Mathematical Modeling of Metabolic-Genetic Networks

Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften *(Dr. rer. nat.)*

vorgelegt von

Neveen Ali Salem Eshtewy

am Fachbereich Mathematik und Informatik

der Freien Universität Berlin

Berlin 2020

Betreuer: PD Dr. Marcus Weber
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

Zweitgutachterin: Prof. Dr. Susanna Röblitz
Department of Informatics
University of Bergen, Norway

Tag der Disputation: 21. July 2020

Acknowledgements

First and foremost, I would like to thank my supervisor Marcus Weber, for giving me the opportunity to complete my PhD thesis in his group, for his helpful inputs, valuable feedback, and discussions. Thanks also to all the former members in the Weber's group.
Special thanks go to Prof. Hassan El-Owaidy whom I have also learned a lot during my Master's degree. I am deeply grateful to Dr. Lena Scholz, for all her guidance, ideas, several interesting discussions, proof-reading this thesis, and for sharing her knowledge and experience with me. I thank Prof. Jens Timmer's group, in particular, Dr. Joep Vanlier for helping me to apply D2D software and for discussion. I am also grateful to Prof. Andreas Kremling for his discussion and comments.
I also want to thank Han Lie for proof-reading some parts of this thesis and valuable comments. Many Thanks to Nada Cvetkovic for her motivation and emotional support. I thank all of my friends in the Berlin area for all various support and sharing all the nice moments: Furqan Muhayodin, Mohamed Omari, Ismail El-Shimy, Zltaka, Mariana, Eric, Ling Sun. Many thanks to Sally Said, Mohammed Brikaa.

I would like to thank the Egyptian Cultural Affairs and Missions Sector for funding the first two years of my PhD studies. I would like to also acknowledge funding from the German Research Foundation (DFG) via the RTG 1772 Computational Systems Biology (CSB). I also want to thank Prof. Edda Klip and Dr. Cordelia Arndt-Sullivan the coordinator of CSB group for their help and support.
My deepest heartfelt appreciation goes to all my family. My deep gratitude to my parents and my sisters, Hana, Nisreen, for their steady love, support, and encouragement. I also thank my brothers, Salman, Ayman, and Hany, for their constant emotional support. Words cannot express how thankful I am to have you all. Deep thanks to my sister Hana, without you, I wouldn't have been able to continue this PhD.

Contents

List of Tables

Abstract

Systems biology deals with the computational and mathematical modeling of complex biological systems. The aim is to understand the big picture of the system's dynamics rather than the individual parts by integrating different sciences, e.g., mathematics, physics, biology, computer science, and engineering. In biological systems, mathematical models of biochemical networks are necessary for predicting and optimizing the behavior of cells in culture. Different mathematical models have been discussed, such as discrete models, continuous models, and hybrid models. In a discrete model, the biological system assumes discrete values. A continuous model uses a system of differential equations to describe the change of concentrations of substances in a cell over time. A hybrid model combines both discrete and continuous models. The main challenge in continuous models is to find the kinetic parameter values. In this thesis, we build a kinetic model of a metabolic-genetic network introduced in Covert et al., 2001 that mimics a discrete model of regulatory flux balance analysis (*rFBA*) which is based on steady-state assumptions. The kinetic model we introduce has unknown parameters, so it is necessary to perform parameter estimation techniques. We perform a parameter estimation technique using data sets generated from a simulation of the rFBA model. In nature, many phenomena of interest are high-dimensional and complex. Thus, model reduction is considered a vital topic in systems biology. Model reduction methods are mathematical techniques that aim to represent a high-dimensional, dynamical system by a low-dimensional system that roughly preserves the main features and characteristics of the original system. The idea of model order reduction is to use the reduced-order model instead of the full-order model in the simulation or optimization of the system to reduce the computational effort and the runtime of the simulations. In this thesis, we discuss two different model reduction methods. The first method assumes a time scale separation, i.e., it assumes two time scales, a fast time scale and a slow time scale, where the fast time scale dynamics converge to a quasi-steady state. The second approach, proper orthogonal decomposition, aims at obtaining low-dimensional approximate descriptions of high-dimensional processes while retaining the most important features of the dynamics. We apply these approaches to different biological system models from the BioModels database.

Zusammenfassung

Die Systembiologie beschäftigt sich mit der computergestützten und mathematischen Beschreibung biologischer Systeme. Ziel ist dabei, die Gesamtheit der Dynamik des Systems zu verstehen, und nicht etwa nur individuellen Teile. In biologischen Systemen werden mathematische Modelle biochemischer Netzwerke benötigt um das Verhalten von Zellen vorher zu sagen und in Zellkulturen zu optimieren. Unterschiedlichste Modelle sind im Diskurs – diskrete Modelle, kontinuierliche Modelle, Hybridmodelle. In einem diskreten Modell nimmt das biologische System diskrete Werte an. Ein kontinuierliches Modell verwendet ein System von Differentialgleichungen um die veränderung der Konzentration von Zellsubstanzen über einen bestimmten Zeitraum zu beschreiben. Ein Hybridmodell kombiniert diskrete und kontinuierliche Modelle. Die größte Herausforderung eines kontinuierlichen Modells ist die kinetische Reaktionsrate zu finden. In dieser Arbeit wurde ein kontinuierliches Modell eines metabolisch-genetischen Netzwerks nach Covert et al. 2001 aufgebaut, welches auf einer steady-state Annahmen beruht und ein diskretes Modell einer regulatorischen Flux-Gleichgewichts-Analyse (rFBA) nachahmt. Da die Parameter des kinetischen Systems unbekannt sind, müssen diese abgeschätzt werden. Dazu werden Datensätze aus rFBA Simulationen generiert, welche zur Kalibrierung verwendet werden. Viele natürliche Systeme, die von Interesse sind, sind hoch-dimensional und komplex. Folglich ist die Reduktion von Modellen ein integraler Bestandteil der Systembiologie. Modellreduktionsmethoden sind mathematische Techniken, die dazu dienen ein hoch-dimensionales, dynamisches System in einem niedriger-dimensionalen System, welches grob die wesentlichen Eigenschaften und Charakteristika erhält, darzustellen. Die Idee hinter der Reduktion einer Modellordnung ist ein System niedrigerer Ordnung an Stelle des vollen Models zur Simulation oder Optimierung des Systems zu verwenden, um somit den rechnerischen Aufwand und die Rechenzeit der Simulationen zu reduzieren. In dieser Arbeit werden zwei verschiedene Methoden zur Modellreduktion diskutiert. In der ersten Methode wird eine Separierung der Zeitskala angenommen – eine langsame und eine schnelle Zeitskala. In der zweiten Herangehensweise, welche 'Proper Orthogonal Decomposition' genannt wird, werden niedrig-dimensionale Näherungen höher-dimensionaler Prozesse generiert, wobei wiederum die wichtigsten Eigenschaften der ursprünglichen Dynamik erhalten bleiben. Diese Methoden werden auf unterschiedliche Modelle der BioModel Datenbank angewendet.

Contributions

In biological systems, mathematical models of biochemical networks are necessary for predicting and optimizing the behavior of cells in culture. Different mathematical models have been introduced, such as discrete models, continuous models, and hybrid models to describe the behavior of a cell. In this thesis, we have addressed two main topics. First, we propose how to build a continuous model for given discrete and hybrid models of a metabolic-genetic network to overcome the limitation and the difficulty of these models. Second, we apply two different model order reduction methods of the kinetic model of biological systems, where the kinetic models of many phenomena of interest are high-dimensional and complex resulting in large computational effort in the simulation.

