied to the plant (Figure 4.1 A). However, uncertainties are inevitable in modeling and simulation of bioprocesses due to the highly variable nature of biological systems. Because of the existing uncertainties and mismatches the model may not be able to predict the real process dynamics accurately, which results in a poor performance of the open-loop control and suboptimal production processes. In practice, improved performance can be obtained by introducing measurement feedback into the open-loop operating strategy [98]. In this direction, in this chapter we consider a closed-loop control of the bioprocess by inte-

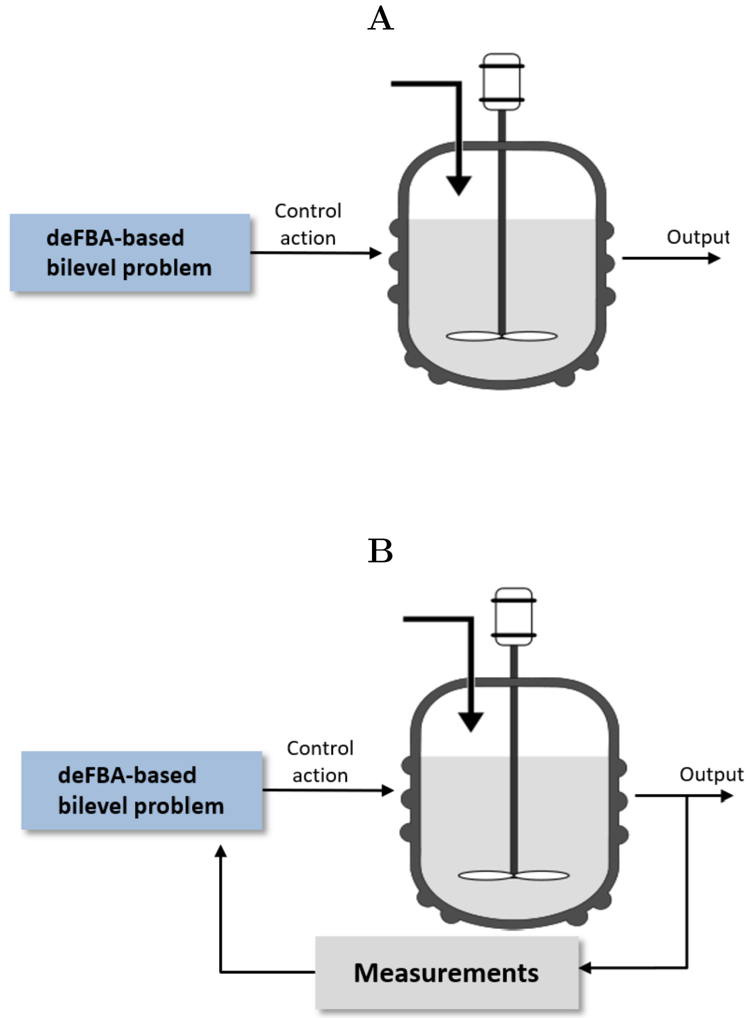


Figure 4.1: Open-loop (A) vs. closed-loop (B) control of the bioprocess.

grating the information from online measurements into the open-loop deFBA-based bilevel optimization (Figure 4.1 B).

4.2 Optimization and control of a fed-batch bioreactor

In this chapter, we consider closed-loop control of the bioprocess for improved productivity specifically in fed-batch process modes.

Many biotechnological processes rely on fed-batch operation modes, where one or more substrates are fed to the bioreactor gradually during the process, while the products remain in the bioreactor until the end of the run. A fed-batch bioreactor allows for better control of the nutrient level for favourable growth conditions. This is in fact the superior cultivation mode when controlling concentrations of a nutrient affect the yield or productivity of the desired product. A common example is the case where some nutrients inhibit the growth of microorganisms even at relatively low concentrations. By adding such substrates properly, their inhibitory effects can be significantly reduced resulting in improved process rates.

In this direction, an efficient production process requires determination of fed-batch operating policies due to their strong effects on cellular metabolism. This means optimal balance between cell growth and product formation and optimal process productivity is highly dependant to keeping an optimal level for nutrients inside the bioreactor. In this direction, we implement the deFBA-based bilevel problem to find the optimal fed-batch operating strategies including the substrate feeding and process-level regulation of metabolic network that achieves maximal productivity for a target metabolite. Afterwards, we implement the deFBA-based bilevel problem within a model predictive control (MPC) [99] which allows for feedback corrections. It involves repeating the bilevel optimization while it is updated by online measurements.

4.2.1 Fed-batch growth simulation

Here we implement the deFBA model to simulate the cellular growth in a fed-batch culture. To address the substrate feeding during the process, the previously presented batch growth model (problem 2.19) is revised accordingly to the following form

$$\underset{V(\cdot)}{\text{maximize}} \quad \int_{t_0}^{t_f} v(t) b^T p(t) dt \tag{4.1a}$$

$$\text{s.t.} \quad \frac{d(vz)}{dt} = z_f F(t) + S_z^z V_z(t), \qquad \frac{d(v)}{dt} = F(t), \tag{4.1b}$$

$$\frac{d(vp)}{dt} = S_p^p V_p(t), \tag{4.1c}$$

$$S_m V(t) = 0, \tag{4.1d}$$

$$\sum_{j \in cat(i)} |V_j(t)/k_{cat,j}| \leq v(t) p_i(t), \qquad i \in \mathbb{E} \tag{4.1e}$$

$$V_{min}(t) \leq V(t) \leq V_{max}(t), \tag{4.1f}$$

$$\varphi_Q b^T p(t) \leq p_Q(t), \qquad Q \in \mathbb{Q}, \tag{4.1g}$$

$$V^{uptake}(t) \leq V_{max}^{uptake} \frac{z_i(t)}{K_m + z_i(t) + z_i^2(t)/K_I}, \tag{4.1h}$$

$$x(t_0) = x_0 = (z_0, p_0), \qquad v(t_0) = v_0, \qquad x(t) \geq 0, \tag{4.1i}$$

where v is the liquid volume and F, z_f are the feed flow rate and the feed concentration, respectively. In the case of growth inhibitions by the substrate (a relevant consideration for the fed-batch mode as explained earlier), additional constraints should be taken into account. Here, growth inhibitions by the substrate is addressed by defining a constraint on substrate uptake flux, V^{uptake} (4.1h) which reduces the maximum substrate uptake rate when high concentrations of the substrate are available. In its general form, (4.1h) is a component-wise inequality constraint, where z_i is the specific substrate with inhibition on its uptake flux V^{uptake}, and K_m and K_I are saturation and inhibition constants. V^{uptake}_{max} is the maximum uptake rate of the inhibiting substrate. For more precise definition within the deFBA, V^{uptake}_{max} can be replaced by $k_{cat}v(t)p(t)$, as the substrate uptake based on the maximum capacity of the corresponding enzyme. It should be mentioned that for simplicity here we assume very low degradation rates for enzymes and neglect the degradation term from mass balance equations (4.1c-d).

The resulted nonlinear dynamic optimization Problem (4.1) simulates the fed-batch growth to be used further for bioreactor optimization and control.

4.2.2 Control problem for productivity maximization

The deFBA model for the fed-batch growth simulation then is used within a bilevel framework to define the open-loop control problem for maximizing the productivity of the target product x_t. Through the problem (4.2), we are looking for a feeding profile such that the substrate concentration is optimal within the bioreactor, along with the optimal timing for the two-stage dynamic control of the metabolism.

$$
\begin{aligned}
\underset{F(t),t_f,t_{reg}}{\text{maximize}} \quad & \frac{x_t(t_f)}{t_f} \\
\text{s.t.} \quad & t_0 \leq t_{reg} \leq t_f, \quad v(t_f) \leq v_{max}, \\
& \underset{V(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} v(t) b^T p(t) dt \\
& \text{s.t.} \quad (4.1\text{b}) \quad \text{to} \quad (4.1\text{h}), \\
& \qquad x(t_0) = x_0 = (z_0, p_0), \qquad v(t_0) = v_0, \qquad x(t) \geq 0, \\
& \qquad V_{reg}(t) = 0 \quad \text{for} \quad t \geq t_{reg}.
\end{aligned}
\tag{4.2}
$$

Fed-batch operating conditions including feed flow rate F, the optimal batch time t_f and the switch time between growth and production stages t_{reg} (by regulation of a target flux V_{reg}) are determined in the outer optimization. For each selection of these control variables, the inner deFBA constrains the fluxes to maximize the biomass integral, subject to the repression of the manipulated flux at time t_{reg}. Recall that using this defined deFBA-based bilevel problem, one can directly represent a manipulation of metabolic pathways V_z and V_m, or manipulating gene expression pathways through adapting the biomass production fluxes V_p.

4.3 Closed-loop control based on MPC

After defining the open-loop problem for the offline control, in the next step we consider online control of the fed-batch bioreactor based on a shrinking-horizon MPC in order to account for plant-model mismatches. The proposed control approach is composed

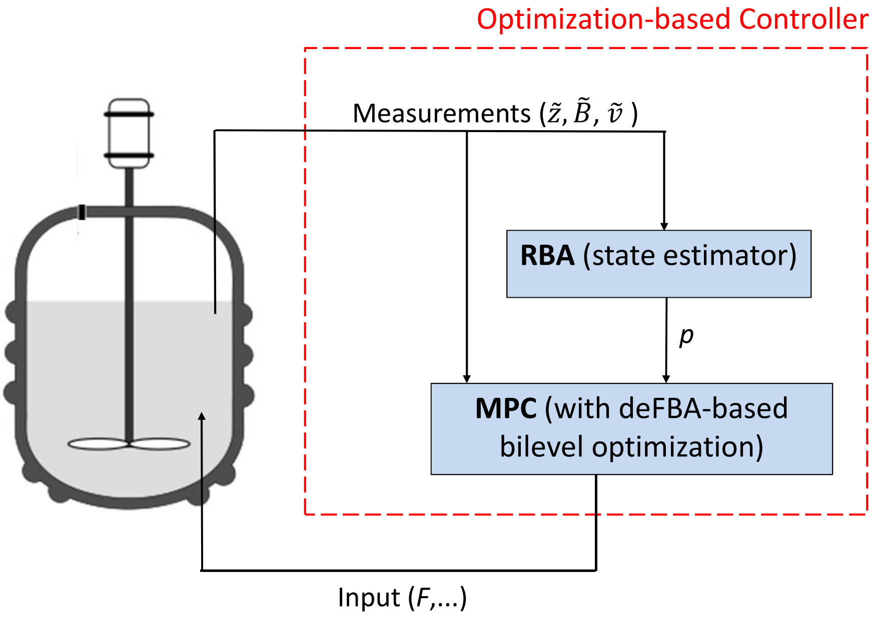

Optimization-based Controller
Measurements ($\tilde{z}, \tilde{B}, \tilde{v}$)
RBA (state estimator)
p
MPC (with deFBA-based bilevel optimization)
Input (F,...)

Figure 4.2: MPC control scheme.

of MPC and RBA as an estimator for estimating the state of biomass components p. Measured and estimated states are then used to initialize the deFBA model employed within MPC.

The control scheme is visualized in Figure 4.2 and explained in detail in following sections.

4.3.1 Model predictive control

Model predictive control is an optimization-based control approach which uses a model to predicts the future behaviour of a dynamic system. MPC computes a trajectory of control inputs by solving an optimization problem at each sampling time [99]. Here, we consider a closed-loop control of the fed-batch bioreactor by introducing feedback to the open-loop problem (4.2), based on the MPC. The optimal trajectory of the control inputs is determined by repeating the bilevel optimization (4.2) while updating the state (x, and v) based on online measurements.

For the MPC optimization, we first fix the batch time t_f equal to its open-loop optimal value, because online adjustment of this variable is considered impractical during the production process. Then as the batch runs for a finite time, a shrinking horizon is defined as $t_f - T$ for the time the optimization is performed over; where T is the current time. A step size h is defined to discretize the prediction horizon to N

time steps:

$$h = \frac{t_f - t_0}{N}, \qquad T = t_0 + ih, \quad i = 0, ..., N.$$

Therefore, the MPC problem has a general form of

$$\underset{\{u_k(t)\}_T^{t_f}}{\text{maximize}} \quad \mathcal{F}(\{x_k(t)\}_T^{t_f}, \{u_k(t)\}_T^{t_f}), \tag{4.3}$$

where $\{u_k(t)\}_T^{t_f}$ and $\{x_k(t)\}_T^{t_f}$ stand for the sequences of control input (here F and t_{reg}) and state variables (here z, p and v) vector, respectively, from the current time T to the final time t_f at the k-th iteration of MPC. $\mathcal{F}$ 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:

$$\begin{aligned}
&\underset{\{F_k(t)\}_T^{t_f}, t_{reg,k}}{\text{maximize}} && x_t(t_f) \\
&\text{s.t.} && T \leq t_{reg,k} \leq t_f, \quad v(t_f) \leq v_{max}, \\
& && \underset{V(\cdot)}{\text{maximize}} \int_T^{t_f} v_k(t) b^T p_k(t) dt \\
& && \text{s.t.} \quad (4.1\text{b}) \quad \text{to} \quad (4.1\text{h}), \\
& && \qquad x_k(T) = \tilde{x}_k(T) = (\tilde{z}_k(T), \tilde{p}_k(T), \quad v_k(T) = \tilde{v}_k(T), \quad x_k(t) \geq 0, \\
& && \qquad V_{reg}(t) = 0 \quad \text{for} \quad t \geq t_{reg,k}.
\end{aligned} \tag{4.4}$$

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 defined nonlinear MPC problem (4.4), the feedback is applied by resetting the controller initial conditions using measured or estimated state variables $\tilde{x}$ (concentrations of substrate and products $\tilde{z}$ and biomass components $\tilde{p}$) and $\tilde{v}$ at each iteration.

4.3.2 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 (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.

We compute the optimal biomass composition by the resource balance analysis (RBA) and use that as an estimate. As explained earlier in Section 2.3.3, 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 μ. 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.

For our application here, the estimation algorithm is defined as:

$$
\begin{aligned}
\underset{\mu,V,p}{\text{maximize}} \quad & \mu \\
\text{subject to} \quad & S_p V_p - \mu v p = 0, \\
& b^T p = \tilde{B}, \\
& S_m V = 0, \\
& \sum_{j \in cat(i)} |V_j / k_{cat,j}| \leq v p_i, \qquad i \in \mathbb{E} \\
& V_{min}(\tilde{z}) \leq V \leq V_{max}(\tilde{z}), \\
& \varphi_Q b^T p \leq p_Q, \qquad Q \in \mathbb{Q}, \\
& V^{uptake} \leq V_{max}^{uptake} \frac{\tilde{z}_i}{K_m + \tilde{z}_i + \tilde{z}_i^2 / K_I}, \\
& V_{reg} = 0, \quad \text{if repressed.}
\end{aligned}
\tag{4.5}
$$

For the measured value of the biomass $\tilde{B}$ at each sampling time with the corresponding volume v, and the state of the manipulated flux V_{reg} (active or repressed), this problem gives an estimate for biomass components, p which is used to initialize the controller states in each iteration. Moreover, measurements on extracellular species ($\tilde{z}$) are used to constrain fluxes. Besides being used in constraining the substrate uptake flux V^{uptake} in the growth inhibition constraint, measurements on $\tilde{z}$ allow to adjust V_{min} and V_{max}; as an example if a metabolite is not present, the corresponding upper flux bound is set to zero.

4.3.3 Linearized MPC

To reduce the computation cost associated with the nonlinear MPC problem in (4.4), we seek a procedure to transform it into a linear problem. Toward this end, we implement a suitable transformation of problem variables followed by a linear approximation of the nonlinear constraint for the substrate uptake.

For external species and biomass components $x = (z, p)$, first we define new variables as $\acute{x}(t) = v(t)x(t)$, which results in a linear objective and dynamics of the model. As the next step, we consider a linear approximation of the growth inhibition constraint (4.1h). To do so, we first replace the general term of V_{max}^{uptake} inside this constraint by $k_{cat}\acute{p}(t)$ for the substrate uptake with maximum enzyme capacity, and then implement the tangent plane approximation to linearize this constraint.