The continuous model we introduce enables us to study the dynamic behavior of every single component of the biological system which helps in understanding the full picture of the system. Additionally, various mathematical theories may be applied to a continuous model, e.g., for parameter estimation, model order reduction, etc. However, the procedure of finding parameter values (e.g., constant reaction rates) for a continuous model is still a challenge in biological systems. This requires us sometimes to apply parameter estimation techniques to obtain parameter values. In contrast, a discrete model of regulatory flux balance analysis (rFBA) uses only a few parameters because rFBA is based on steady-state assumptions. However, even though rFBA uses a few parameters, its main limitation is that it is not able to predict the dynamic behavior of every component which the continuous model is able to achieve (Chapter 2).
The metabolic-regulatory network is considered as an example of a hybrid system. However, the hybrid model becomes complicated if the number of regulatory rules is large because the modeling of regulatory rules using the boolean functions and detecting of events becomes difficult. We have introduced a continuous model of the hybrid system of a metabolic-regulatory network. We have shown that by using our continuous model, we can more easily obtain the same results as those produced by a hybrid model for a metabolic regulatory network (Chapter 3).

In the second primary topic, we have explained a model order reduction methods (MOR). The MOR methods are used to reduce the computational complexity of high dimensional systems by approximating them with lower-dimensional systems while retaining the important information and properties of the full order system. First, we have discussed the time scale separation technique, a major class of MOR methods in system biology. The time scale separation technique assumes two time scales, a fast time scale and a slow time scale, where the fast time scale dynamics converge to a quasi-steady state. The available literature was not clear enough about how to satisfy the required conditions of this technique to biological systems. We have clarified and explained the idea behind the technique and applied it to different metabolic-genetic networks. In our study, we checked numerically that the required conditions were satisfied (Chapter 4). The advantage of the time scale separation technique is that it preserves the dynamics of the original system. However, it is difficult to apply it to a large scale system since the required conditions (Tikhonov's theorem) of this technique are difficult to satisfy. The technique requires the separation of slow and fast components which is difficult to do in large scale systems. This difficulty motivated us to use another technique that overcomes all the difficulties of the time scale separation technique. The technique which we use is called proper orthogonal decomposition (POD). The POD technique works by projecting a high-dimensional system onto a lower-dimensional space. The POD technique works well for large-scale systems, but the dynamics are given for a surrogate model that is obtained via a projection of the original model. Thus, it is difficult to predict the contribution of a specific single component.
We successfully applied the POD technique to the different biological systems from the BioModels database. We have used the POD to compute a reduced-order model for different initial conditions of the dynamical system. Using different initial conditions in the time scale separation technique will usually fail, because the algebraic equations of the fast variables may not be solvable if inconsistent initial values are prescribed (Chapter 5).

Chapter 1

Introduction

Systems biology is an integrated approach to study the behavior of biological processes. It tries to put all the pieces of a biological system together to understand how all the pieces interact [21]. Research in systems biology is based on an interdisciplinary approach studying the whole living system rather than individual components [5; 82]. It integrates different sciences, e.g., mathematics, biology, computer science, engineering, and physics, to understand the various interacting roles involved, see [49; 79; 149].

Systems biology relies on simplified mathematical models of cells and organisms. Often, mathematical models are used to study complicated objects, such as population dynamics [70; 112], regulatory networks [20], or metabolic networks [37]. The goal is to develop models that capture the essence of various interactions while allowing their output to be more fully understood [109]. Mathematical modeling is considered a powerful tool for gaining an understanding of the functioning of the large and complicated systems where the good models show how the process works and then predict what may follow [1; 40; 86].

The mathematical model of a dynamic system can be characterized by a set of differential equations [3; 126] which are about continuous flows or difference equations [77; 87] which are useful in describing discrete problems. These equations, form the mathematical model of the phenomena of interest. The parameters and variables that define the equations can be measured or estimated using experimental data. Different forms of mathematical models have been discussed, such as discrete models, continuous models, and hybrid models. In a discrete model, the biological system assumes discrete values [80; 135]. A continuous model uses a system of differential equations to describe the change of concentrations of substances in a cell over time [71; 74; 143]. A hybrid model combines both discrete and continuous models [8; 65]. A continuous model is considered to be more effective and provides a useful level of

description of the biological system. It is considered a convenient mathematical model to which many theorems can be applied, e.g., model reduction and parameter estimation. On the other hand, a continuous model can be complicated because it requires more information about kinetic parameters which still poses a challenge in biological study. Often, discrete models are used in biological systems to describe gene expression networks. The gene expression networks usually are described by the boolean network model [76; 146] where genes switch between two states "on" or "off".

Here, in this thesis, we focus on the study of a mathematical model that is described by a system of ordinary differential equations. We are interested in a continuous model that can predict the full picture of the system behavior to which we can apply different theories. In our study, we discuss how to express discrete and hybrid models by continuous models and predict almost the same result as for the discrete models in Chapters 2,3. In addition, we will study model reduction methods of continuous mathematical models in Chapters 4,5.

In the following sections, we will discuss the different mathematical models that are used and the types of kinetic laws that are used for describing specific reactions.

1.1 Kinetic modeling of chemical networks

In kinetic theory [90; 139], different types of kinetics are used to build up a system of ordinary differential equations. Examples include *mass action* kinetics, which was introduced by Guldberg and Waage in the 19th century [46; 61] and *Michaelis-Menten* kinetics [104], [73] which is considered to be one of the best models for the reactions that are catalyzed by enzymes. The formulas for Michaelis-Menten kinetics are derived from mass-action kinetics of enzymatic reactions. In addition, there is the *Monod equation*, which has the same form as the Michaelis-Menten equation [96; 108]. The Monod equation is used for describing the growth of microorganisms. *Hill equations* [53; 66] were introduced by A. Hill in 1910 to describe the equilibrium relationship between oxygen tension and the saturation of haemoglobin [67]. It is considered the simplest example of a mathematical representation of a regulated reaction. Most of these kinetics will be discussed later in this chapter.
In this section, we will give an overview of the kinetic modeling of biochemical reactions. Firstly, we introduce a simple structure of a chemical network, as shown in Figure 1.1.

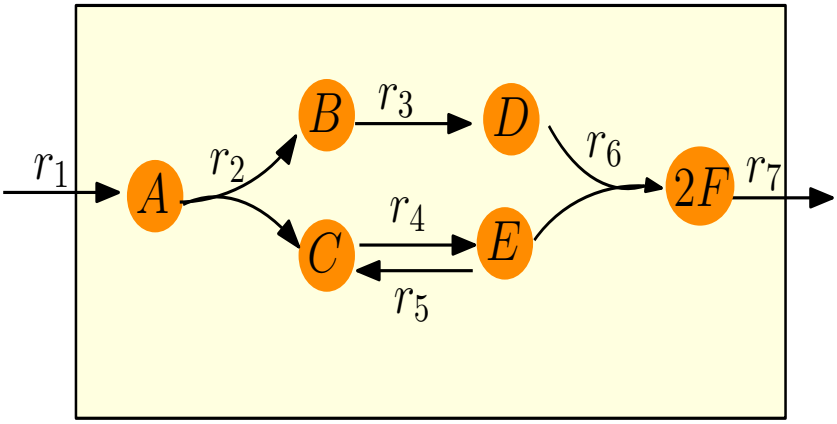


Figure 1.1: A simple structure for a metabolic network involving a set of species: A, B, C, D, E, and $2F$. It shows seven reactions as $r_1 : 0 \longrightarrow A$, $r_2 : A \longrightarrow B + C$, $r_3 : B \longrightarrow D$, $r_4 : C \longrightarrow E$, $r_5 : E \longrightarrow C$, $r_6 : D + E \longrightarrow 2F$, $r_7 : 2F \longrightarrow 0$. The metabolites on the left-hand side of arrows indicate the reactants, and the ones on the right-hand side indicate the products.

The general formula of the set of chemical reactions $r_i,\ i \in \{1, 2, ..., n_r\}$ is defined as follows

$$r_i : \sum_{j=1}^{n_m} \xi_{ji} M_j \longrightarrow \sum_{j=1}^{n_m} \eta_{ji} M_j,$$

where n_r and n_m are the number of reactions and species, respectively. The coefficients ξ_{ji} and η_{ji} are called the *stoichiometric coefficients*, they denote the proportion of substrate and product molecules involved in the reaction, see [136]. The network has seven reactions ($n_r = 7$) and six species ($n_m = 6$). For instance, the reaction r_2 will be written as

$$r_2 : A \longrightarrow B + C,$$

where the stoichiometry coefficients of this reaction are

$$\xi_{12} = 1,\ \ \xi_{22} = 0,\ \ \xi_{32} = 0,\ \ \xi_{42} = 0,\ \ \xi_{52} = 0,\ \ \xi_{62} = 0,$$

$$\eta_{12} = 0,\ \ \eta_{22} = 1,\ \ \eta_{32} = 1,\ \ \eta_{42} = 0,\ \ \eta_{52} = 0,\ \ \eta_{62} = 0,$$

where $\xi_{12} = 1$ is the proportion of the substrate M_1 involve in reaction r_2, i.e., only one molecule of A participates in the reaction r_2. We can introduce the chemical network mathematically by a *stoichiometric matrix* , which contains important information about the structure of the metabolic network [123].
A stoichiometric matrix $[S = S_{ji}]$ is a $n_m \times n_r$ matrix such that the rows correspond to the species and the columns to the reactions, with a coefficient defined by

$$S_{ji} = \eta_{ji} - \xi_{ji}.$$

The stoichiometric matrix of the network which is shown in Figure 1.1 is given as follows

$$S = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{ccccccc} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix} \end{array} \in \mathbb{R}^{6\times 7}.$$

1.1.1 Dynamic behavior of chemical networks

At the beginning of any reaction in a network, the species start with initial concentration values. A *reaction rate v* describes the change in concentration of reactants over time. In other words, a reaction rate measures how quickly reactants are changed into products [29]. A simple characterization of the rate of a chemical reaction is given by the *law of mass action* [60], which states that the rate of a chemical reaction is proportional to the product of the concentrations of the reactants [46]. The reaction rate is thus defined as follows

$$v_i(M(t)) = k_i \prod_{j=1}^{n_m} M_j^{\xi_{ji}}(t), \quad \text{for all} \quad i = 1, ..., n_r,$$

where $M(t) = [M_1(t), M_2(t), ..., M_{n_m}(t)]$, k_i is the reaction rate constant for the ith reaction. The dimensions of the rate constant depend on the number of reactants. The rate constant for a single-reactant reaction has dimensions of $time^{-1}$ and in the case of two reactants, the rate constant has dimensions of $concentration^{-1} \cdot time^{-1}$.

The modeling of chemical reaction systems is described by systems of ordinary differential equations that represent the rate of change of metabolite concentrations as follows

$$\textit{The rate of change of } [M] = S \cdot \textit{ (rate of reaction).}$$

In general, the system of differential equations associated with chemical reactions is given by

$$\frac{dM(t)}{dt} = S \cdot v(M(t)),$$

where $M = [M_1(t), M_2(t), ..., M_{n_m}(t)]$ is the vector of metabolite concentrations, m is the number of metabolites in the dynamical system, $v = [v_1(t), v_2(t), ..., v_{n_r}(t)]$ is the vector of reaction rates, n_r is the number of reactions in the network, and $S \in \mathbb{R}^{n_m \times n_r}$ is the stoichiometric matrix.

Another kinetic to describe the reaction is called Michaelis-Menten kinetics. The Michaelis-Menten kinetics are often used to account for enzymatic dynamics. Using these kinetics, the expression of the reaction rates depends on the concentration of enzyme (e) and substrate (M_j). The reaction rate v is given by

$$v = \frac{v_{max} \cdot M_j}{K_M + M_j},$$

where $v_{max} := k_{cat} \cdot e$ is the maximum uptake rate. The turnover number k_{cat} [157] is the number of times each enzyme site converts a substrate to product per unit time. The unit of turnover number is $time^{-1}$. The Michaelis-Menten constant K_M is defined as the substrate concentration at half of the maximum velocity [92]. Thus, the unit of K_M is the same as the unit of the concentration of the substrate.
The reaction rate constants of many biological systems of interest are unknown and difficult to obtain experimentally. This motivates the need for model calibration, which is an important topic in systems biology [19; 80]. Model calibration is the process of estimating parameter values from given data sets [12; 124] and it is considered a big challenge topic in mathematical modeling, see [10; 69]. We will discuss in detail the methods and techniques of model calibration in Section 1.6.

Some mathematical approaches to systems biology avoid the use of kinetic rate constants by making the assumption that the dynamical system is in a steady state [27; 136]. Examples include flux balance analysis (*FBA*) [114], [48], resource balance analysis (*RBA*) [56], models of metabolism and expression (*ME*) [93] and the dynamic flux balance analysis (dFBA) of Palsson [147] or Mahadevan [99]. In addition, there are also the dynamicME approach [158] and the dynamic enzyme-cost flux balance analysis (*deFBA*) [152]. In the following section, we discuss the approaches of *FBA* and *dFBA* which are mentioned in Palsson et al. [147].

1.2 Constraint based modeling

Metabolic networks are often analyzed in a steady-state case. However, In real systems, the cell is not in a steady-state, but is in a dynamic state. Constraint-based modeling is based on simplifying the network by overcoming the rate constants values, which is considered to be the difficult part in mathematical modeling. Several approaches are used to study and analyze metabolic networks, such as the FBA approach, which was used in several recent applications [42; 43]. The *FBA* is a mathematical approach that is used to analyze the flow of metabolite through a metabolic network. The metabolic network is constrained based on the stoichiometry of the metabolic reactions and there is no need for constant rate values. The *FBA* approach [114] is defined as

$$\begin{aligned} max \quad & c^\intercal v \\ s.t \quad & Sv = 0 \\ & lb \leq v \leq ub, \end{aligned} \tag{1.1}$$

where S is the stoichiometric matrix, v represents the vector of fluxes (reaction rates), c is a vector of weights indicating how much each reaction v contributes to the objective, and lb, ub are lower and upper bound flux rates for every reaction, respectively. The expression $c^\intercal v$ is called the objective function which will be maximized. For instance, if one aims to maximize the growth rate associated to one reaction, e.g., biomass (biomass can be defined as a composition of chemical components) production, then c will be a vector with 1 in the entry corresponding to the index of the biomass reaction and 0 in all other entries [114].

Varma and Palsson [147] further developed the FBA to optimize the biomass growth by predicting the time course of concentration of biomass and external metabolites M_{ext} in a medium and taking into account that the behavior inside the cell is at a quasi-steady state. The dynamical behavior of biomass and external metabolites concentration is described by differential equations in the following form

$$\begin{aligned} \frac{d}{dt}X(t) &= \mu(t)X(t), \quad X(t_0) = X_0 \\ \frac{d}{dt}M_{ext}(t) &= -v_{ext} \cdot X(t), \quad M_{ext}(t_0) = M_{ext,0} \end{aligned}$$

where X is the cell density (biomass), μ is the growth rate, $M_{ext,0}$ is the initial substrate concentration of external metabolites, and v_{ext} denotes the substrate uptake. Palsson et al.