Basically, if we have a function $\mathbb{U} = f(\mathbb{X}, \mathbb{Y})$, $f : \mathbb{R}^2 \to \mathbb{R}$ and if f is differentiable at point (a, b), then the graph of the linearization $L((\mathbb{X}, \mathbb{Y}); (a, b))$ is the tangent plane to the graph of $\mathbb{U} = f(\mathbb{X}, \mathbb{Y})$ at the point $(a, b, f(a, b))$ in $\mathbb{R}^3$, such that:

$$L(\mathbb{X}, \mathbb{Y}) = f(a, b) + \frac{\partial f}{\partial \mathbb{X}}(a, b)(\mathbb{X} - a) + \frac{\partial f}{\partial \mathbb{Y}}(a, b)(\mathbb{Y} - b). \tag{4.6}$$

Here, we follow the same procedure to approximate the nonlinear function in the constraint (4.1h)

$$f(\acute{p}, \acute{z}) = k_{cat}\acute{p}\frac{\acute{z}/v}{K_m + \acute{z}/v + (\acute{z}/v)^2/K_I},$$

by linearizing that at the point $(\acute{p}(T), \acute{z}(T))$

$$L(\acute{p}(T), \acute{z}(T) = f(\acute{p}(T), \acute{z}(T)) + f_{\acute{p}}(\acute{p}(T), \acute{z}(T))(\acute{p} - \acute{p}(T)) + f_{\acute{z}}(\acute{p}(T), \acute{z}(T))(\acute{z} - \acute{z}(T)).$$

This results the inequality constraint (4.1h) as

$$V^{uptake}(t) \leq L(\acute{p}(T), \acute{z}(T)). \tag{4.7}$$

The supporting idea for the linear approximation of this constraint is that as in MPC the states are reset at each control time step, the constraint can also be approximated at the start point of each MPC iteration (time T). In this way, the approximated constraint is satisfied only within one control time step and then will be updated for the next time step. These consequent linear approximations of (4.1h) possible in the closed-loop optimization, allow us to have a reasonable approximation for the nonlinear constraint and to work with an overall linear model for the MPC.

4.3.4 Overall MPC algorithm

The overall control algorithm is presented as Algorithm 4.1 (according to Figure 4.2). It should be noted that before the iterations of the MPC algorithm, the RBA algorithm is used to estimate the composition of the initial biomass, $p(0)$.

4.3.5 Numerical solution

The solve bilevel optimization problems for the offline (4.2) and online (4.4) control, the inner deFBA model is treated by the collocation approach (as described in Section 2.3.2), considering $N = 10$ time intervals and $K = 2$ collocation points within each interval for discretization of variables (V, $\dot{x}$ and x). The decision variable F determined from the outer optimization is considered to change at the boundaries of the N time intervals, as $(F_1, F_2, ..., F_N)$.

For the MPC, the controller sampling time is considered to be equal to the time

Algorithm 4.1: Overall control algorithm

```
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 $\tilde{v}_k(T)$,
   a. linearize the nonlinear constraint (4.1h) as equation (4.7).
   b. solve the dynamic optimization problem (4.4) over $[T,\ t_f]$, which yields the
      optimal control sequence $(F_k, t_{reg,k})$.
2. a. Apply the first move of $(F_k)_T^{T+h}$ to the plant during $[T,\ T+h]$.
   b. Apply the flux regulation ($V_{reg} = 0$) if $T \leq t_{reg,k} \leq T+h$.
3. $T = T + h$
4. Obtain new measurements: $\tilde{z}_{k+1}(T)$, the total biomass $\tilde{B}_{k+1}(T)$ and $\tilde{v}_{k+1}(T)$.
5. Solve the RBA problem (4.5) with measured total biomass $\tilde{B}_{k+1}(T)$,
   which gives estimated biomass components $p_{k+1}(T)$.
6. $k = k + 1$, iterate.
```

intervals for variables discretization ($\frac{t_f - t_0}{N}$). As the state variable $x(t)$ is discretized at the boundaries of the N time intervals, it can be reset at each sampling interval based on the measurements.

The overall bilevel optimization problems for the open-loop control and the MPC are then solved in MATLAB. The outer optimization is performed using *fmincon.* The inner deFBA is solved using *linprog* for the LP optimizations of the MPC, and using IPOPT for the NLP of the open-loop optimization, for each iteration of the outer problem's solution.

For the open-loop optimization, several initial conditions have been evaluated for the outer optimization to avoid obtaining a local minimum of problem's variables. Also, for an efficient performance within the MPC, the optimal control inputs obtained from the 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.

4.4 Application: fed-batch growth of *E. coli* for improved ethanol productivity

As an implementation, we consider a fed-batch growth of *E. coli* for converting glucose to ethanol with maximum productivity. For that, we use the small-scale metabolic-genetic network model of *E. coli* derived in previous chapter (Section 3.5.1) with

Table 4.1: Initial conditions and model parameters.

Parameter	value
$v(0)$	0.7 l
GLC(0)	20 mM
GLC_f	50 mM
$b^T p(0)$	0.005 gl^{-1}
K_m, K_I	0.015, 10 mM
v_{max}	1.5 l

glucose as the only available substrate which is considered to have growth inhibitory effects at high concentrations.

For enhanced production of ethanol, we include a static knockout of the acetate pathway as the competing co-product (by deletion of the corresponding gene *ackA*), along with an optimal aerobic-anaerobic switch. Through that, a high growth rate is achieved by maintaining aerobic conditions (growth stage) while at an optimal switch time, the oxygen supply to the culture is removed and the culture then grows anaerobically to produce ethanol.

Model parameters and initial values are listed in Table 4.1, including initial liquid volume $v(0)$, glucose GLC(0), and biomass ($b^T p(0)$) concentrations, feed glucose concentration GLC_f, maximum allowable liquid volume v_{max}, and saturation and inhibition constants (K_m, K_I) for glucose uptake.

4.5 Results and discussion

To have maximal ethanol productivity through the *E. coli* model, the open-loop optimization problem (4.2) is firstly applied to obtain the optimal batch time, aerobic-anaerobic switch time, and the feeding pattern. The obtained optimal batch time (12.3 hr) is then used as t_f within the closed-loop optimization problem (4.4), while the regulation time and the feeding rate are adjusted online to compensate for plant-model mismatches.

To check the accuracy of implementing the linearized MPC problem, we compare the results obtained by the open-loop optimization to those from the closed-loop optimization without considering any modelling error or mismatches. The predictions by these two implementations are quite similar. Both implementations yield almost a same switch time (6.1 hr). As shown in Figure 4.3, there is a small difference in the feeding pattern. The glucose, ethanol and biomass concentration profiles (Figure 4.4) are slightly different, but both optimizations result in almost the same final values of species (including 45.8 mM ethanol). Based on the results we can argue applying the linearized model for the closed-loop controller as an approximation of the nonlinear

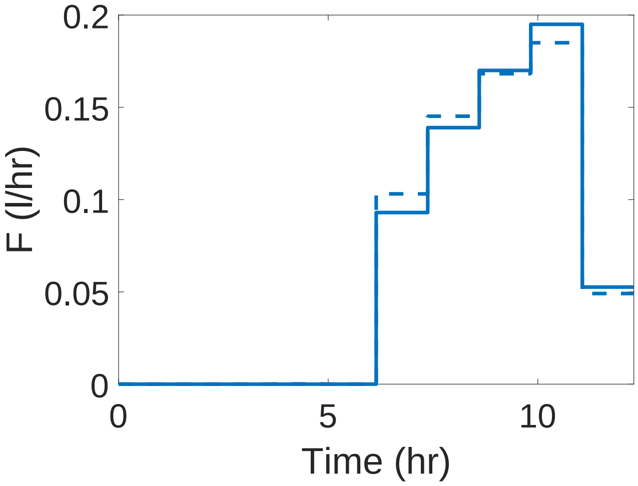


Figure 4.3: Feeding rate from the closed-loop (solid line) and open-loop (dashed line) control considering no parametric errors.

model works well.

4.5.1 MPC performance with parametric errors

To evaluate the performance of the closed-loop controller over the open-loop one, we consider some modelling errors. To do so, we consider plant-model mismatch in terms of two model parameters, the catalytic constant k_{cat} of the enzyme 'pts', responsible for glucose uptake (reaction 1) and glucose feed concentration (GLC_f). For the controller we use the nominal values of these two parameters as in Tables 3.2, and 4.1, while for the plant simulation perturbations are added.

First, we consider -20% errors in the k_{cat} of glucose uptake enzyme in the plant model. The MPC is then implemented with the final batch time of 12.3 hr. Figures 4.5 and 4.6 demonstrate the performance of the closed-loop control over the open-loop one. In order to compensate for decreased glucose uptake rate due to the reduced enzyme catalytic constant, MPC adjusts the feeding pattern and the regulation time for maximal ethanol production. The closed-loop controller predicts a longer aerobic phase (7.4 hr) compared to the nominal case (6.1 hr) to have sufficient amount of biomass produced (0.96 g/L) for higher rate of the process. The final ethanol concentration is obtained as 43 mM (64.5 mmol). This concentration value of ethanol is reasonably consistent with the nominal result of 45.8 mM (68.7 mmol).

On the other hand, the open-loop controller shows a relatively poor performance with the perturbed parameters, compared to the closed-loop one. As there is a reduced substrate uptake rate (due to the reduced catalytic efficiency of the corresponding enzyme), only 0.7 g/l biomass is produced during the batch. Therefore, a large amount of substrate remains unconsumed and as the result only 32.3 mM (48.5 mmol) of ethanol is produced.

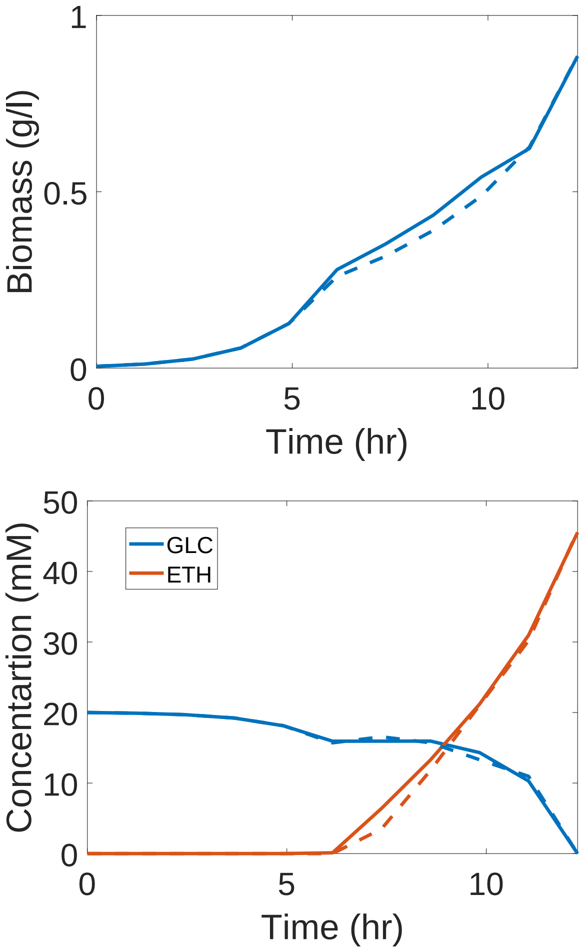


Figure 4.4: Biomass, glucose and ethanol concentration profiles resulted from the closed-loop (solid line) and open-loop (dashed line) control considering no parametric errors.

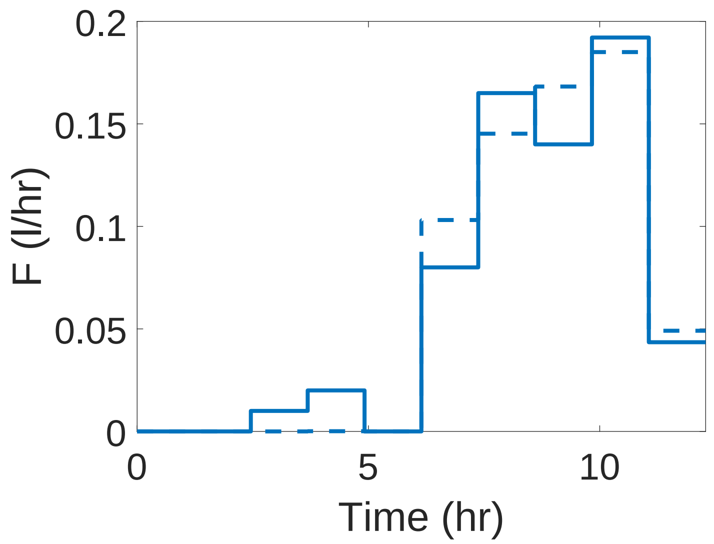


Figure 4.5: Feeding rate from closed-loop (solid line) and open-loop (dashed line) control for -20% error in the k_{cat} of glucose uptake enzyme.

For further investigation of the closed-loop controller performance, we consider +20% change in glucose feed concentration in the plant simulation. Figure 4.7 represents the trajectories for the control input from the open-loop control and MPC. A comparison of species concentrations resulted from MPC and the open-loop control is also shown in Figure 4.8.

With the increased amount of glucose in the feed, the online controller adjusts the substrate feeding and the aerobic-anaerobic switch time to appropriately balance between the biomass growth and ethanol production. Therefore, by an optimal feed input and initial aerobic growth of 6.65 hr predicted by MPC, all the substrate is consumed and 52 mM ethanol is produced. As shown in Figure 4.8, this closed-loop value for final ethanol (52 mM) is significantly higher than the case of implementing the open-loop optimal profile, where there is an incomplete consumption of substrate and only 43 mM ethanol is produced.

As mentioned earlier, in this work $N = 10$ iterations are considered for the MPC optimizations. For this case study, an average 59 s is required for each MPC iteration. However, if one consider the implementation of the original nonlinear deFBA model, the required time increases significantly (an average 4390 s for each iteration). In fact, using a linearized model within the MPC saves the computational time vastly but of course at the expense of linear approximation errors.

4.6 Conclusion

In this chapter, the deFBA-based bilevel optimization problem has been implemented for fed-batch bioreactor control. Through that, fed-batch operating policies including

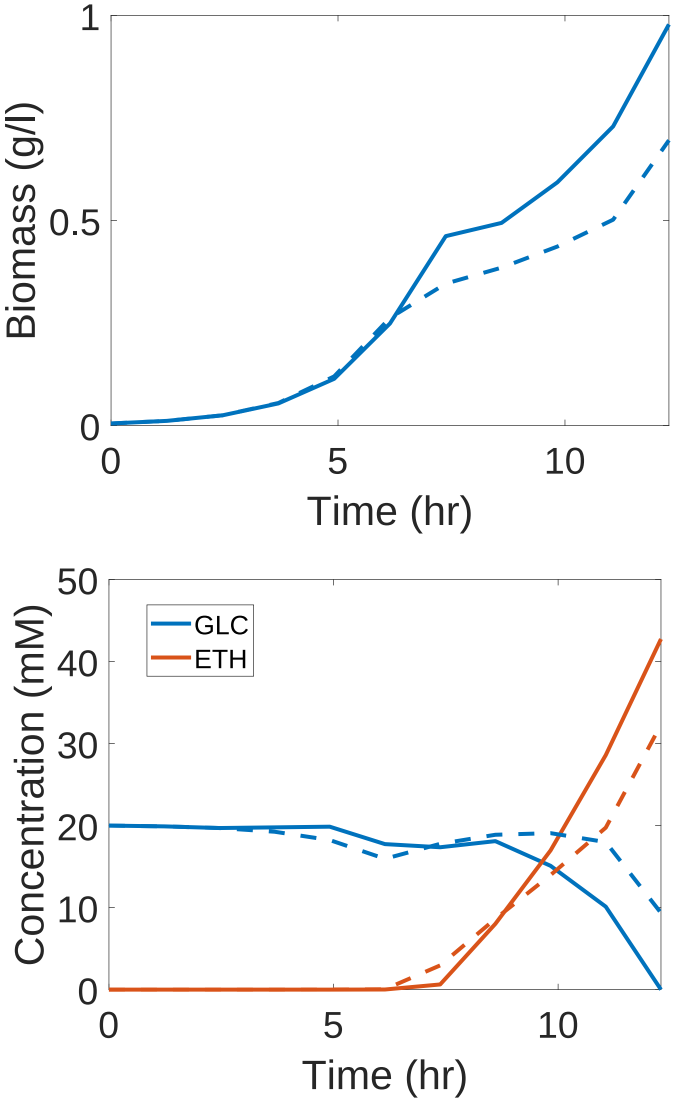


Figure 4.6: Biomass, glucose and ethanol concentration profiles resulted from closed-loop (solid line) and open-loop (dashed line) control for -20% error in the k_{cat} of glucose uptake enzyme.

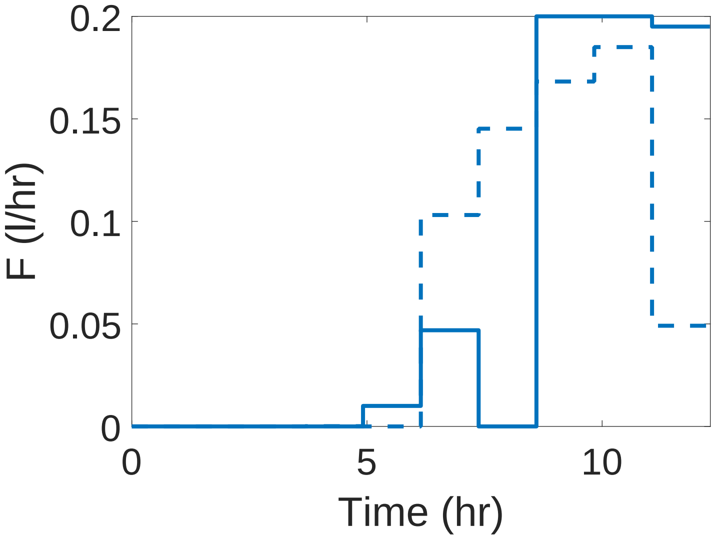


Figure 4.7: Feeding rate from closed-loop (solid line) and open-loop (dashed line) control for +20% error in glucose feed concentration, GLC_f.

the substrate feeding profile and the optimal time for a process-level regulation of metabolism were determined such that the productivity of a target product is maximized. Then, a closed-loop implementation of the open-loop strategies has been considered based on model predictive control in order to handle uncertainties in modelling parameters.