(see e.g., [33; 147]) studied the dynamic profiles of cell growth and external concentrations by dividing the experimental time into small time steps Δt, specifying initial concentrations, and predicting the biomass and by-product concentrations for the next step by an iterative algorithm:

$$X_{i+1} = X_i \cdot e^{\mu_i \Delta t},$$
$$M_{ext,i+1} = M_{ext,i} + \frac{v_{ext,i}}{\mu_i} X_i (1 - e^{\mu_i \Delta t}),$$

for $i = 0, 1, ..., N$ where N is the number of time step, X_i is an approximation for $X(t_i)$, μ_i is an approximation for $\mu(t_i)$, and so on for $M_{ext,i}$, $v_{ext,i}$. The values for μ_i and $v_{ext,i}$ in each iteration step are obtained from the optimization problem solved in the FBA (1.1).

Since the FBA model does not include the regulation of gene expression, the approach leads to incorrect prediction in many cases related to the regulation of gene expression. In 2001, Covert & Palsson [33] developed a method that combines transcriptional regulatory rules with the FBA to generate time profiles and simulate the corresponding behavior under certain conditions. This approach is called the *regulatory flux balance analysis* (rFBA). In the rFBA models, the objective function (e.g., biomass growth) is optimized while regulatory constraints reduce the dimension of the solution space. The rFBA models are solved in an iterative way, alternating between flux balance analysis and applying the boolean rules. Due to the steady-state assumption of flux balance analysis, the rFBA can handle only external metabolite concentrations. In Chapter 2, our contribution is to build a kinetic model that mimics the rFBA model, to overcome the limitation of the rFBA model. We study the transformation of the rFBA model into a kinetic model, which includes concentrations for both internal and external metabolites. In addition, we perform parameter estimation techniques.

1.3 Modeling gene regulation networks

The study of metabolic networks is one of the main areas of research in understanding the behavior of metabolites and flux distribution in metabolic networks. The development of biological systems, especially in genomics, has provided detailed insight into genetic networks of many microorganisms. Metabolic-genetic models present the combination of metabolism and gene expression. In these models, different metabolite reactions are interconnected by various forms of protein interactions [89].

We now review the central dogma [34; 81] of molecular biology, which states that protein synthesis depends on two phases: transcription and translation. See Figure 1.2.

Transcription: Transcription process starts when *RNA* polymerase binds to a section of *DNA* known as a promoter and copies messenger *RNA* (*mRNA*).

Translation: The translation is the process in which ribosomes in the cytoplasm synthesize proteins after the transcription of DNA to RNA in the cell's nucleus.

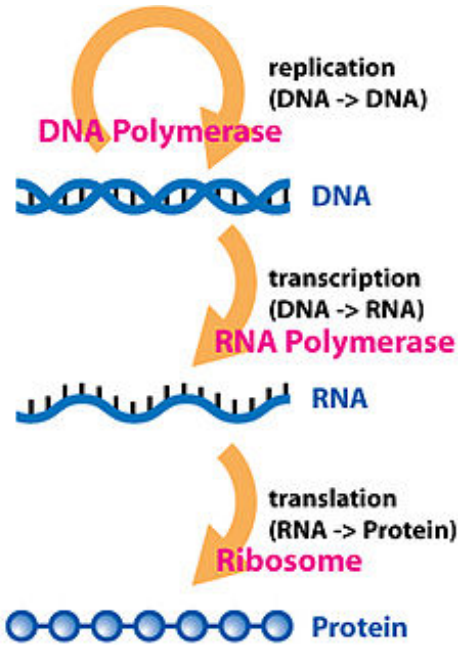
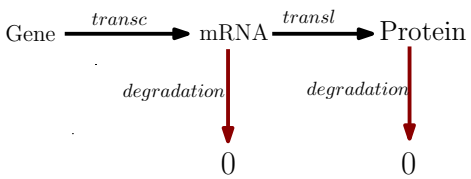


Figure 1.2: The central dogma of molecular biology describes the two-step process, transcription and translation, by which the information in genes flows into proteins: $DNA \rightarrow RNA \rightarrow protein$. Figure is taken from [35], License: $CCBY - SA3.0$.

The central dogma could be summarized [137] as follows:



1.3.1 Discrete modeling of gene regulation networks

Since it is known that the gene regulatory and metabolic processes depend on one another, a few studies addressed their interplay on a large scale model. In 2008, Samal et al. studied the transcriptional regulatory network of E. coli metabolism as the boolean network model [127]. Covert et al. [32; 33] discussed gene regulation through the core carbon metabolic network of E. coli, such that the transcriptional regulatory rules are presented by the boolean Logic equation [76; 146].

Figure 1.3 shows a simple system that involves one gene G which is transcribed by a process *trans* to produce an enzyme E. The enzyme catalyzes the reaction *rxn* which converts the species A to the species B. The presence of species B will inhibit the $trans$ process, see [33]. The logic equation which describes this procedure is

$$\text{trans= IF (G) AND NOT (B)}$$
$$\text{rxn= IF (A) AND (E)}$$

where $trans$ and *rxn* have boolean values 0 or 1.

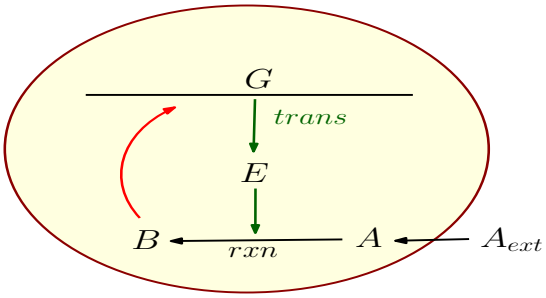


Figure 1.3: A simple regulatory circuit. Gene G is transcribed by a process trans to produce an enzyme E. Then this enzyme catalyzes a reaction *rxn* which converts substrate A into product B. Product B then represses transcription of G. The figure is inspired from [33].

1.3.2 Continuous modeling of gene regulation networks

In the previous subsection, we showed how the boolean logic functions describe the expression of regulatory genes in the discrete model. In this subsection, we present the continuous version of gene regulation. For instance, consider that the transcription of a gene is activated by activator m, see Figure 1.4. Then the gene regulation can be described as shown in the

diagram below.

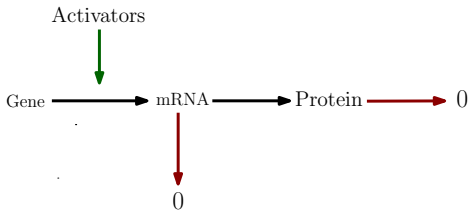


Figure 1.4: A structure illustrates how the activator induce the gene to produce mRNA then produce Protein [137].

Assume that the regulator (Activator) m controls the synthesis of *mRNA*. The dynamics of *mRNA* and proteins are described by ordinary differential equations using mass action kinetics and Hill function as

$$\frac{dmRNA}{dt} = k_1 \frac{m^h}{K^h + m^h} - d_1 \cdot mRNA,$$
$$\frac{dProtein}{dt} = k_2 \cdot mRNA - d_2 \cdot protein,$$

where k_1 is a maximal transcription rate, K is an activation coefficient with units of concentration. A Hill coefficient h is a dimensionless parameter that was derived by A. V. Hill in 1910, and reinvented by J. Wyman several decades later [15; 55], see Figure 1.5. The scalar parameters d_1 and d_2 are degradation rates for *mRNA* and *Proteins*, respectively [137].

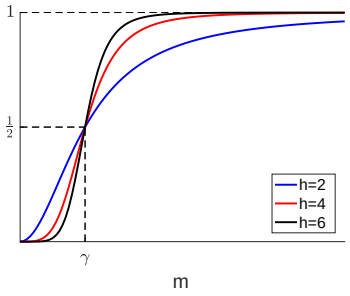


Figure 1.5: The figure shows the graph of the Hill function $\Gamma(m) = \dfrac{m^h}{\gamma^h + m^h}$ with different values of h, where h is the Hill coefficient, γ is a threshold value, and m is the substrate concentration.

So far, we have presented models that are discrete or continuous. In the following section, we discuss another type of model, the so-called *hybrid* models that combine continuous and discrete models.

1.4 Hybrid model

Hybrid systems are a collective term for dynamical systems presenting both continuous variables (described by a differential equation) and discrete states (logical functions) [59; 91]. A formal model for a hybrid system is known as a *h*ybrid automaton. In [7; 8; 65; 118], a *hybrid automaton* $\mathscr{H}$ is defined a tuple $(Y, Mode, Edge, Event, Init, Inv, Dyn, Jump)$, where

- Y is a finite set of real variables and $\dot{Y}$ represents the first derivatives of the variables;
- *Mode* is a finite set of $\{\mathsf{M}_1, \mathsf{M}_2, ..., \mathsf{M}_{n\mathsf{M}}\}$ discrete states, $n\mathsf{M}$ is a number of modes;
- *Event* is the set of event names;
- *Edge* ($\boldsymbol{E}$) is a finite set of labeled edges that represents discrete changes of one mode to another in the hybrid system that are labeled by event names;
- *Init* is a set of initial conditions of continuous variables Y when the hybrid system starts from mode M, *Inv* is an invariant condition which constrains the possible valuations of the continuous variables Y when the control of the hybrid system is in mode M, and *Dyn* is a set of differential equations that specifies the dynamics of variables in every mode;
- *Jump* defines the possible discrete actions of $\mathscr{H}$ and specifies the change of the variable values after transition $\boldsymbol{e} \in \boldsymbol{E}$ has taken place.

To clarify the hybrid system and its components, we give an example of a simple metabolic-regulatory network for describing a hybrid system.

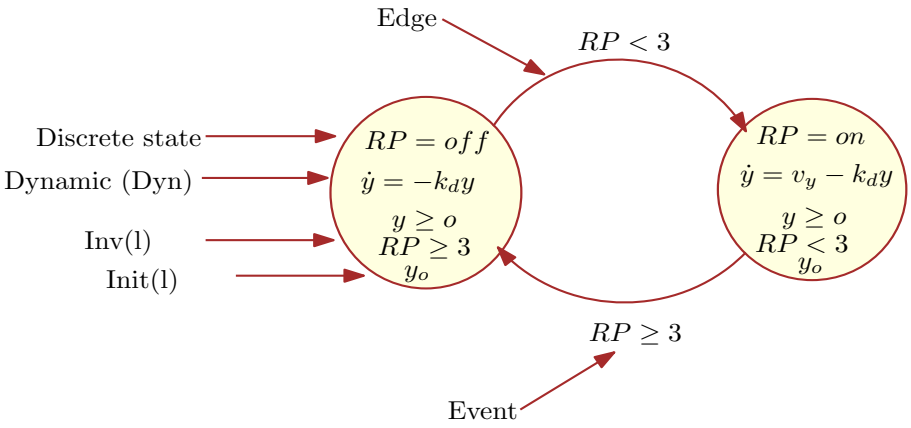


Figure 1.6: Hybrid automaton model of a metabolic-regulatory network. Assume a regulatory protein (*RP*) that has binary values $\{0,1\}$. In the on-mode, the rate change of protein y is described by the protein synthesis rate v_y and degradation term $k_d y$ where k_d is the degradation constant. In the off-mode, the rate change of protein y is described by degradation term $k_d y$ only. The transition from on-mode to off-mode depends on the value of *RP*, which is called an *event*.

Hybrid modeling topics are an area of active research; see [117; 140; 141] for some overviews and surveys. Some of this research focuses on integrative and systems biology [98; 106] and complexity in biology [31]. In addition, hybrid models have been applied in cancer modeling [102; 128; 154], and to tumor growth models [51; 122].

In Chapter 3, our contribution is to describe a continuous version of the hybrid model of a metabolic-regulatory network that is introduced in [95], because a continuous model is considered more convenient and easy to handle. In addition, different theories can be applied to continuous models, e.g., parameter estimation, model reduction, etc. In addition, the hybrid model becomes more complicated in case the number of regulatory proteins is enormous because the modeling of regulatory proteins and detecting of events becomes difficult. In the above example, we have one regulatory rule and the number of the modes is 2 according to the formula 2^n where $n = 1$ is the number of regulatory rules. It means that the increase in the number of regulatory rules leads to the increase in the number of modes and events which make the hybrid system is difficult to handle.

1.5 Model reduction methods

Kinetic models of metabolic networks are necessary for predicting and optimizing the behavior of cells in culture. However, most of these models are high-dimensional because of a large number of species and reactions, as well as kinetic rate constants [52]. Model order reduction (*MOR*) aims to reduce the computational complexity and computational time of large-scale dynamical systems, by using approximate models of much lower dimension that can produce almost the same input-output response characteristics [129]. A survey of MOR methods is given in [16].

In the history of mathematics, the approximation of a complicated function with a simpler formulation already exists a long time ago. In *1807*, *Fourier* published the idea that a function can be approximated with a few trigonometric terms [129]. In linear algebra, the initial step in the topic came from *Lanczos* (*1893-1974*). The aim is to reduce a matrix in a tridiagonal form [11].

In system biology, methods based on time scale separation are a major class of MOR methods. Such methods are based on singular perturbation analysis, which depends on the identification of quasi-steady-state conditions for fast reactions, and the derivation of non-linear models of the slow dynamics independent of the fast reaction rate expressions [104]. This approach was applied to a model of erythrocytes metabolism by assuming all reversible reactions occur at a faster time scale than the irreversible ones [130]. Also, it was applied to a metabolic-genetic network model, where metabolism reaction is very fast compared to enzyme reactions, which are considered to be slow. That means the difference between the time scales of enzymes and metabolites leads to a steady-state of metabolism reactions to adapt to the changing of the cell, see [89].

Another approach is called *proper orthogonal decomposition*, which is a MOR technique that works by projecting high-dimensional data onto a lower-dimensional space [78; 150]. The idea of the proper orthogonal decomposition is to reduce a large number of interdependent variables to a much smaller number of variables while retaining as much as possible of the variation in the original variables.

The advantage of the time scale separation method is that it preserves the dynamics of the original system. On the other hand, it is difficult to apply to a large scale system because the required conditions (Tikhonov's conditions) of this technique are sometimes difficult to satisfy. In addition, it requires the separation between slow and fast components. In contrast, the POD method works well for large scale systems but the dynamics are given for

a surrogate model that is obtained via a projection of the original model. Thus, it is difficult to predict the contribution of a specific single component.

In Chapter 4, our contribution is to discuss a model reduction by time scale separation technique for a metabolic-genetic network. By following the work of [62], we clearly described the boundary layer system which is needed to show the asymptotic stability of the fast variables. We apply the time scale separation technique to different examples of a metabolic-genetic network. For instance, systems that describe a scenario of unimolecular and bimolecular reactions. While, in Chapter 5, our contribution is to discuss the *Proper Orthogonal Decomposition* (POD) method for model reduction and to apply that technique to different biological systems from the BioModels database. In addition, we compare the time cost of the simulation of the full and reduced models. Also, we predict the behavior of the metabolic-genetic network for different scenarios with the same reduced order model.

1.6 Model fitting

This section aims at giving an overview of statistical methods used for parameter estimation or model fitting on which the "Data2Dynamics" toolbox based. We will apply this toolbox on a case study in Chapter 2.

Mathematical modeling in systems biology may require solving a *forward problem* or an *inverse problem* [159]. The forward problem (inputs to outputs) in the mathematical model means that given a model with known inputs and parameter values, compute the system state (output), see Figure 1.7. In contrast, model calibration (outputs to inputs) uses the measured system states and other available information to find the unknown model inputs [115]. This procedure is called the inverse problem of the model prediction [13; 145].