For estimating the state of biomass components required for updating the initial condition of the deFBA model at each iteration of the MPC, the control algorithm includes the RBA as an state estimator. Further through the next chapter, relevant evaluations show that this estimation algorithm can be considered as a suitable approach for estimating the state of biomass components.

Via the example of a fed-batch fermentation of *E. coli* for maximal ethanol productivity, the advantages of feedback corrections to compensate for modelling errors are evaluated by comparing the performance of the closed-loop control over an open-loop one. Considering parametric errors, it was shown that the substrate feeding and metabolic regulation strategy determined through the closed-loop control allow for better consumption of substrate and higher production of the target product within the fed-batch operating mode.

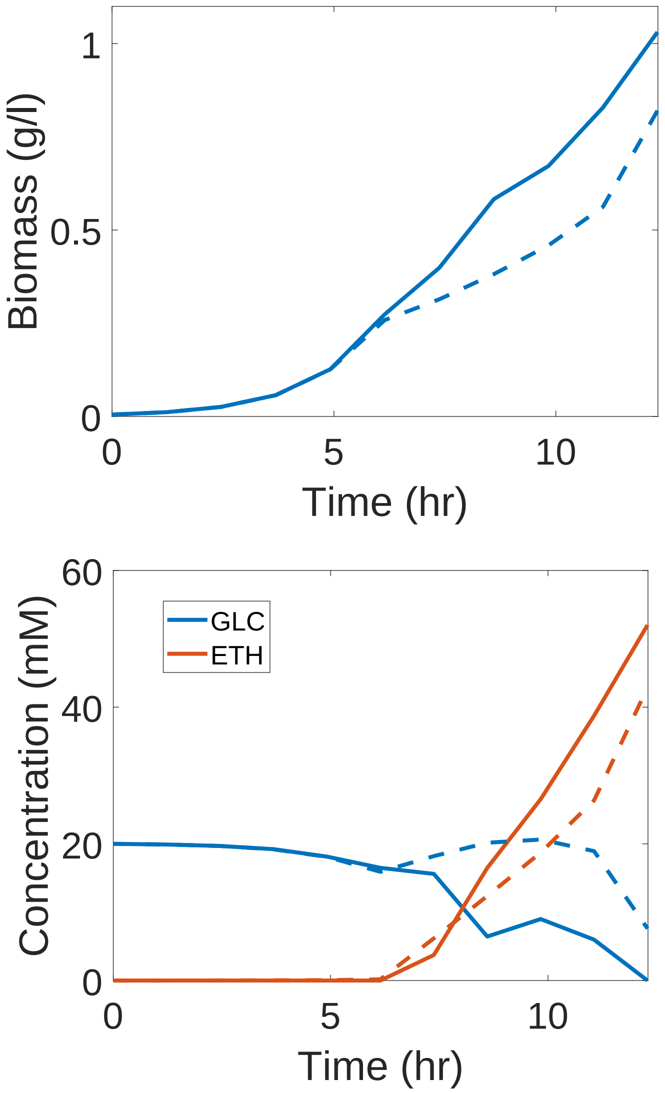


Figure 4.8: Biomass, glucose and ethanol concentration profiles resulted from closed-loop (solid line) and open-loop (dashed line) control for +20% error in glucose feed concentration, GLC_f.

5 Adaptive predictive control of the bioprocess using constraint-based modeling and estimation

This chapter presents an approach for adaptive control of the bioprocess. An adaptive control system allows to compensate for changes in the system dynamics by adjusting the controller characteristics. Based on that, it includes units for measuring (or estimating) the process dynamics and adjusting the controller characteristics accordingly.

Here, an adaptive MPC approach based on the developed deFBA-based bilevel optimization is proposed which allows to adjust the model for each transient mode, using the data obtained during the process. This means it considers online adjustments of not only states but also modeling parameters which are uncertain or time-variable.

This approach is applied to an oxygen-limited (microaerobic) batch growth of *E. coli* for ethanol production. Depending on the degree of oxygen limitation in the microaerobic growth, the cells are transferred into different metabolic modes which lead to changes in dynamics of the system. Through a simulation-based study, we demonstrate that the proposed MPC approach with the adaptive deFBA model can be an efficient approach to control such a process.

5.1 Adaptive control of the bioprocess

In the previous chapter, it was shown that introducing measurement feedbacks into the open-loop deFBA-based bilevel optimization allows for improved performance of the bioreactor control. Repeating the bilevel optimization with updated states based on online measurements resulted in improved 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, here we consider a closed-loop implementation of the deFBA-based control based on adaptive MPC.

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 [100]. Due to the very dynamic nature of biological systems, adaptive MPC can be a suitable approach for a flexible bioprocess control.

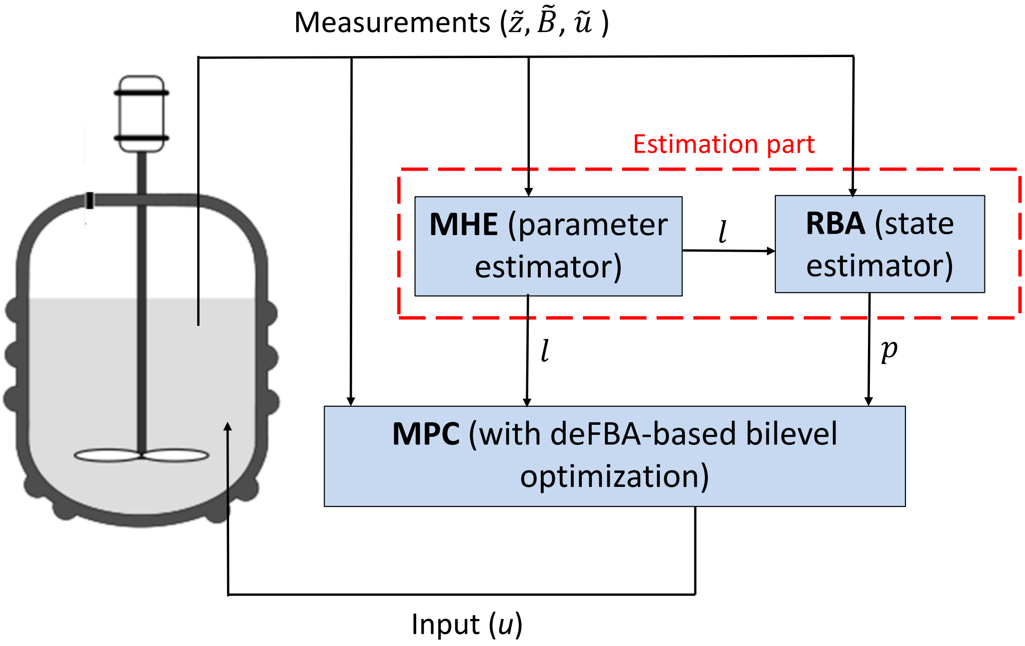

Measurements ($\tilde{z}, \tilde{B}, \tilde{u}$)
Estimation part
MHE (parameter estimator)
l
RBA (state estimator)
l
p
MPC (with deFBA-based bilevel optimization)
Input (u)

Figure 5.1: Adaptive predictive control scheme.

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. To do so, MPC is combined with a moving horizon estimation (MHE) algorithm to estimate uncertain modelling parameters for different growth modes of the organism online. This adaptive approach allows to adapt the model for each transient mode, using the data obtained during the process.

5.2 MPC based on adapted metabolic-genetic model

The proposed control approach is composed of MPC and an estimation part including moving horizon estimation (MHE) and resource balance analysis (RBA) algorithms. 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 5.1 and explained in detail in the following sections.

5.2.1 Adaptive model predictive control

Before starting with the adaptive MPC, let's recall the bilevel optimization for maximal productivity in a batch growth developed in Chapter 3, which is used here as the open-loop control problem:

$$\begin{aligned}
&\underset{t_f,u(t)}{\text{maximize}} && x_t(t_f)/t_f && \text{(5.1a)}\\
&\text{s.t.} && \underset{V(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} b^T p(t)dt && \text{(5.1b)}\\
& && \text{s.t.} \quad \dot{x}(t) = \begin{pmatrix}\dot{z}(t)\\ \dot{p}(t)\end{pmatrix} = SV(t), && \text{(5.1c)}\\
& && S_m V(t) = 0, && \text{(5.1d)}\\
& && \sum_{j\in cat(i)} |V_j(t)/k_{cat,j}| \le p_i(t), \qquad i \in \mathbb{E} && \text{(5.1e)}\\
& && V_{min}(t) \le V(t) \le V_{max}(t), && \text{(5.1f)}\\
& && \varphi_Q b^T p(t) \le p_Q(t), \qquad Q \in \mathbb{Q} && \text{(5.1g)}\\
& && x(t_0) = x_0 = (z_0, p_0), && \text{(5.1h)}\\
& && x(t) \ge 0, && \text{(5.1i)}\\
& && V_{reg}(t) = u(t). && \text{(5.1j)}
\end{aligned}$$

For this part, instead of finding the optimal time for repression of a target flux (which so far has been considered for the regulation control input), we aim to find the profile of the target flux which is dynamically manipulated. It means this formulation does not only address the flux repression, but also allows for temporal down-regulation or up-regulation of the target flux. Therefore, in the outer optimization the optimal profile of the manipulated flux u and the optimal batch time t_f are determined to get the maximum productivity of the target metabolite $x_t(t_f)/t_f$. In the inner problem the unregulated metabolic fluxes are optimally distributed for biomass production, while the manipulated flux V_{reg} is set to the control profile u determined from the outer optimization.

For the MPC, same as what has been done in the previous chapter, we fix the batch time equal to the optimal batch time t_f obtained from the open-loop optimization (5.1), and then perform MPC optimizations over a shrinking horizon from the current time T to the end of the batch t_f.

For $k = 1$ to N, each individual optimization problem over the shrinking batch time $[T, t_f]$ is defined as:

$$
\begin{aligned}
\underset{\{u_k(t)\}_T^{t_f}}{\text{maximize}} \quad & x_t(t_f) \\
\text{s.t.} \quad & \underset{V(\cdot)}{\text{maximize}} \int\limits_T^{t_f} b^T p_k(t)dt \\
& \text{s.t.} \quad (5.1c) \text{ to } (5.1g), \\
& \qquad x_k(T) = \tilde{x}_k(T) = (\tilde{z}_k(T), \tilde{p}_k(T)), \quad x_k(t) \geq 0, \\
& \qquad V_{reg}(t) = u_k(t).
\end{aligned} \tag{5.2}
$$

Based on this problem (5.2), the feedback is applied by resetting the controller initial conditions using measured and estimated state variables $\tilde{x}$ (substrate, products and biomass components concentrations) at each iteration.

Within an adaptive MPC besides updating the state based on the online data, the uncertain parameters are also estimated and updated in each iteration of the MPC. 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.

5.2.2 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 [101, 102]. 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:

$$
\begin{aligned}
\underset{l_k}{\text{minimize}} \quad & \sum_{i=0}^{N_{mhe}} (B_k(T-ih) - \tilde{B}_k(T-ih))^2 + (z_k(T-ih) - \tilde{z}_k(T-ih))^2 \\
\text{s.t.} \quad & \qquad l_k^{min} \leq l_k \leq l_k^{max}, \\
& \underset{V(\cdot)}{\text{maximize}} \int\limits_{T-N_{mhe}h}^{T} b^T p_k(t) dt \\
& \text{s.t.} \quad \text{(5.1c) to (5.1g)}, \\
& \qquad x_k(T-N_{mhe}h) = \tilde{x}_k(T-N_{mhe}h) = (\tilde{z}_k(T-N_{mhe}h), \tilde{p}_k(T-N_{mhe}h)), \\
& \qquad V_{reg}(t) = \tilde{u}_k(t).
\end{aligned}
\tag{5.3}
$$

In the estimation problem (5.3), 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 for the possibility of input disturbances which can be measured we use a more general formulation here. Note that in (5.3), the constraints (5.1c) to (5.1g) 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 $T - ih$, $i = 0, ..., N_{mhe}$, 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.

5.2.3 State estimation by resource balance analysis

As explained in the previous chapter, in practice it is only possible to measure the total biomass concentration, and an algorithm is needed to estimate the state of biomass components for updating the states at each MPC iteration. Here, again we implement RBA to compute the optimal biomass composition and use that as the estimate. For our current application, the estimation algorithm is defined as:

$$
\begin{aligned}
\underset{\mu,V,p}{\text{maximize}} \quad & \mu \\
\text{subject to} \quad & S_p V_p - \mu p = 0, \\
& b^T p = \tilde{B}, \\
& S_m V = 0, \\
& \sum_{j \in cat(i)} |V_j / k_{cat,j}| \leq p_i, \qquad i \in \mathbb{E} \\
& V_{min}(\tilde{z}) \leq V \leq V_{max}(\tilde{z}), \\
& \varphi_Q b^T p \leq p_Q, \qquad Q \in \mathbb{Q} \\
& V_{reg} = \tilde{u}.
\end{aligned}
\tag{5.4}
$$

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. Estimated p is then 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}.

5.2.4 Adaptive MPC algorithm

The overall control algorithm is summarized as Algorithm 5.1 (in accordance to Figure 5.1). 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 $p(0)$ (while still no regulation of the metabolic-genetic network is applied).

5.3 Application: improved ethanol productivity in microaerobic growth of *E. coli*

As the case study, we consider a batch growth of *E. coli* for converting glycerol into ethanol under microaerobic conditions.

Glycerol is an attractive carbon source for fermentation processes as besides its availability and low prices, it allows for higher yields of fermentation products than those obtained from common sugars such as glucose [103]. Although the anaerobic fermentation of glycerol by *E. coli* is an excellent platform for the synthesis of products such as ethanol, the microorganism can not metabolize glycerol fermentatively unless rich nutrients such as tryptone and yeast extract are present.

However, in [104] it was shown that the microaerobic growth condition (growth with limited amount of oxygen) allows for efficient production of ethanol from glycerol in *E. coli* without the need for additional expensive nutrients. This is a specifically interesting growth condition to be implemented in our study, as the structure of the

Algorithm 5.1: Adaptive predictive control algorithm

```
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 (5.2) over $[T,\, t_f]$ which,
yields the optimal control sequence $\{u_k\}_T^{t_f}$.
2. Apply the first move of control input $(u_k)_T^{T+h}$ 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 (5.3) with inputs $(\tilde{u}, \tilde{x})_{T-N_{mhe}h}^{T}$ which
yields the estimated parameter $l_{k+1}$.
6. Solve the state estimation problem (5.4) 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.
```

model changes in correspondence to the degree of oxygen limitation. Within this case study of the microaerobic growth of *E. coli*, the proposed MPC algorithm is applied to regulate the metabolic modes in order to maximize ethanol production by adjusting the oxygen supply to the culture.

5.3.1 Network description

We consider a reduced metabolic-genetic network model of *E. coli* growing on glycerol developed from the *in silico* model of the central *E. coli* metabolism [1, 105]. 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 *E. coli* network with glycerol as the carbon source is similar to what was presented in Chapter 3 (Section 3.5.1).