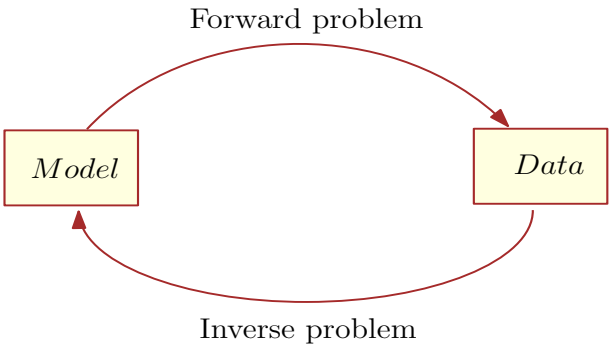


Figure 1.7: A simple definition of *forward* and *inverse* problems.

Parameter estimation is defined as the process of using observations from a dynamical system to develop mathematical models that represent the system characteristics properly. More precisely, the main goal of parameter estimation is to determine the parameter values for a model such that simulated data from this model best match experimental observations, see [107; 144].

With advances in computing, algorithms and numerical tools have been developed for simulating complicated chemical and biological processes. These developments concerning optimization and modeling focus on methods, and algorithms of computational optimization [121] such as least square regression [6; 18], maximum likelihood estimation [45], Gauss-Newton Method [39; 85; 113], Markov Chain Monte Carlo Methods (MCMC) [22; 54], etc. These methods are used for estimating unknown parameters of mathematical models.

1.6.1 Some basic concepts of statistics

Here, we will introduce some definitions from statistics that are used in the following sections. First, we briefly present the concept of a random variable. Subsequently, we present some basic concepts, e.g., a probability density function, likelihood, and maximum likelihood estimation.

A *random variable* Z is a variable whose values are numerical outcomes of a random phenomenon. In other words, it is a numerical quantity whose value depends on chance, [156]. There are two types of random variables, discrete and continuous. A discrete random variable has values that form a discrete data set, like the number of cars on the street. A continuous random variable has values that form a continuous data set, like the height of students in a class.

A *probability distribution Pr* is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence [156].

A *probability density function* of a continuous random variable Z is a function $P(z)$ such that for any two numbers a and b

$$Pr(a \leq Z \leq b) = \int_a^b P(z)dz,$$

which means that the probability of Z in the interval $[a,b]$ is the area under the graph of the density function [151], see Figure 1.8. In addition, $P(z)$ should satisfy these properties:

1. $P(z) \geq 0$ for all $z \in [a,b]$,
2. $\int_{-\infty}^{\infty} P(z)dz = 1$.

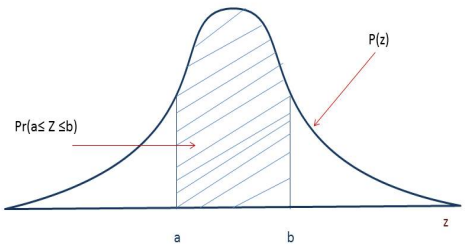


Figure 1.8: Probability density function

There are different types of probability distributions. A commonly used probability distribution is the normal distribution, which has the probability density function

$$P(z) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\ \ exp\ (\frac{-(z-\hat{\mu})^2}{2\,\sigma^2}),$$

where $\hat{\mu}$ is the mean and σ^2 is the variance. The normal distribution with mean $\hat{\mu}$ and variance σ^2 is often denoted by $N(\hat{\mu},\sigma^2)$ [4; 116].

The probability density function $P(z)$ of the normal distribution is characterized by the parameters $\hat{\mu}$ and the standard deviation σ. Thus, we can write the probability density function as $P(z,\hat{\mu},\sigma)$. All the parameters can be written as a single vector of parameters $\theta \in \mathbb{R}^{n_\theta}$, where n_θ is the number of parameters.

In practice, we often have unknown parameters θ, and the goal is to estimate θ values from some given data z. Thus, we can estimate the unknown parameters by maximizing the function

$$L(z|\theta) = P(z|\theta),$$

for the given observations z, where $L(z|\theta)$ is called the *likelihood function* [17]. This approach is called the *maximum likelihood estimation* [44; 105; 125]. It is a method for estimating parameters of a model for given data by maximizing a likelihood function such that the estimate which explains the data best, will be the best estimator.

In the following section, we will discuss in detail how to estimate the unknown parameters for a given data set using maximum likelihood estimation.

1.6.2 Parameter estimation method

As we have discussed in the previous chapter, the dynamics of the species concentrations in chemical networks can be described by a system of ordinary differential equations:

$$\dot{x}(t) = f(x(t), \theta) = S \cdot v(x(t), \theta), \quad x(t_0) = x_0, \tag{1.2}$$

where $x(t) \in \mathbb{R}^{n_m}$ is the vector of concentration of n_m species and $\theta \in \mathbb{R}^{n_\theta}$ is a vector of parameters that contains all unknown constants determining the dynamics, e.g., kinetic parameters and Hill coefficients. The function f can be decomposed into the stoichiometric matrix $S \in \mathbb{R}^{n_m \times n_r}$ and reaction rate $v \in \mathbb{R}^{n_r}$, where n_r is the number of reactions, v could be expressed by the rate law of mass action or the Michaelis-Menten kinetics.

In most cases, not all of the species concentrations can be measured directly. The dynamic states x are linked to measurements via observation function $g_k(x(t_i), \theta) \in \mathbb{R}$ such that

$$y_k(t_i) = g_k(x(t_i), \theta) + \hat{\varepsilon}_{ki}, \tag{1.3}$$

where $y_k(t_i)$, $i = 1, ..., d$ is the measured data, d is the number of experimental data $y_k(t_i)$ for each observable $k = 1, ..., s$ and s is the number of observable, measured at time points t_i. The measurement noise $\hat{\varepsilon}_{ki}$ is assumed to be normally distributed such that $\hat{\varepsilon}_{ki} \sim N(0, \sigma_k^2)$. Thus,

from equation (1.3), $y_k(t_i)$ is also normally distributed $y_k(t_i) \sim N(g_k(t_i), \sigma_k^2)$ with mean $g_k(t_i)$ and variance σ_k^2, see [84].

For the estimation of the parameters θ, we use maximum likelihood estimation (*MLE*). Under the assumption that the random variables $y_k(t_i)$ are independent and identically distributed, the maximum likelihood function $L(y|\theta)$ [84] is expressed as

$$L(y|\theta) = \prod_{k=1}^{s} \prod_{i=1}^{d} P(y_k(t_i)|\theta),$$

where $P(y|\theta)$ is a probability density function is defined as:

$$P(y_k(t_i)|\theta) = \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\ exp\left(\frac{-(y_k(t_i) - g_k(x(t_i),\theta))^2}{2\,\sigma_k^2}\right).$$

The parameters can be estimated numerically by

$$\hat{\theta} = arg\ \max_{\theta}\ L(y|\theta).$$

The log-likelihood function is

$$log\ L(y|\theta) = log\ \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}} + log\left(exp\left(\frac{-(y_k(t_i) - g_k(x(t_i),\theta))^2}{2\,\sigma_k^2}\right)\right).$$

Then we obtain

$$log\ L(y|\theta) = const + \sum_{k}^{s} \sum_{i}^{d} \frac{-(y_k(t_i) - g_k(x(t_i),\theta))^2}{2\,\sigma_k^2}.$$

If we take two times the negative log-likelihood $-2\ logL(y|\theta)$ of the data y for the parameter θ then we get the following:

$$-2\ log\ L(y|\theta) = const + \sum_{k}^{s} \sum_{i}^{d} \frac{(y_k(t_i) - g_k(x(t_i),\theta))^2}{\sigma_k^2},$$

we represent $-2\ log\ L(y|\theta)$ by $-2\ LL(y|\theta)$. The maximum likelihood estimate of the unknown parameter is given by

$$\hat{\theta} = arg \min_{\theta} \sum_{k}^{s} \sum_{i}^{d} (y_k(t_i) - g_k(x(t_i), \theta))^2 / \sigma_k^2,$$

where $-2LL$ is called the Chi-square χ^2 [58; 155] or goodness of fit statistic. The goodness of fit statistic $-2LL$ is much easier to interpret than the likelihood and more efficient to compute numerically.

Here, we clarify how to generate parameter samples to perform parameter estimation techniques. In statistical sampling, a *Latin square* is a square grid containing sample positions, such that there is only one sample in each column and each row. It was inspired by Leonhard Euler (1707–1783) [28; 153]. *Latin hypercube sampling* (LHS) is a generalization of the Latin square-based method to an arbitrary number of dimensions and was introduced by McKay in [64; 103]. It is a method for generating fewer samples and is used to decrease computational complexity. Latin hypercube sampling is used to guarantee that each parameter estimation run starts in a different place in the high-dimensional parameter space.

1.7 Thesis Structure

This thesis focuses on mathematical modeling of metabolic-genetic network and model order reduction techniques for a kinetic model of different biological systems.
We introduce in Chapter 2, a kinetic model that mimics the regulatory flux balance analysis of a metabolic-genetic network. Our kinetic model is built to overcome the limitation of the rFBA model. Due to the steady-state assumption of flux balance analysis the rFBA model can handle only external metabolite concentrations, whereas our kinetic model includes concentrations for both internal and external metabolites.
We then introduce in Chapter 3 a continuous model that mimics the hybrid system of a metabolic regulatory network. The hybrid model becomes complicated if the number of regulatory proteins is large, since the modeling of regulatory proteins and detecting of events becomes difficult. In other words, the increase in the number of regulatory rules leads to an increase in the number of modes and events exponentially, which makes the hybrid model is difficult to handle. Our continuous model is introduced to overcome the difficulty of the hybrid model. We show that by using our continuous model, we can obtain the same results as the results produced by the hybrid model for a metabolic regulatory network.
We continue in Chapter 4 with a model reduction method using the time scale separation

technique. We clearly explain the required conditions that are needed to apply the technique. We apply the technique to different examples of a metabolic-genetic network. We check numerically that the required conditions are satisfied.
In Chapter 5, we discuss model order reduction using the proper orthogonal decomposition method. We apply the method to different biological systems from the BioModels database. In addition, we compute the reduced order model for different initial conditions of the dynamical system. Finally, we end with some concluding results of the thesis in Chapter 6.

Chapter 2

Kinetic modeling of metabolic-genetic networks

In this chapter, we build a kinetic model that mimics the regulatory flux balance analysis (rFBA) model that is introduced in [33]. Our kinetic model is designed to overcome the limitation of the rFBA model. Due to the steady-state assumption of flux balance analysis, the rFBA model can handle only external metabolite concentrations, whereas our kinetic model includes concentrations for both internal and external metabolites. Another difference between the models is that the rFBA framework aims at an optimization over discrete time steps in order to approximate the dynamics of external metabolites from stoichiometric constraints. By contrast, the kinetic model gives a full picture of the dynamical system by studying the dynamics of every component over a continuous time interval. The kinetic model requires more parameters, while rFBA model uses only a few parameters. Due to the large number of parameters, the kinetic model relies heavily on parameter estimation techniques. We perform the parameter estimation technique using the D2D toolbox [119; 120] with two different data sets for the diauxic switch and aerobic/anaerobic-diauxie scenarios. Then we validate the kinetic model using a different data set for another scenario. In addition, we discuss different kinetic models for the regulatory proteins in order to show diauxic growth behavior.

2.1 Regulatory flux balance analysis (rFBA)

Integrated modeling of metabolism and gene regulation continues to be a challenging problem in computational system biology. In [33], Covert & Palsson developed a method that combines transcriptional regulatory rules with the FBA approach to generate time profiles

and simulate the corresponding behavior under certain conditions. This approach is the *regulatory flux balance analysis* (rFBA). The rFBA is a dynamic modeling approach that shows the effects of transcriptional regulation on metabolism. In the rFBA model, the objective function (e.g., Biomass growth) is optimized while regulatory constraints reduce the dimension of the solution space. After the constraints have been applied, the original solution may remain in the smaller solution space. The transcriptional regulatory rules can be represented by the boolean functions [127]. In this section, we will briefly introduce the main concepts of the regulatory flux balance analysis. The general form of the rFBA model is defined by

$$\begin{aligned} &max \qquad c^{\mathsf{T}} v \\ &s.t \qquad Sv = 0, \\ &\qquad g_i(\mathfrak{t}) = f_i(M_{ext}(\mathfrak{t}), v_{int}(\mathfrak{t})),\ i = 1,2,...,n_r, \\ &v_{min,i} \cdot g_i(\mathfrak{t}) \leq v_i \leq v_{max,i} \cdot g_i(\mathfrak{t}), \end{aligned} \tag{2.1}$$

for discrete times $\mathfrak{t}$, where n_r is a reaction number and $v_{min,i}$, $v_{max,i}$ are lower and upper bound flux rates for every reaction, respectively. In our study, the objective function $c^{\mathsf{T}} v$ is the growth rate. In (2.1) the regulation function $g_i(\mathfrak{t})$ depends on the presence or the absence of several important factors, e.g., external metabolites M_{ext} and internal fluxes v_{int}. The function $g_i = 1$ means that the reaction is constrained like in FBA, while $g_i = 0$ implies that the reaction flux is zero (i.e., $v_i = 0$). By using the algorithm of Palsson in [147] as denoted in Chapter 1, we obtain the concentration of Biomass and external metabolites in the discrete time steps.

The COBRA toolbox [63] can perform the optimization over the discrete-time step of the rFBA model after providing the initial condition values and several parameters, e.g., $v_{min,i}$, $v_{max,i}$. It generates the value of Biomass and external metabolite concentrations for every time step. This is shown in section 2.4. In the following section, we will present the object of study, the metabolic-genetic network, from the perspective of rFBA and kinetic models.

2.2 Metabolic-genetic network

The most discussed microorganism in the current literature is the bacterium Escherichia coli (*E. coli*) [24], which is often chosen as a “model” organism. This is because of its central carbon metabolism. Many mathematical models have been developed to study the

metabolism of E. coli, including the biochemical reactions that are involved in its growth and cell division [9; 41; 94; 100; 101]. Therefore, mathematical models of cellular systems of E. coli that can describe the interactions between metabolic networks and gene expression are important for understanding biological phenomena [36; 82].

In this section, we will study mimicking core metabolic network of E. coli introduced in [33]. We will present the reaction rates and the regulatory proteins of a simplified metabolic network as depicted in Figure 2.1. The network contains 20 reactions and 19 metabolites: 7 external metabolites $(C_1, C_2, O_{ext}, F_{ext}, H_{ext}, E_{ext}, D_{ext})$ and 11 internal metabolites $(A, B, C, F, H, E, D, G, ATP, NADH, O_2)$. We use a symbol X to represent the Biomass in the mathematical model.