All reactions with corresponding enzymes and their catalytic constants are given in Tables 5.1 and 5.2. The catalytic constants of enzymes are extracted from the enzyme database BRENDA [61] for *E. coli*, and multiplied by an overall scaling factor f to match model predictions to available experimental data. The initial biomass composition $p(0)$ is computed by the RBA to yield the maximum aerobic growth rate

Table 5.1: 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	*glpK-D*	$1998 \cdot f$
2	GLY + PEP + NAD$^+$ $\rightarrow$ T3P + NADH + PYR	*dhaK*	$1447 \cdot f$
3	2T3P $\rightarrow$ G6P	*fba*	$630 \cdot f$
4	G6P + 2NAD$^+$ $\rightarrow$ RU5P + 2NADH + CO2	*gnd*	$1326 \cdot f$
5	RU5P $\rightarrow$ X5P	*rpe*	$78000 \cdot f$
6	RU5P $\rightarrow$ R5P	*rpi*	$3000 \cdot f$
7	X5P + R5P $\rightarrow$ S7P + T3P	*tkt*	$3402 \cdot f$
8	S7P + T3P $\rightarrow$ E4P + G6P	*tal*	$780 \cdot f$
9	E4P + X5P $\rightarrow$ T3P + G6P	*tkt*	$3402 \cdot f$
10	T3P + ADP + NAD$^+$ $\rightarrow$ PEP +NADH + ATP	*eno*	$3162 \cdot f$
11	PEP + ADP $\rightarrow$ PYR + ATP	*pyk*	$3960 \cdot f$
12	PYR + CoA $\rightarrow$ AcCoA + FOR	*pfl*	$768 \cdot f$
13	PYR + CoA + NAD$^+$ $\rightarrow$ AcCoA + NADH + CO2	*pdh*	$29160 \cdot f$
14	FOR $\rightarrow$ CO2	*fhl*	$169980 \cdot f$
15	PEP + CO2 $\rightarrow$ OAA	*ppc*	$32400 \cdot f$
16	OAA + AcCoA + NAD$^+$ $\rightarrow$ AKG + CoA + NADH + CO2	*acn*	$318 \cdot f$
17	AKG + ADP + 3NAD$^+$ $\rightarrow$ OAA + 3NADH + ATP + CO2	*sucCD*	$2684 \cdot f$
18	AcCoA + ADP $\rightarrow$ ATP + CoA + ACT	*ackA*	$800 \cdot f$
19	AcCoA + 2NADH $\rightarrow$ ETH + CoA + 2NAD$^+$	*adhE*	$942 \cdot f$
20	ATP $\rightarrow$ ADP		
21	AKG + ATP + NADH $\rightarrow$ AA + ADP + NAD$^+$	*gln*	$360 \cdot f$
22	O2 + 2qH2 $\rightarrow$ 2q + 8H$^+$	*cyo*	$18000 \cdot f$
23	NADH + q $\rightarrow$ qH2 + 4H$^+$ + NAD$^+$	*nuo*	$2220 \cdot f$
24	4H$^+$ + ADP $\rightarrow$ ATP	*atpH*	$3300 \cdot f$
25	ACT + ATP + CoA $\rightarrow$ ADP + AcCoA	*acs*	$3996 \cdot f$

on glycerol. The complete deFBA model is then specified by the reactions in Tables 5.1 and 5.2 in combination with the biomass composition constraint for the quota compound (as in Chapter 3)

$$0.55b^T p \leq Q. \tag{5.5}$$

Initial nutrient conditions are summarized in Table 5.3.

In the next step, we try to evaluate the prediction accuracy of the considered metabolic-genetic network of *E. coli* for oxygen-limited growth conditions. To this aim, we calibrate the model based on the available experimental data for the growth of *E. coli* on glycerol for different degrees of oxygen limitation during a continuous process mode.

As a parameter for the model adjustment, we consider the already introduced parameter f as a factor which is multiplied to all catalytic constants to scale them. This parameter is then tuned such that the mismatch between model predictions on

Table 5.2: 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 → *glpK-D* + 6628ADP	180610	0.43	0.000044
27	3657AA + 14628ATP → *dhaK* + 14628ADP	398610	0.2	0
28	1696AA + 6784ATP → *fba* + 6784ADP	184860	0.42	0.0000061
29	1275AA + 5100ATP → *gnd* + 5100ADP	138980	0.56	0.0000032
30	225AA + 900ATP → *rpe* + 900ADP	24520	3.2	0.000000008
31	438AA + 1752ATP → *rpi* + 1752ADP	47740	1.64	0.00000091
32	1326AA + 5304ATP → *tkt* + 5304ADP	144530	0.54	0.00000041
33	634AA + 2536ATP → *tal* + 2536ADP	69110	1.14	0.0000018
34	1923AA + 7692ATP → *eno* + 7692ADP	209610	0.37	0.000025
35	1880AA + 7520ATP → *pyk* + 7520ADP	204920	0.38	0.000014
36	1766AA + 7064ATP → *pfl* + 7064ADP	192490	0.41	0
37	42096AA + 168384ATP → *pdh* + 168384ADP	4588460	0.02	0.0000015
38	2837AA + 11348ATP → *fhl* + 11348ADP	309230	0.25	0
39	3532AA + 14128ATP → *ppc* + 14128ADP	384990	0.2	0.0000006
40	1943AA + 7772ATP → *acn* + 7772ADP	211790	0.37	0.000108
41	4565AA + 18260ATP → *sucCD* + 18260ADP	497580	0.16	0.0000082
42	4291AA + 17164ATP → *ackA* + 17164ADP	467720	0.17	0
43	35640AA + 142560ATP → *adhE* + 142560ADP	3884760	0.02	0
44	5675AA + 22700ATP → *gln* + 22700ADP	618580	0.13	0.000034
45	1291AA + 5164ATP → *cyo* + 5164ADP	140720	0.56	0.0000071
46	4282AA + 17128ATP → *nuo* + 17128ADP	466740	0.17	0.000076
47	4895AA + 19580ATP → *atpH* + 19580ADP	533550	0.15	0.000116
48	652AA + 2608ATP → *acs* + 2608ADP	68130	1.1	0
49	7459AA + 9132R5P + 29836ATP → R + 29836ADP	2292420	0.1	0.000015
50	25AA + 27G6P + 10R5P + 36E4P + 12T3P + 100PEP + 283PYR + 274AcCoA + 178OAA + 3110ATP + 1361NADH → Q + 274CoA + 3110ADP + 1361NAD$^+$	65000	28.8	0.0059

cellular rates and the experimentally measured ones is minimized. For that, we use the *E. coli* growth data for four levels of respiration (representing fully aerobic to highly oxygen-limited growth). The scaling factor value is determined by minimizing the squared norm of the deviations between the model outputs and the experimental measurements on glycerol uptake, acetate formation and the overall cellular growth rates.

The obtained values of the scaling factor f for different levels of respiration are presented in Table 5.4. From that it can be noticed that a different scaling factor is obtained for the aerobic and microaerobic modes. The scaling factor for the aerobic growth is higher (0.9) compared to the oxygen-limited growth (approximately 0.6 for all oxygen-limited modes), which in fact suggests lower rate of *E. coli* metabolism in oxygen limitation conditions.

In this direction, Figure 5.2 shows the rates of glycerol uptake and acetate production as well as the overall growth rate predicted by the calibrated deFBA model along with their experimentally measured values for different microaerobic levels.

5.3.2 Simulation cases for adaptive MPC

In this step, we define simulation cases for implementing the adaptive MPC algorithm (Algorithm 5.1), in which the metabolic-genetic model of *E. coli* is used to maximize the ethanol productivity. A metabolic regulation is considered by manipulating the oxygen uptake rate OUR (V_{22} according to the numbering of metabolic pathways in Table 5.1).

We first describe the parameter variation that is used to generate artificial process data. 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 [106]. Previous experimental studies indicate such a pattern for microaerobic *E. coli* growth on glycerol as well, suggesting lower metabolic activity of *E. coli* in oxygen limitation conditions compared to a fully aerobic growth. As shown in the previous section, through validating the metabolic-genetic network with the available experimental data, 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

Table 5.3: Initial nutrient and initial biomass $b^T p(0)$.

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

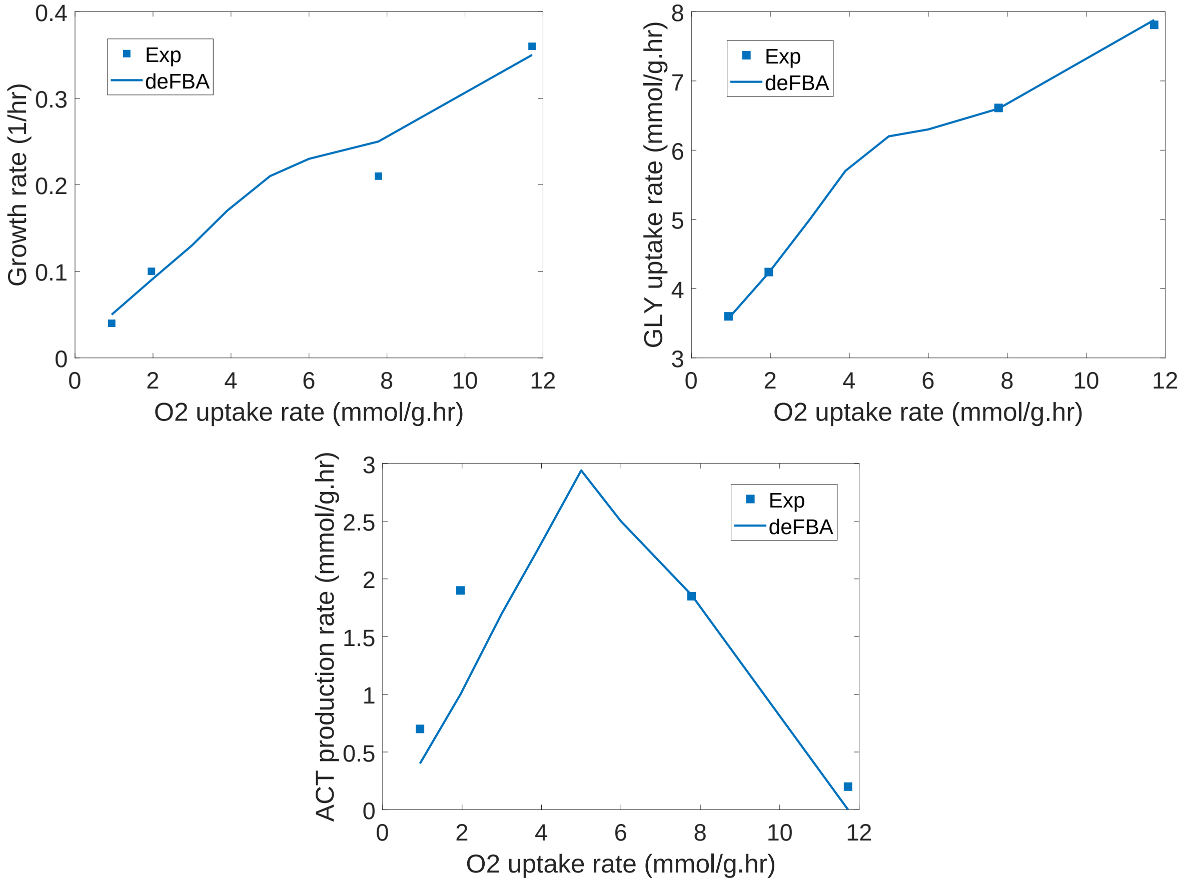


Figure 5.2: Oxygen-limited growth of *E. coli*: rates of cellular growth, glycerol uptake and acetate formation simulated by the deFBA model together with the experimental data.

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 5.5. Details of implementing each case are presented in the following sections.

5.3.2.1 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.

5.3.2.2 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

Table 5.4: Obtained scaling factor values for different oxygen-limited growth modes through the model validation.

Oxygen uptake rate/ mmol(g.biomass)$^{-1}$hr^{-1}	Scaling factor f
11.8 (aerobic)	0.88
7.78	0.58
1.96	0.6
0.94	0.61

Table 5.5: 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

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, 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 (5.3) 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 5.3. 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 adaptation 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 represent the process dynamics at each stage. However, a simultaneous scheme can be also considered for MHE in which a larger set of identified constants are estimated simultaneously through a single problem.

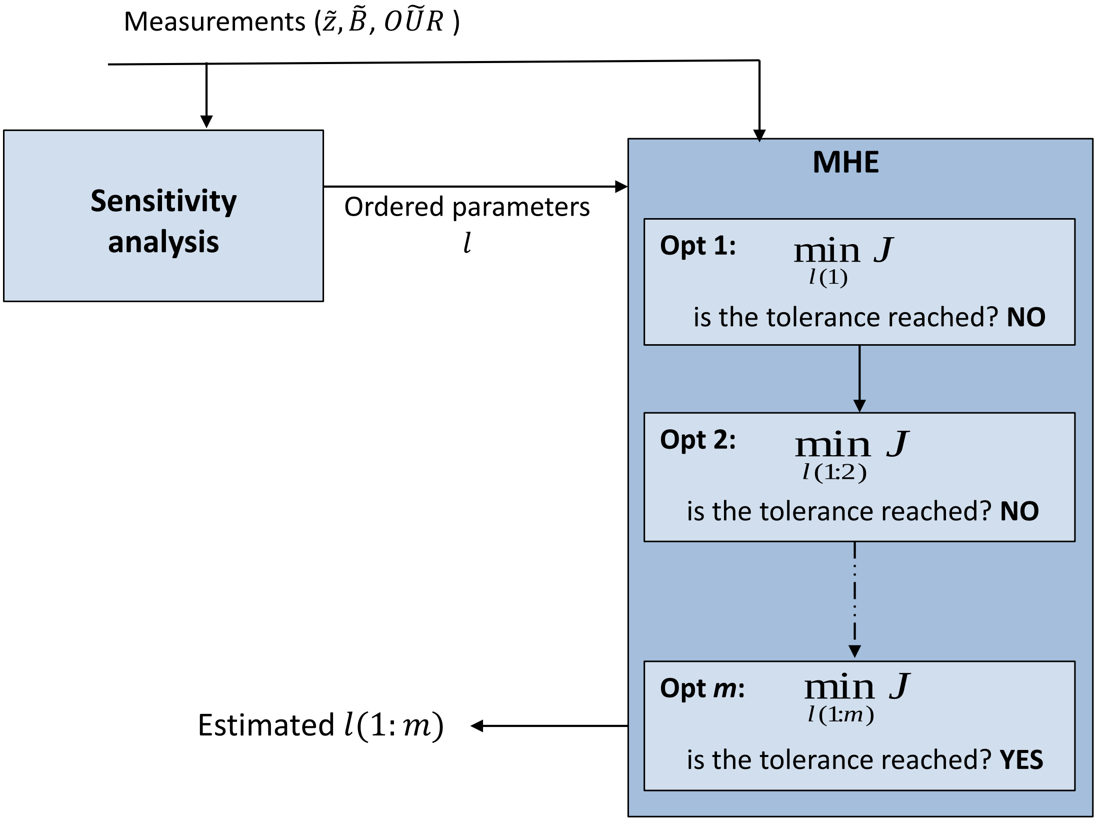

Measurements ($\tilde{z}, \tilde{B}, \widetilde{OUR}$)
Sensitivity analysis
Ordered parameters l
MHE
Opt 1: $\min_{l(1)} J$
is the tolerance reached? NO
Opt 2: $\min_{l(1:2)} J$
is the tolerance reached? NO
Opt m: $\min_{l(1:m)} J$
is the tolerance reached? YES
Estimated $l(1:m)$

Figure 5.3: Rate constants selection and estimation procedure.

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

5.4 Results and discussions

To have maximal ethanol productivity through the introduced *E. coli* model, the open-loop optimization problem (5.1) is firstly applied to obtain the optimal batch time t_f and oxygen uptake rate $OUR(t)$. By implementing the open-loop problem (5.1) on the model discussed in Section 5.3, the final batch time is obtained as 14.9 hr for maximal ethanol productivity. The optimization results are shown in Figure 5.4. 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 (5.2), but the OUR is adjusted online to yield the maximum final ethanol concentration while compensating for the plant-model mismatch. This is implemented on the three cases defined in Section 5.3.2.

5.4.1 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 5.5 to 5.7 show the overall performance of the adaptive MPC (as in Algorithm 5.1) 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 5.5 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

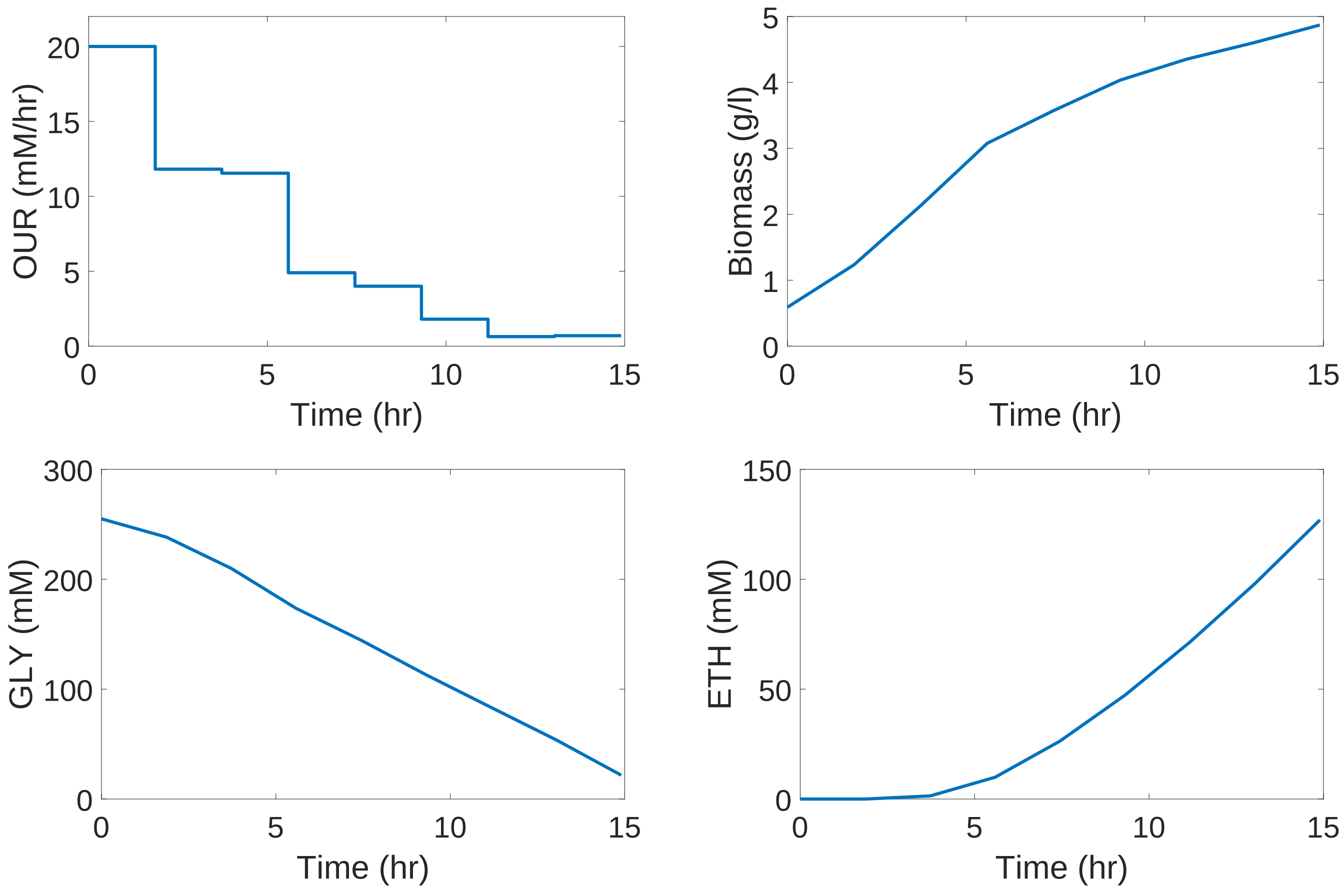


Figure 5.4: Optimal OUR pattern and concentration profiles of biomass, glycerol and ethanol, resulted from the open-loop optimization.

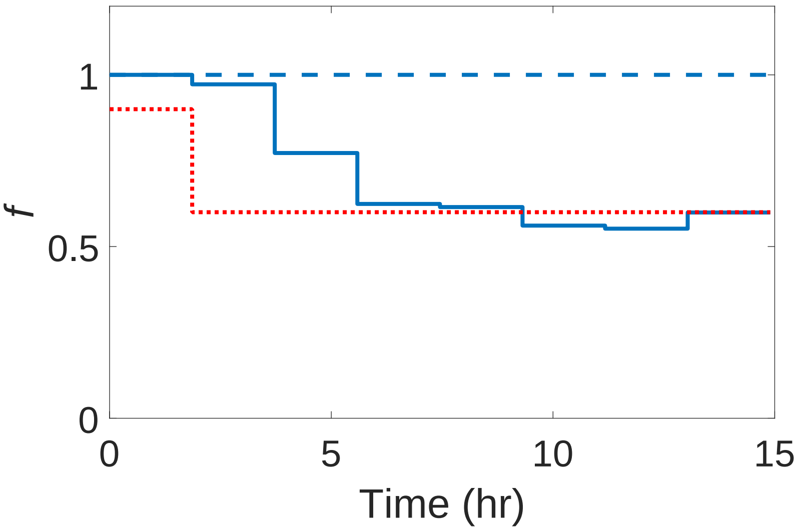


Figure 5.5: Scaling factor values estimated in adaptive MPC/Case 1 (solid line) and used in non-adaptive MPC (dashed line). The real system values are shown in red.

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).

5.4.2 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. 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 5.3.2.2) 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 $k_{cat,new} = k_{cat}/5$ and $k_{cat,new} = 5k_{cat}$, 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

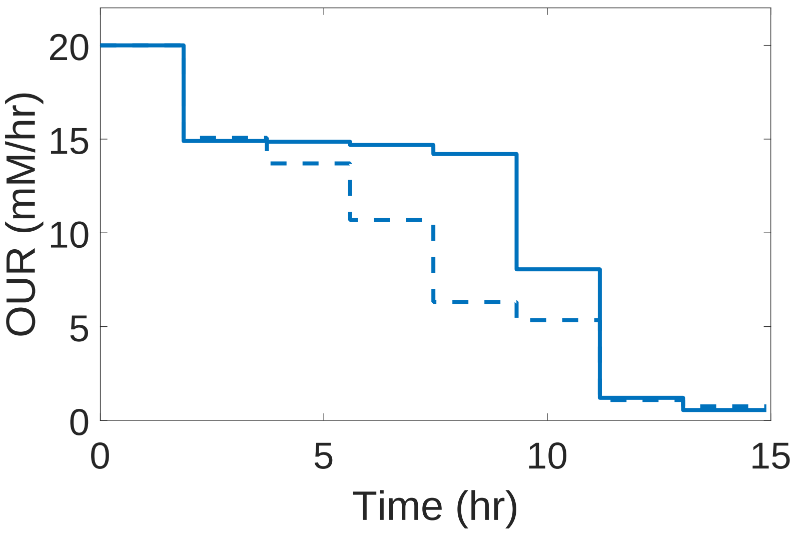


Figure 5.6: *OUR* pattern from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

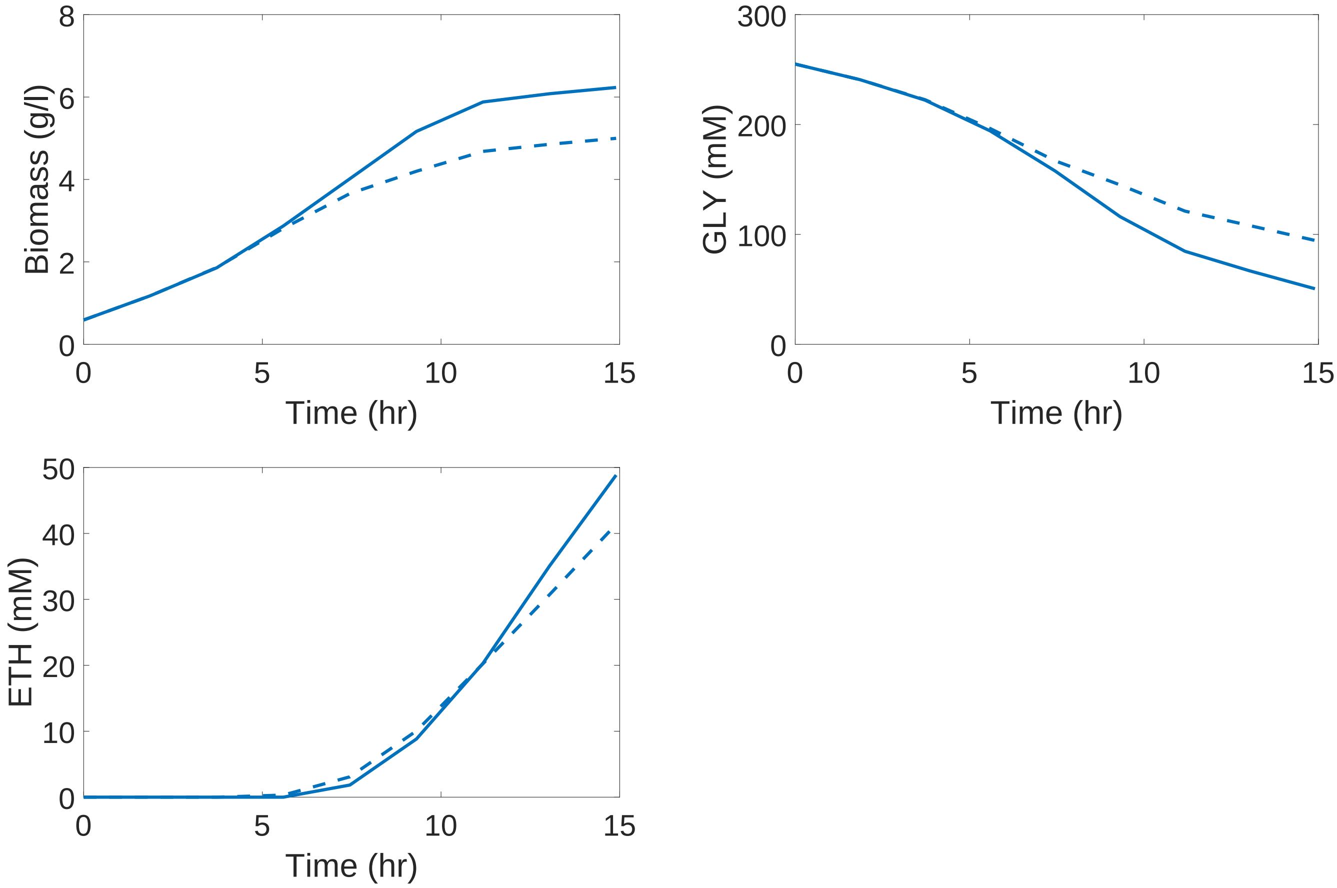


Figure 5.7: Biomass, glycerol and ethanol concentration profiles resulted from adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

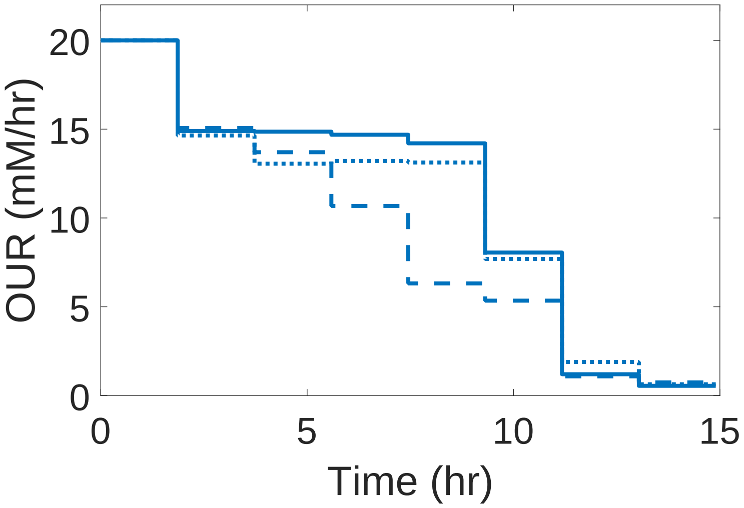


Figure 5.8: OUR pattern from adaptive MPC/Case 2 (dotted line), adaptive MPC/-Case 1 (solid line) and non-adaptive MPC (dashed line).

MPC without model adaptation (in which nominal values of catalytic constants are used), shown in Figures 5.8 and 5.9. As can be seen in Figure 5.8, 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 5.6: 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

Table 5.6 shows the final concentration of ethanol from adaptive MPC (Case 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 5.10 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

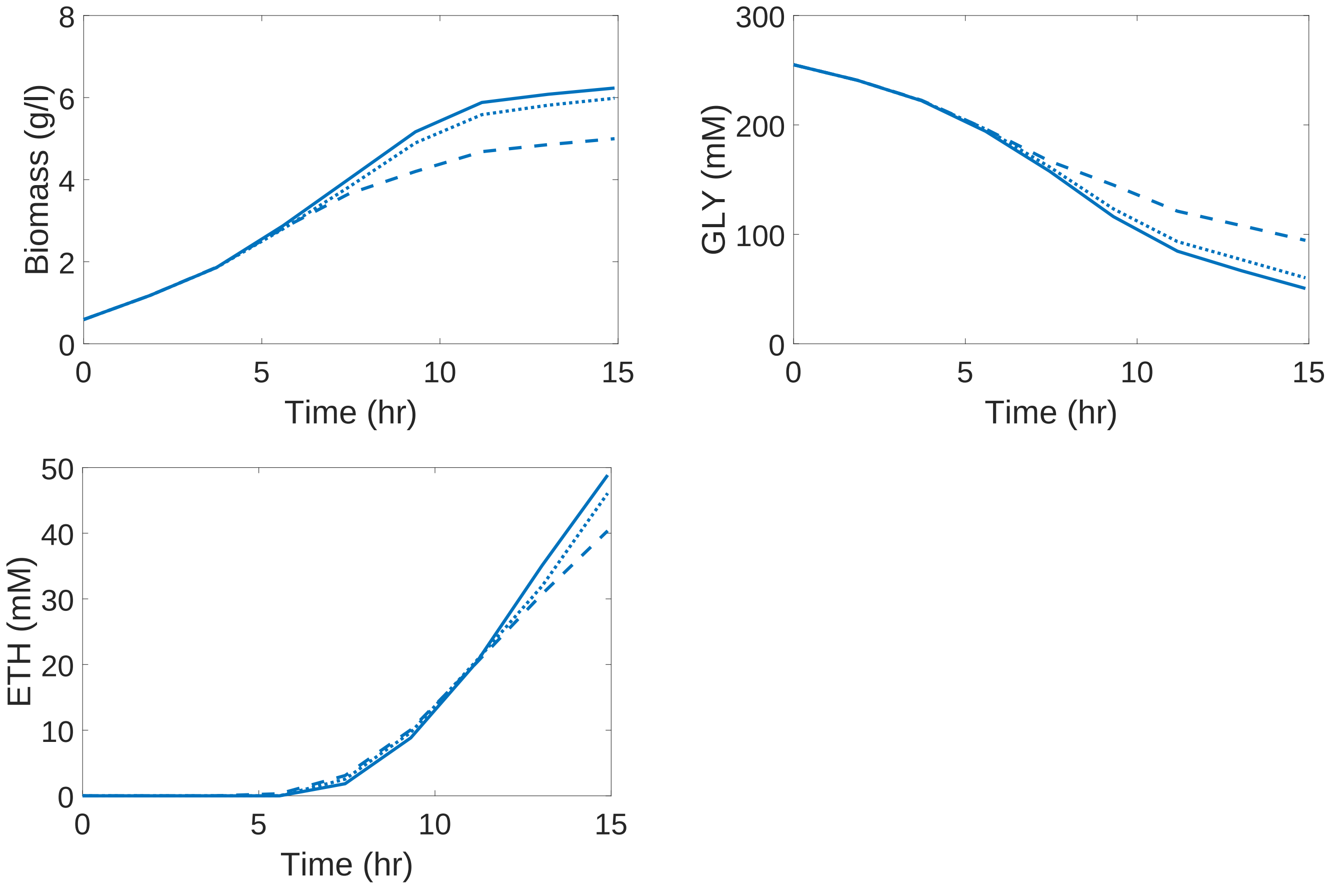


Figure 5.9: Biomass, glycerol and ethanol concentration profiles resulted from adaptive MPC/Case 2 (dotted line), adaptive MPC/Case 1 (solid line) and non-adaptive MPC (dashed line).

Table 5.7: Estimated values of the identified parameters (k_{cat}) at each MPC iteration, with their nominal and plant values

Enzyme with k_{cat} adjustment	*glpK-D*	*dhaK*	*pfl*	*acn*	*ackA*	*adhE*	*gln*	*nuo*	*atpH*
Nominal value	1998	1447	768	318	800	942	360	2220	3300
Aerobic growth									
Plant value	1798	1302	691	286	720	848	324	1998	2970
Estimated (Iteration 1)	-	-	-	-	-	-	249	2156	3199
Microaerobic growth									
Plant value	1199	868	461	191	480	565	216	1332	1980
Estimated (Iteration 2)	1047	-	-	307	-	-	220	1683	1882
Estimated (Iteration 3)	-	888	-	179	483	-	208	1252	-
Estimated (Iteration 4)	-	-	-	-	-	-	-	1232	1601
Estimated (Iteration 5)	-	836	457	-	475	577	-	-	-
Estimated (Iteration 6)	-	846	-	-	-	563	-	-	-
Estimated (Iteration 7)	-	-	-	-	-	549	-	-	-

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). Corresponding to Figure 5.10, Table 5.7 represents the k_{cat} values of identified enzymes estimated in each MPC iteration along with their nominal and plant values. 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 5.11. During 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, [104], regarding the transition between respiratory and oxygen-limited utilization of glycerol). That is why the catalytic constant of enzyme *glpK-D* is the relevant param-

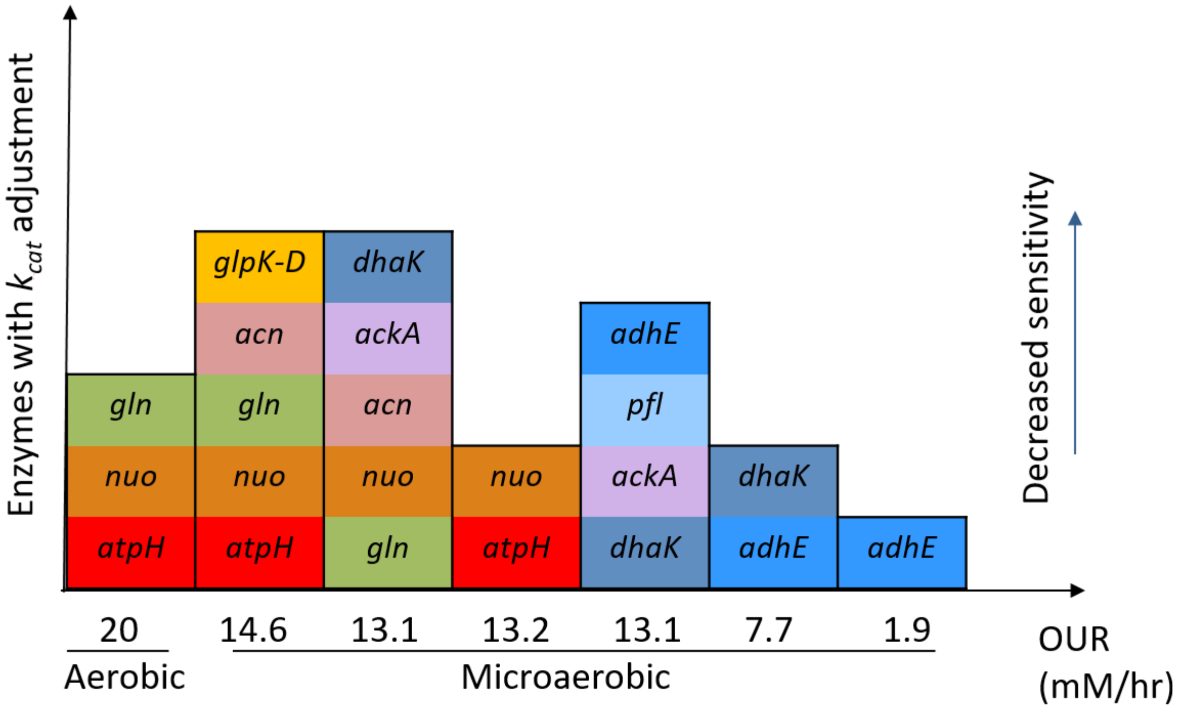


Figure 5.10: Enzymes with estimated constants at different OUR levels of the process corresponding to the MPC iteration ($k = 1, ..., N - 1$): *glpK-D* (reaction 1), *dhaK* (reaction 2), *pfl* (reaction 12), *acn* (reaction 16), *ackA* (reaction 18), *adhE* (reaction 19), *gln* (reaction 21), *nuo* (reaction 23), *atpH* (reaction 24).

eter in the earlier stage of the process while the constant of enzyme *dhaK* is relevant in later stages. 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.

5.4.3 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.

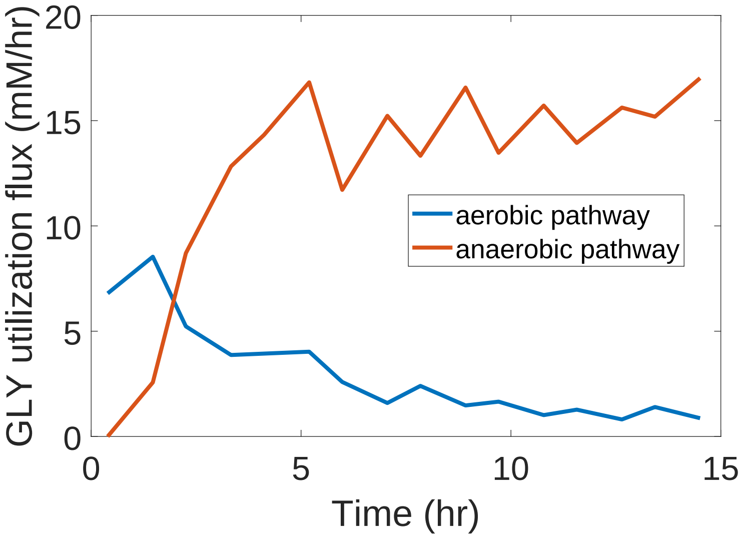


Figure 5.11: Aerobic and anaerobic glycerol utilization fluxes during the considered microaerobic process (resulted from the open-loop optimization).

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 5.3.2.2), 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 5.12 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 5.12 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.

5.4.4 Computational limitations

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.

For MPC with the parameter selection, 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 [107]) 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.

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.

5.4.5 MPC with and without biomass state estimation

Next, the effect of using the RBA method as an approximate steady state estimation within the adaptive MPC is evaluated. In previous sections of this chapter, MPC

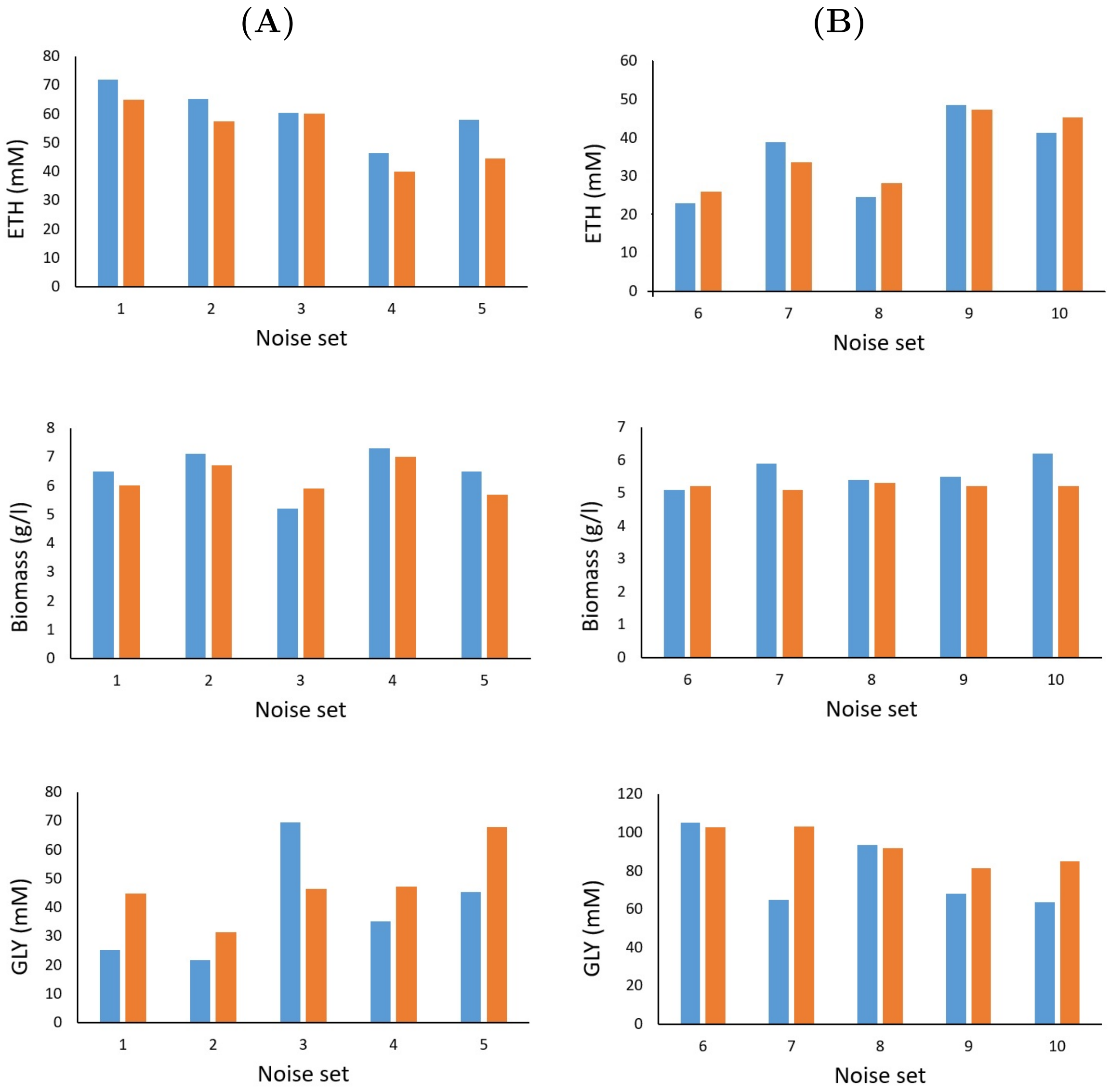


Figure 5.12: Ethanol, biomass and glycerol concentration terminal values resulting from adaptive MPC (blue bars) and non-adaptive MPC (orange bars) for different noise sets applied on k_{cat} values with the adjustment of (A) influential k_{cat} values and (B) the scaling factor f value.

used the RBA algorithm through the problem (5.4) 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 5.13, 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 5.14 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 5.4.2 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.

5.5 Conclusions

In this chapter, we have implemented a combination of model predictive control and moving horizon estimation for an adaptive and flexible control of the bioprocess, based on the developed bilevel optimization framework. In particular, improving bioprocess productivity in microaerobic growth regimes was considered.

By considering several simulation cases, we have demonstrated the usefulness of

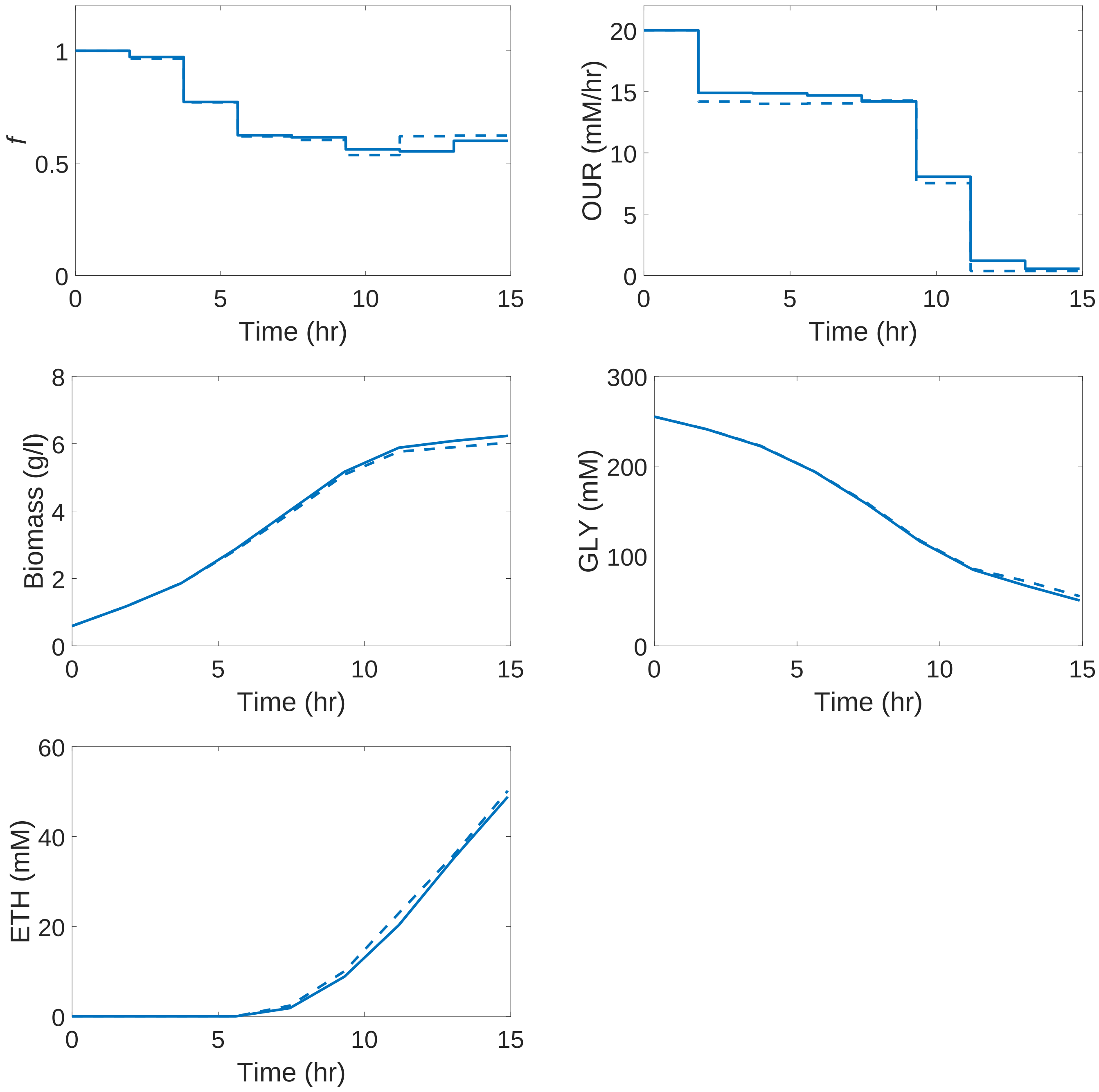


Figure 5.13: Performance of adaptive MPC/Case 1 with state estimation (solid line) and with full state information (dashed line).

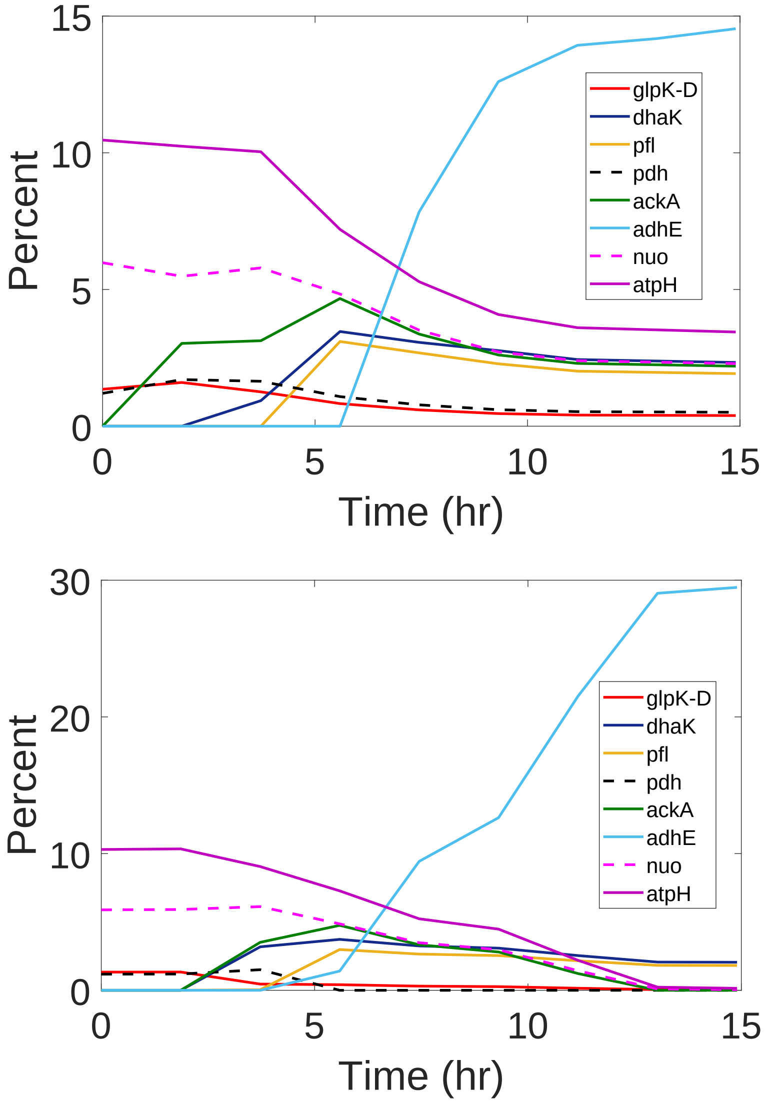


Figure 5.14: Biomass composition in adaptive MPC/Case 1 measured from the plant model (top) and estimated by the RBA (bottom).

the proposed adaptive control approach to address different biological states during the process and to handle associated uncertainties. It was shown that 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.

Moreover, the results of our simulation study showed 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 bioprocess. Here, constraint-based models are considered beneficial for our optimization and control purposes 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.

6 Conclusion and summary

Although Bioprocess optimization and control are commonly done based on basic unstructured models, the simple representation of cellular metabolisms in this models can not reflect the real bioprocess due to the complexity of the biological systems. Compared to them, constraint-based approaches are useful for bioprocess analysis and control as they address the full metabolic versatility by considering intracellular components connected to each other and to the environment through metabolic fluxes. In this work, we approached the model-based optimization and control of bioprocesses by exploiting the capabilities of constraint-based models and metabolic-genetic network models in particular. Beside cellular metabolism, metabolic-genetic network models include details on biomass synthesis and gene expression e.g. enzymes synthesis and degradation.

Within this thesis, the deFBA model as a dynamic metabolic-genetic network model has been implemented to efficiently derive optimal manipulation strategies for improving bioprocess productivity. The details on gene level included in this model allows for direct temporal manipulation of gene expression (which could not be addressed via previous studies) and makes them a suitable approach to be applied for model-based bioprocess analysis and control. However, models with additional metabolic details involve larger sets of model parameters which needs to be identified correctly. This causes a higher level of model uncertainties for detailed metabolic-genetic models compared to simple unstructured models, which limits their application in bioreactor control. Therefore, in this thesis we made efforts to present an adaptive model-based approach based on metabolic-genetic network models which allows for deriving flexible and robust process control strategies, while different levels of cellular manipulation (process and gene levels) are feasible. Below we summarize the thesis contents which have been addressed in several chapters.

In Chapter 2, we reviewed methods for constraint-based modeling of cellular metabolism including the flux balance analysis (FBA) and its extended versions. Then metabolic-genetic network models from stationary to dynamic ones are presented in details with a focus on the dynamic enzyme-cost FBA (deFBA) model, as the underlying approach for model-based bioprocess optimization and control in this thesis.

In Chapter 3, we have focused on the application of FBA-based models in bioprocess analysis and optimization especially for improving bioprocess productivity. In this direction, we have formulated a bilevel optimization approach based on the deFBA model to identify optimal control strategies for our target engineering objective (productivity). This approach has been implemented to maximize ethanol productivity in a batch growth of *E. coli* using a derived metabolic-genetic network model. Based on the results obtained in Chapter 3, we argue that the proposed linear deFBA-based bilevel approach is a promising approach in predicting suitable genetic and process level regulations in order to guide metabolic engineering and bioprocess optimization, as it accounts for gene expression (including enzymes production and degradation) besides cellular metabolism and allows to directly regulate genetic networks.

Consequently in Chapters 4 and 5, we have focused on flexible model-based control of bioprocesses for obtaining control strategies robust against process uncertainties. In the fourth chapter, a closed-loop control of the bioprocess is addressed. The bilevel deFBA-based approach has been implemented within a control algorithm based on model predictive control which allows for feedback corrections based on online measurements. By applying the control approach to a case study for improved ethanol productivity in a fed-batch growth of *E. coli*, the importance of online control for reliable model-based control of the bioprocess has been shown, as it allows for compensating plant-model mismatches.

In chapter 5, we improved the flexibility of the proposed deFBA-based control approach by integrating online adaptation schemes which are suitable to control highly uncertain biological processes. To this aim, an adaptive control algorithm has been proposed composed of model predictive control and moving horizon estimation, and an algorithm for estimating unknown states (here for biomass components). The adaptive approach could allow for online adaptation of the underlying model (deFBA) by estimating uncertain and variable model parameters in different stages of the process. As there is a relatively larger number of parameters in constraint-based modeling of metabolic-genetic networks, parameter selection for identifying relevant parameters to be used for model adaptation is an important step to be considered. In this direction, the proposed control algorithm includes steps for parameter selection by performing sensitivity analysis at each iteration of the process to identify parameters sufficient for model adaptation at each specific growth mode. As an application, microaerobic batch growth of *E. coli* for improved ethanol productivity from glycerol has been considered for evaluating the performance of the proposed control approach. Considering several simulation cases with different levels of uncertainties, the capabilities of the proposed adaptive approach to control highly dynamic and uncertain processes have been highlighted.

Overall, we argue that the proposed deFBA-based control scheme is a suitable approach to control time-varying 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.

Although only simulation-based studies have been considered during this study, as a further work a practical implementation of the proposed control approach is a highly recommended step. An experimental application of the proposed model-based approach for both open-loop and closed-loop control allows to evaluate the results for productivity improvements by suggested dynamic genetic and process level strategies as well as the improvements by feedback corrections, theoretically achieved in this study. Moreover, the performance of the adaptive closed-loop control needs be tested and investigated experimentally (e.g. for the considered case study; the microaerobic growth of *E. coli*) to recognize possible limitations of being applied in practice, guiding for further modifications of the proposed approach.

Besides, it is important to extend the applicability of the proposed approach for larger (genome scale) models. Although dealing with larger networks for optimization and control sounds feasible, it may enforce some challenges due to the higher model complexity. For instance when considering a large metabolic-genetic network within the adaptive MPC algorithm, proper selection of uncertain parameters of course requires integrating a preciser parameter identifier than a local sensitivity analysis implemented within this work. Moreover, additional cares are also required to justify computational limitations for this case, as a larger set of parameters are required to be identified and estimated through the proposed sequential MHE, which makes the estimation part computationally very expensive.

Furthermore, dealing with such non-convex bilevel optimizations can be prohibitive for MPC implementations from computational point of view, especially in case of nonlinear optimization problem (as implemented in Chapter 4 for the fed-batch bioreactor control). In this direction, modifications can be considered to speed up the calculations, such as efficient transformation of the bilevel optimization problem to a one-level problem, or selection of suitable algorithms for numerical solutions.

Moreover, modifications on the underlying model itself can be taken into account. Although the deFBA model includes details on biomass components, if available integrating any additional gene level data and regulatory information is still plausible to improve prediction capabilities. Such data can be integrated within the model to additionally constrain metabolic fluxes during different growth modes and to address corresponding regulatory mechanism. For example, it was experimentally observed that there is a switch between contribution of different enzymes responsible for respiration (cytochrome bo3 ubiquinol oxidase and cytochrome bd-I ubiquinol oxidase) at different oxygen-limited levels, while such an observation can not be addressed only based on the optimization principle and additional flux constraints are required to address this switch.

As an interesting extension of the deFBA model, one can also consider a combination of relevant objective functions (e.g. maximization of the overall ATP production or minimization of the substrate consumption beside considering the growth maximization), with the possibility to switch between different objective functions during different growth phases [108]. For instance, a weighted sum of relevant objectives can be considered in the model, such that the weight coefficient of each objective is determined at each MPC iteration based on the available measurement data. This in fact allows for proper adjustments in contribution of objectives at each growth mode e.g. during the exponential growth of the cell or in nutrient scarcity conditions.

A dFBA-based control problem for maximal productivity

Here, we present a dFBA-based bilevel problem for optimizing bioprocess productivity based on [17]. In the outer optimization the optimal batch time t_f and the optimal time of regulation for the manipulated flux t_{reg}, are determined for maximum productivity of a target product z_t. In the inner problem, the DOA approach of dFBA is implemented to determine the optimal metabolic flux distribution for a cellular objective optimization (here maximization of the biomass integral). The formulated dynamic bilevel problem with the dFBA model is as:

$$
\begin{aligned}
\underset{t_f, t_{reg}}{\text{maximize}} \quad & \frac{z_t(t_f)}{t_f} \\
\text{s.t.} \quad & t_0 \le t_{reg} \le t_f, \quad t_{min} \le t_f \le t_{max}, \\
& \underset{v(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} X(t)dt \\
& \text{s.t.} \quad \dot{z}(t) = S_e v(t) X(t), \\
& \qquad \dot{X}(t) = \mu(t) X(t), \\
& \qquad S_{int} v(t) = 0, \\
& \qquad v_{min}(t) \le v(t) \le v_{max}(t), \\
& \qquad v^{uptake}(t) \le \frac{v_{max}^{uptake}}{1 + z(t)/K_I}, \\
& \qquad z(t_0) = z_0, \quad X(t_0) = X_0, \\
& \qquad z(t) \ge 0, \quad X(t) \ge 0, \\
& \qquad v_{reg}(t) = 0 \quad \text{for} \quad t \ge t_{reg}.
\end{aligned}
\tag{A.1}
$$

After linearization of the inner dFBA model using the flux variable transformation ($V(t) = v(t)X(t)$) and the linear approximation of the growth inhibition constraint for v^{uptake} (by considering two tangent planes $L_1(X, z)$ and $L_2(X, z)$ of its nonlinear function), the dynamic bilevel problem can be described as:

$$\begin{aligned}
&\underset{t_f, t_{reg}}{\text{maximize}} && \frac{z_t(t_f)}{t_f} \\
&\text{s.t.} && t_0 \le t_{reg} \le t_f, \quad t_{min} \le t_f \le t_{max}, \\
& && \underset{V(\cdot)}{\text{maximize}} \int_{t_0}^{t_f} X(t) dt \\
& && \text{s.t.} \quad \dot{z}(t) = S_e V(t), \\
& && \qquad \dot{X}(t) = V_{growth}(t), \\
& && \qquad S_{int} V(t) = 0, \\
& && \qquad v_{min}(t) X(t) \le V(t) \le v_{max}(t) X(t), \\
& && \qquad V^{uptake}(t) - L_1(X(t), z(t)) \le m\lambda(t), \\
& && \qquad V^{uptake}(t) - L_2(X(t), z(t)) \le m(1 - \lambda(t)), \\
& && \qquad z(t_0) = z_0, \quad X(t_0) = X_0, \\
& && \qquad z(t) \ge 0, \quad X(t) \ge 0, \\
& && \qquad V_{reg}(t) = 0 \quad \text{for} \quad t \ge t_{reg}.
\end{aligned} \tag{A.2}$$

For solving the resulted bilevel problem, the inner linearized dFBA is approximated by discretization of dynamic variables in the time domain (similar to the procedure explained in Section 2.3.2). With N time intervals and K collocation points within each interval, $V(t)$ and $\dot{z}(t)$ are discretized at collocation points while the state variable z is discretized at the boundaries of time intervals. To keep the growth inhibition inequality constraints active during the entire time interval, the binary variable $\lambda(t)$ is considered to be applied at each collocation point. In this way, the dFBA problem is transformed into a mixed-integer linear program (MILP) with the variable vectors W and λ, defined as

$$W = (V_{1,1}, V_{1,2}, ..., V_{N,K}, \dot{z}_{1,1}, \dot{z}_{1,2}, ..., \dot{z}_{N,K}, z_1, z_2, ..., z_N),$$

$$\lambda = (\lambda_{1,1}, \lambda_{1,2}, ..., \lambda_{N,K}).$$

B MATLAB implementation of RBA algorithm

The implemented algorithms in this work are based on metabolic-genetic network models. In this section we provide the MATLAB scripts of the RBA algorithm, as the base and steady state formulation of metabolic-genetic network models. However, the details provided here can be helpful enough in better understanding of implementations for dynamic algorithms used within this study as well.

The included RBA simulation code corresponds to metabolic and genetic networks of *E. coli* growing on glycerol, preseneted in Tables 5.1 and 5.2.

M-file 1:

```
function Xp= RBA(B0)     % B0=total biomass concentration

Nu=50;        %total number of reactions
Num=25;       %number of metabolic reactions
Nup=25;       %number of biomass production reactions
Np=25;        %number of macromolecules
Nxint=21;     %number of internal metabolites

%variable vector W includes fluxes V, macromolecules p, growth rate mu
% V=W(1:Nu); p=W((Nu+1:Nu+Np); mu=W(Nu+Np+1)

%weights of macromolecules (g/mmol)
bT=100*[1.8061     3.9861    1.8486     1.3898     0.2452     0.4774
    1.4453     0.6911     2.0961     2.0492     1.9249    45.8846    3.0923
    3.8499   2.1179     4.9758     4.6772    38.8476     6.1858     1.4072
    4.6674     5.3355    0.6813     22.9242     0.65];

%equality constraints: Sm*V=0  and bT*p=B0
Aeq=zeros(Nxint+1,Nu+Np+1);
beq=zeros(Nxint+1,1);

%FBA constraint: Sm*V=0
Sm=input('load the stoichiometric matrix');
Aeq(1:Nxint,1:Nu)=Sm;

%biomass amount in the system: bT*p=B0
Aeq(Nxint+1,Nu+1:Nu+Np)=bT;
beq(Nxint+1,1)=B0;

%inequality constrains: enzyme capacity and biomass composition
AA=zeros(25+1,Nu+Np+1);
```

```
BB=zeros(25+1,1);

%enzyme capacity constraints  V/kcat-p<0 in matrix form => M1*V-M2*p<0
kcat(1:Num)=60*[1998 1447  630 1326 78000 3000 3402 780 3402 3162 3960 768 29160 169980 32400 318 2684 800 942 inf 360 18000 2220 3300 3996]';  %catalytic constants for metabolic reactions

kcat(Num+1:Nu)=60*[0.43 0.2 0.42 0.56 3.2 1.64 0.54 1.14 0.37 0.38 0.41 0.02 0.25 0.2 0.37 0.16 0.17 0.02 0.13 0.56 0.17 0.15 1.1 0.1 28.8]';  %translation rates for macromolecules production reactions

M1=zeros(25,Nu);
for i=1:8
    M1(i,i)=1/kcat(i);
end
M1(7,9)=1/kcat(9);
for i=9:24
    M1(i,i+1)=1/kcat(i+1);
end
for i=Num+1:Nu
M1(25,i)=0.01/kcat(i);
end
AA(1:25,1:Nu)=M1;

M2=zeros(25,Np);
for i=1:18
    M2(i,i)=-1;
end
for i=19:24
    M2(i+1,i)=-1;
end
AA(1:25,Nu+1:Nu+Np)=M2;

%biomass composition constraint : 0.65*bT*p<Q in matrix form => CC*p<0
CC(1,1:Np)=0.65*bT;
CC(1,Np)=(0.65-1)*b(Np);
AA(25+1,Nu+1:Nu+Np)=CC;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%optimization by ipopt through OPTI toolbox
lb=zeros(Nu+Np+1,1);                %lower bounds
ub=1000*ones(Nu+Np+1,1);            %upper bounds
W0=ones(Nu+Np+1,1);                 %initial value
fun=@(W) objectivefun(W);           %objective function: mu
nlcon=@(W) nonlinearcon(W);         %nonlinear constraint: Sp*Vp-mu*p=0
cu=zeros(Np,1);                     %lower bound of the nonlinear constraint
cl=zeros(Np,1);                     %upper bound of the nonlinear constraint

opts = optiset('solver','ipopt','maxiter',3000,'display','iter');

```

```
%Build OPTI Problem
Opt = opti('fun',fun,'eq',Aeq,beq,'ineq',AA,BB,'nl',nlcon,cl,cu,'bounds',lb,ub,'x0',W0,'options',opts);

%Solve NLP
[W,fval,exitflag] = solve(Opt)

Vm=W(1:Num);           %metabolic fluxes
Vp=W(Num+1:Nu);        %biomass production fluxes
p=W(Nu+1:Nu+Np)';      %concentration of macromolecules(biomass components)
mu=W(Nu+Np+1);         %growth rate

end
```

M-file 2:

```
%defining the objective function:  maximize mu
function objective= objectivefun( W )
mu=W(Nu+Np+1);
objective=-mu;
end
```

M-file 3:

```
%defining the nonlinear constraintÂ´: Sp*Vp-mu*p=0
function c = nonlinearcon( W )

V=W(1:Nu);
p=W(Nu+1:Nu+Np);
mu=W(Nu+Np+1);

Sp=0.01*diag(ones(1,Np));
c(1:Np)=(Sp*V(Num+1:Nu))-((mu*p));
end
```

Bibliography

[1] Markus W. Covert and Bernhard O. Palsson. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *The Journal of Biological Chemistry*, 277:28058–28064, 2002.

[2] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28:245–248, 2010.

[3] Naveen Venayak, Nikolaos Anesiadis, William R Cluett, and Radhakrishnan Mahadevan. Engineering metabolism through dynamic control. *Current Opinion in Biotechnology*, 34:142 – 152, 2015.

[4] Steffen Klamt, Oliver Hädicke, and Axel von Kamp. *Stoichiometric and Constraint-Based Analysis of Biochemical Reaction Networks*, pages 263–316. Springer International Publishing, Cham, 2014.

[5] Amit Varma and Bernhard O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and environmental microbiology*, 60:3724–3731, 1994.

[6] Charles G. Sinclair, Bjorn Kristiansen, and John D. Bu'Lock. *Fermentation kinetics and modelling.* Open University Press; Taylor & Francis, 1987.

[7] H.M. Tsuchiya, A.G. Fredrickson, and R. Aris. Dynamics of microbial cell populations. *Advances in Chemical Engineering*, 6:125–206, 1966.

[8] Michael A Henson. Dynamic modeling of microbial cell populations. *Current Opinion in Biotechnology*, 14(5):460–467, 2003.

[9] Arnold G. Fredrickson. Formulation of structured growth models. *Biotechnology and Bioengineering*, 67(6):720–725, 2000.

[10] Fabio R. Sidoli, A. Mantalaris, and Steven P. Asprey. Modelling of mammalian cells and cell culture processes. *Cytotechnology*, 44:27–46, 2004.

[11] Mihai Caramihai and Irina Severin. Bioprocess Modeling and Control. In *Biomass Now - Sustainable Growth and Use*. IntechOpen, 2013.

[12] Jörg Schubert, Rimvydas Simutis, Michael Dors, Ivo Havlik, and Andreas Lübbert. Bioprocess optimization and control: Application of hybrid modelling. *Journal of Biotechnology*, 35(1):51–68, 1994.

[13] Coleman Brosilow and Babu Joseph. *Techniques of Model-based Control.* Prentice-Hall International Series in the Physical and Chemical Engineering Sciences. Prentice Hall, 2002.

[14] Amit Varma and Bernhard O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12:994–998, 10 1994.

[15] 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.

[16] Radhakrishnan Mahadevan, Anthony P. Burgard, Iman Famili, Steve Van Dien, and Christophe H. Schilling. Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals. *Biotechnology and Bioprocess Engineering*, 10(5):408–417, 2005.

[17] Banafsheh Jabarivelisdeh and Steffen Waldherr. Improving bioprocess productivity using constraint-based models in a dynamic optimization scheme. *IFAC-PapersOnLine*, 49(26):245 – 251, 2016.

[18] 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.

[19] 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.

[20] Banafsheh Jabarivelisdeh, Lisa Carius, Rolf Findeisen, and Steffen Waldherr. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Computers & Chemical Engineering*, 135:106744, 2020.

[21] Gregory N. Stephanopoulos, Aristos A. Aristidou, and Jens Nielsen. Metabolic flux analysis. In *Metabolic Engineering*, pages 309–351. 1998.

[22] Jeremy S. Edwards, Rafael U. Ibarra, and Bernhard O. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19(2):125–130, 2001.

[23] Adam M. Feist, Markus J. Herrgård, Ines Thiele, Jennie L. Reed, and Bernhard O. Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7:129–143, 2009.

[24] Markus W. Covert, Christophe H. Schilling, and Bernhard O. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of theoretical biology*, 213:73–88, 12 2001.

[25] Sriram Chandrasekaran and Nathan D. Price. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 107(41):17845–17850, 2010.

[26] Christopher S. Henry, Linda J. Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, 92(5):1792 – 1805, 2007.

[27] Andreas Hoppe, Sabrina Hoffmann, and Hermann-Georg Holzhütter. Including metabolite concentrations into flux balance analysis: Thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC systems biology*, 1:23, 2007.

[28] Anna Blazier and Jason Papin. Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in physiology*, 3:299, 2012.

[29] Tomer Shlomi, Moran Cabili, Markus Herrgard, Bernhard O. Palsson, and Eytan Ruppin. Network-based prediction of human tissue-specific metabolism. *Nature biotechnology*, 26:1003–1010, 2008.

[30] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, 2002.

[31] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10:291–305, 2012.

[32] Mohsen Razzaghi, Jalal Nazarzadeh, and KY Nikravesh. A collocation method for optimal control of linear systems with inequality constraints. *Mathematical Problems in Engineering*, 3(6):503–515, 1998.

[33] Lorenz T. Biegler. An overview of simultaneous strategies for dynamic optimization. *Chemical Engineering and Processing: Process Intensification*, 46(11):1043–1053, 2007.

[34] Adam L. Meadows, Rahi Karnik, Harry Lam, Sean Forestell, and Brad Snedecor. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metabolic Engineering*, 12(2):150 – 160, 2010.

[35] Robert Flassig, Melanie Fachet, Kai Höffner, Paul Barton, and Kai Sundmacher. Dynamic flux balance modeling to increase the production of high-value compounds in green microalgae. *Biotechnology for Biofuels*, 9:165, 2016.

[36] Timothy J. Hanly and Michael A. Henson. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnology and Bioengineering*, 108(2):376–385, 2011.

[37] 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.

[38] Jared L Hjersted and Michael A Henson. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. *Biotechnology Progress*, 22:1239–48, 10 2006.

[39] Michael Henson and Timothy Hanly. Dynamic flux balance analysis for synthetic microbial communities. *IET Systems Biology*, 8:214–229, 2014.

[40] Jan Schellenberger, Richard Que, Ronan Fleming, Ines Thiele, Jeff Orth, Adam

Feist, Daniel Zielinski, Aarash Bordbar, Nathan Lewis, Sorena Rahmanian, Joseph Kang, Daniel Hyduke, and Bernhard Palsson. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols*, 6:1290–1307, 2011.

[41] Jose Gomez, Kai Höffner, and Paul Barton. DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinformatics*, 15:409, 2014.

[42] Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210 – 1218, 2011.

[43] Joshua A Lerman, Daniel Hyduke, Haythem Latif, Vasiliy Portnoy, Nathan Lewis, Jeff Orth, Alexandra Schrimpe-Rutledge, Richard Smith, Joshua N Adkins, Karsten Zengler, and Bernhard O. Palsson. In silico method for modelling metabolism and gene product expression at genome scale. *Nature Communications*, 3:929, 2012.

[44] Ines Thiele, Neema Jamshidi, Ronan M. T. Fleming, and Bernhard O. Palsson. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLoS Computational Biology*, 5:130 – 141, 2009.

[45] 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.

[46] Alexandra-M. Reimers, Henning Lindhorst, and Steffen Waldherr. A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites*, 7(3):47, 2017.

[47] Steven A. Frank. The trade-off between rate and yield in the design of microbial metabolism. *Journal of Evolutionary Biology*, 23(3):609–613, 2010.

[48] Jacques Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 1949.

[49] Anthony P. Burgard, Shankar Vaidyaraman, and Costas D. Maranas. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, 17(5):791–797, 2001.

[50] Madalena Chaves and Diego A Oyarzún. Dynamics of complex feedback architectures in metabolic pathways. *Automatica*, 99:323–332, 2019.

[51] Navtej Juty, Rao Ali, Mihai Glont, Sarah Keating, Nicolas Rodriguez, Maciej Swat, Sarala Wimalaratne, Henning Hermjakob, Nicolas Novére, Camille Laibe, and Vijayalakshmi Chelliah. BioModels: Content, features, functionality, and use. *CPT: Pharmacometrics and Systems Pharmacology*, 4:55–68, 2015.

[52] Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard Palsson, and Nathan Lewis.

BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44:D515–D522, 2016.

[53] Alexandra-M. Reimers, Henning Knoop, Alexander Bockmayr, and Ralf Steuer. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences*, 114(31):E6457–E6465, 2017.

[54] Annika Röhl and Alexander Bockmayr. A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks. *BMC Bioinformatics*, 18:2, 2017.

[55] Philipp Erdrich, Ralf Steuer, and Steffen Klamt. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Systems Biology*, 9:48, 2015.

[56] Dennis A. Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. GenBank. *Nucleic Acids Research*, 41:D36–D42, 2013.

[57] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2016.

[58] Minoru Kanehisa and Susumu Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[59] Akihiro Nakao, Maki Yoshihama, and Naoya Kenmochi. RPG: the Ribosomal Protein Gene database. *Nucleic Acids Research*, 32:D168–D170, 2004.

[60] 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.

[61] 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.

[62] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, Olga Krebs, Meik Bittkowski, Elina Wetsch, Isabel Rojas, and Wolfgang Müller. SABIO-RK– database for biochemical reaction kinetics. *Nucleic Acids Research*, 40:D790–D796, 2012.

[63] Edward J O'Brien, Joshua A Lerman, Roger Chang, Daniel Hyduke, and Bernhard O. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9:693, 2013.

[64] Jo Maertens and Peter A Vanrolleghem. Modeling with a view to target identification in metabolic engineering: A critical evaluation of the available tools.

Biotechnology Progress, 26:313–331, 2010.

[65] Ali R. Zomorrodi, Patrick F. Suthers, Sridhar Ranganathan, and Costas D. Maranas. Mathematical optimization applications in metabolic networks. *Metabolic Engineering*, 14(6):672 – 686, 2012.

[66] 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.

[67] Nathan S Mosier and Michael R Ladisch. Modern biotechnology: connecting innovations in microbiology and biochemistry to engineering fundamentals. Hoboken: John Wiley & Sons, 2009.

[68] Vasiliy A Portnoy, Daniela Bezdan, and Karsten Zengler. Adaptive laboratory evolution - harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology*, 22(4):590 – 594, 2011.

[69] 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.

[70] Harry Yim, Robert Haselbeck, Wei Niu, Catherine Pujol-Baxley, Anthony Burgard, Jeff Boldt, Julia Khandurina, John Trawick, Robin E Osterhout, Rosary Stephen, Jazell Estadilla, Sy Teisan, Brett Schreyer, Stefan Andrae, Tae Hoon Yang, Sang Yup Lee, Mark Burk, and Steve Van Dien. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology*, 7:445–452, 2011.

[71] Naama Tepper and Tomer Shlomi. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, 26:536–543, 2009.

[72] Kiran Raosaheb Patil, Isabel Rocha, Jochen Förster, and Jens Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 6:308, 2005.

[73] Priti Pharkya, Anthony P Burgard, and Costas Maranas. OptStrain: A computational framework for redesign of microbial production systems. *Genome Research*, 14:2367–2376, 2004.

[74] Priti Pharkya and Costas D. Maranas. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering*, 8(1):1–13, 2006.

[75] Joonhoon Kim and Jennifer L Reed. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Systems Biology*, 4:53, 2010.

[76] Sridhar Ranganathan, Patrick F. Suthers, and Costas D. Maranas. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. 6(4):e1000744, 2010.

[77] Desmond Lun, Graham Rockwell, Nicholas J Guido, Michael Hartmann Baym, Jonathan Adam Kelner, Bonnie Berger, James E Galagan, and George Church. Large-scale identification of genetic design strategies using local search. *Molecular Systems Biology*, 5:296, 2009.

[78] Oliver Haedicke and Steffen Klamt. CASOP: A computational approach for strain optimization aiming at high productivity. *Journal of Biotechnology*, 147(2):88 – 101, 2010.

[79] Kai Zhuang, Laurence Yang, William R Cluett, and Radhakrishnan Mahadevan. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySScO strategy for strain design. *BMC Biotechnology*, 13:8, 2013.

[80] Jaoon Young Hwan Kim and Hyung Joon Cha. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: improved foreign protein production. *Biotechnology and Bioengineering*, 83:841–853, 2003.

[81] Han-Saem Cho, Sang Woo Seo, Young Mi Kim, Gyoo Yeol Jung, and Jong Moon Park. Engineering glyceraldehyde-3-phosphate dehydrogenase for switching control of glycolysis in *Escherichia coli*. *Biotechnology and Bioengineering*, 109(10):2612–2619, 2012.

[82] Apoorv Gupta, Irene Reizman, Christopher Reisch, and Kristala Prather. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology*, 35(3):273–279, 2017.

[83] Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173 – 1183, 2013.

[84] John Hawkins, Spencer Wong, Jason Peters, Ricardo Almeida, and Lei Qi. Targeted transcriptional repression in bacteria using CRISPR interference (CRISPRi). *Methods in Molecular Biology*, 1311:349–362, 2015.

[85] Kapil G. Gadkar, Radhakrishnan Mahadevan, and Francis J. Doyle III. Optimal genetic manipulations in batch bioreactor control. *Automatica*, 42(10):1723 – 1733, 2006.

[86] Nikolaos Anesiadis, William R. Cluett, and Radhakrishnan Mahadevan. Dynamic metabolic engineering for increasing bioprocess productivity. *Metabolic Engineering*, 10(5):255 – 266, 2008.

[87] Nikolaos Anesiadis, Hideki Kobayashi, William R Cluett, and Radhakrishnan Mahadevan. Analysis and design of a genetic circuit for dynamic metabolic

engineering. *ACS Synthetic Biology*, 2:442–452, 2013.

[88] Timothy J. Hanly, Aubrey R. Tiernan, and Michael A. Henson. Validation and optimization of a yeast dynamic flux balance model using a parallel bioreactor system. *IFAC Proceedings Volumes*, 46(31):113 – 118, 2013.

[89] Lisha K. Parambil and Debasis Sarkar. In silico analysis of bioethanol overproduction by genetically modified microorganisms in coculture fermentation. *Biotechnology Research International*, 2015:1–11, 2015.

[90] Joshua K. Michener, Kate Thodey, Joe C. Liang, and Christina D. Smolke. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metabolic Engineering*, 14(3):212 – 222, 2012.

[91] Jin I. Kim, Jeffery D. Varner, and Doraiswami Ramkrishna. A hybrid model of anaerobic *E. coli* GJT001: Combination of elementary flux modes and cybernetic variables. *Biotechnology Progress*, 24(5):993–1006, 2008.

[92] Xiongfeng Dai, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Patsalo, Hiroyuki Okano, James Williamson, Kurt Fredrick, Yi-Ping Wang, and Terence Hwa. Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nature Microbiology*, 2:16231, 2017.

[93] David P. Clark. The fermentation pathways of *Escherichia coli*. *FEMS Microbiology Letters*, 63(3):223 – 234, 1989.

[94] Neha Munjal, Anu J. Mattam, Dibyajyoti Pramanik, Prem Sh. Srivastava, and Syed Sh. Yazdani. Modulation of endogenous pathways enhances bioethanol yield and productivity in *Escherichia coli*. *Microbial Cell Factories*, 11(1):145, 2012.

[95] Irene M. Brockman and Kristala L.J. Prather. Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metabolic Engineering*, 28:104–113, 2015.

[96] Björn-Johannes Harder, Katja Bettenbrock, and Steffen Klamt. Temperature-dependent dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by *Escherichia coli*. *Biotechnology and Bioengineering*, 115(1):156–164, 2017.

[97] Alexandra-M. Reimers. *Understanding metabolic regulation and cellular resource allocation through optimization*. PhD thesis, Free University, Berlin, Germany, 2017.

[98] 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.

[99] James B. Rawlings, David Q. Mayne, and Moritz M. Diehl. Model Predictive Control: Theory, Computation, and Design. *Nob Hill Pub*, 2009.

[100] Veronica Adetola, Darryl DeHaan, and Martin Guay. Adaptive model predic-

tive control for constrained nonlinear systems. *Systems and Control Letters*, 58(5):320–326, 2009.

[101] James B. Rawlings. *Moving Horizon Estimation*, pages 1–7. Springer London, London, 2013.

[102] Peter Klaus Findeisen. *Moving Horizon State Estimation of Discrete Time Systems*. University of Wisconsin–Madison, 1997.

[103] Syed Sh. Yazdani and Ramon Gonzalez. Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Current Opinion in Biotechnology*, 18(3):213–219, 2007.

[104] Guyton Durnin, James Clomburg, Zeno Yeates, Pedro J. Alvarez, Kyriacos Zygourakis, Paul Campbell, and Ramon Gonzalez. Understanding and harnessing the microaerobic metabolism of glycerol in *Escherichia coli*. *Biotechnology and Bioengineering*, 103(1):148–161, 2009.

[105] 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.

[106] Svetlana Alexeeva, Klaas J. Hellingwerf, and M. Joost Teixeira de Mattos. Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *Journal of Bacteriology*, 185(1):204–209, 2003.

[107] Yunfei Chu and Juergen Hahn. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE Journal*, 53(11):2858–2870, 2007.

[108] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3(1):119, 2007.

Schriftliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Insbesondere habe ich nicht die Hilfe einer kommerziellen Promotionsberatung in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Banafsheh Jabarivelisdeh Magdeburg, den 5. März 2021

Publications by the author

Peer-reviewed international journals:

- **Banafsheh Jabarivelisdeh**, Lisa Carius, Rolf Findeisen, Steffen Waldherr, “Adaptive predictive control of bioprocesses with constraint-based modeling and estimation“, *Computers & Chemical Engineering*, 135: page 106744, 2020.
- **Banafsheh Jabarivelisdeh**, Steffen Waldherr, “Optimization of bioprocess productivity based on metabolic-genetic network models with bilevel dynamic programming“, *Biotechnology and Bioengineering*, 115(7): pages 1829-1841, 2018.
- **Banafsheh Jabbari**, Nassim Tahouni, Abtin Ataei, M. Hassan Panjeshahi, “Design and optimization of CCHP system incorporated into Kraft process, using pinch analysis, with pressure drop consideration“, *Applied Thermal Engineering*, 61(1): pages 88-97, 2013.

Peer-reviewed proceedings:

- **Banafsheh Jabarivelisdeh**, Rolf Findeisen, Steffen Waldherr, “Model Predictive Control of a Fed-batch Bioreactor Based on Dynamic Metabolic-Genetic Network Models“, *In Proceedings of Foundations of Systems Biology in Engineering (FOSBE)*, pages 34-37, Chicago, USA, 2018.
- **Banafsheh Jabarivelisdeh**, Steffen Waldherr, “Improving Bioprocess Productivity Using Constraint-Based Models in a Dynamic Optimization Scheme“, *In Proceedings of Foundations of Systems Biology in Engineering (FOSBE)*, pages 245-251, Magdeburg, Germany, 2016.
- **Banafsheh Jabbari**, Nassim Tahouni, M. Hassan Panjeshahi, “Optimal Design of a Cogeneration system in a Kraft Industry Using Genetic Algorithm“, *Chemical Engineering Transactions*, 29:pages 19-24, 2012.
- **Banafsheh Jabbari**, Nassim Tahouni, M. Hassan Panjeshahi, “Optimum Design of an Absorption Heat Pump Integrated with a Kraft Industry Using Genetic Algorithm“, *World Academy of Science, Engineering and Technology*, 6: pages 923-929, 2012.
- **Banafsheh Jabbari**, Nassim Tahouni, Mahdi Rezaei, M. Hassan Panjeshahi, “Improving Energy Efficiency in Pulp and Paper Industry, Using a CCHP System“, *7th International Chemical Engineering Congress & Exhibition*, Kish, Iran, 2011.

Banafsheh Jabarivelisdeh Magdeburg, den 5. März 2021