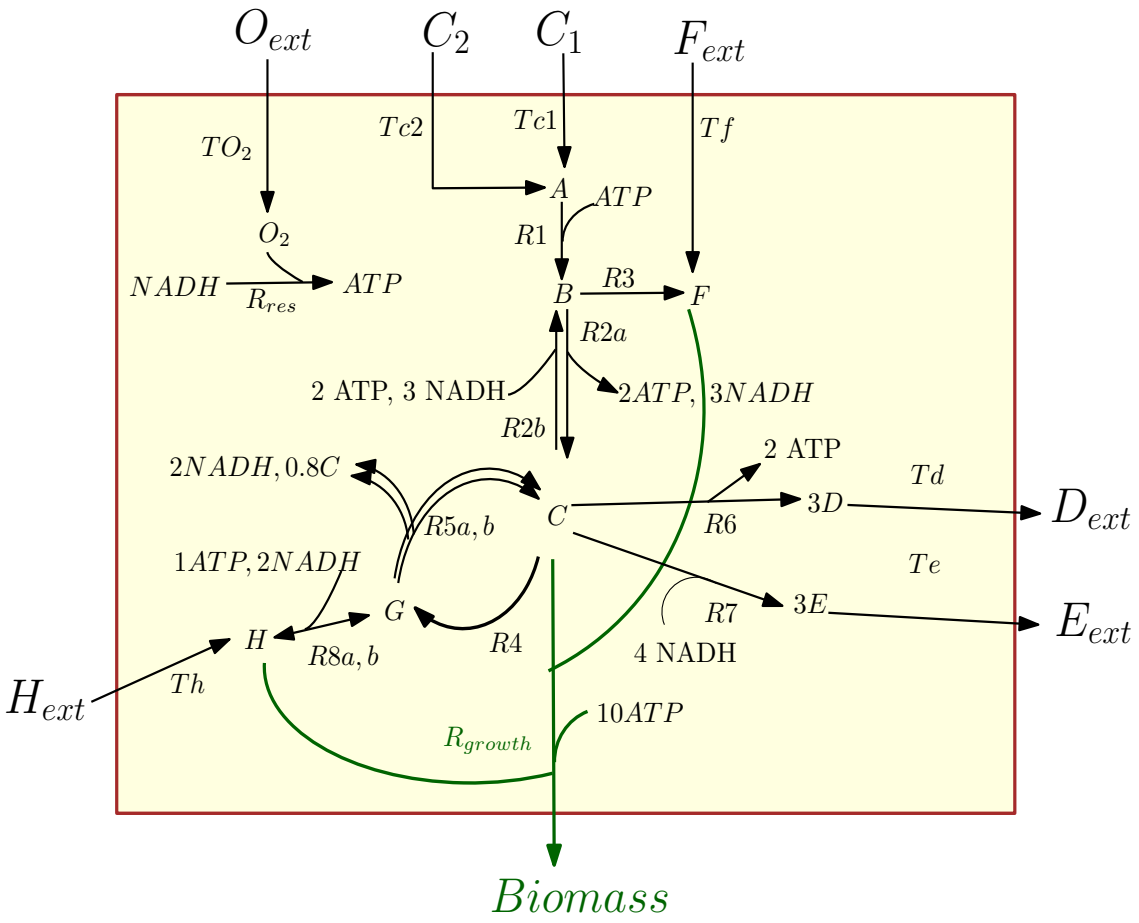


Figure 2.1: A simplified core carbon metabolic network.

In Table 2.1, we describe all chemical reactions of metabolism and regulatory proteins corresponding to every reaction.

Reaction	Reaction's name	Regulation
Exchange reactions		
$C_1 \rightarrow$ A	$Tc1$	
$C_2 \rightarrow$ A	$Tc2$	IF Not RPc_1
$F_{ext} \rightarrow$ F	Tf	
$O_{ext} \rightarrow O_2$	To_2	
D $\rightarrow D_{ext}$	Td	
E $\rightarrow E_{ext}$	Te	
$H_{ext} \rightarrow$ H	Th	
Intracellular metabolite reactions		
A + ATP $\rightarrow$ B	R1	
B $\rightarrow$ C + 2 ATP+ NADH	R2a	IF Not RPb
C + 2 ATP + NADH $\rightarrow$ B	R2b	
B $\rightarrow$ F	R3	
C $\rightarrow$ G	R4	
G $\rightarrow$ 0.8 C + 2 NADH	R5a	IF Not RPO_2
G $\rightarrow$ 0.8 C + 2 NADH	R5b	IF RPO_2
C $\rightarrow$ 3 D + 2 ATP	R6	
C + 4 NADH $\rightarrow$ 3 E	R7	IF Not RPb
G + ATP + 2 NADH $\rightarrow$ H	R8a	IF Not RPh
H $\rightarrow$ G + ATP + 2 NADH	R8b	
$NADH + O_2 \rightarrow$ ATP	Rres	IF Not RPO_2
Growth Processes		
C+ F+ H + 10 ATP $\rightarrow$ Biomass	Growth	

Table 2.1: Reactions and regulatory rules for the simplified metabolic network.

The transcriptional regulatory rules are used with *FBA* to generate time profiles of cell growth, substrate utilization, and by-product secretion for organisms. The regulatory rules are presented by Thomas' boolean formalism [146].

The four regulatory rules of the network are summarized in the following table:

Regulatory proteins		Transcriptional regulation	
RPc_1	IF (C_1)	tTc_2	IF NOT (RPc_1)
RPh	IF ($v_{Th} > 0$)	$tR8a$	IF NOT (RPh)
RPb	IF ($v_{R2b} > 0$)	$tR2a$, $tR7$	IF NOT (RPb)
RPO_2	IF Not (O_{ext})	$tRres$, $tR5a$	IF NOT (RPO_2)

Table 2.2: Boolean logic formula of regulatory protein rules.

The first logic formula in the above table illustrates that the presence of the metabolite C_1 will activate the regulatory protein RPc_1; the activation of RPc_1 (on) inhibits the transcriptional regulation tTc_2 (off). Every regulatory protein is defined by the boolean function with values 0 or 1 (off, on) in the rFBA model. In the following section, we will discuss the kinetic model that describe the reactions and regulatory proteins of the network.

2.3 Kinetic modeling of metabolic-genetic network

Dynamical studying of models is the study of the relationship between components of the network that are dynamically interrelated. An ordinary differential equation (*ODEs*) model allows the qualitative simulation of changes in concentration over time. In general, a network has external (e.g., nutrients) and internal metabolites, and Biomass reactions (growth rate reaction). The growth rate equation has been studied by many researchers [38; 72; 131]. The general formula of Biomass is given by

$$\frac{dX}{dt} = \mu X,$$

where μ is the growth rate and X is the Biomass, we refer to [36] for more information. The external metabolites depend on the Biomass. It means that if there is no Biomass, there is no nutrient consumption. Therefore, we use the Biomass term in external metabolite equations. In addition, we maintain that the extracellular components are given based on the reactor volume, while intracellular components are given on the basis of the cellular dry weight.

In summary, the differential equations of a metabolic network can be given as follows:

$$\begin{aligned} \frac{dM_{ext}}{dt} &= S_{ext} v X, \\ \frac{dM_{int}}{dt} &= S_{int} v, \\ \frac{dX}{dt} &= \mu X, \quad \text{with} \quad \mu = \sum_{i=1}^{n_b} \omega_i v_i, \end{aligned} \tag{2.2}$$

where M_{ext} and M_{int} are the vectors of external and internal metabolite concentrations, respectively, $S = \begin{bmatrix} S_{ext} \\ S_{int} \end{bmatrix}$ is the stoichiometric matrix split into external and internal parts, v is the vector of reaction flux, and X is the Biomass. The factor μ describes the growth rate and is composed of the reaction fluxes v_i, for all reactions $i = 1, \ldots n_b$, which produces Biomass, multiplied with the corresponding yield coefficients ω_i. The yield coefficients expressed as the mass of cells or product formed per unit mass of substrate consumed, the standard notion is *gDW* produced/*g* substrate used. Notice that in our model $n_b = 1$.

Regulatory rules in the kinetic model

As we discussed in Chapter 1, the regulatory proteins can be described by the boolean function with values 0 or 1 in a discrete model or by the Hill function with a range between $[0, 1]$ in a continuous model, see section 1.3. In our kinetic model, we use the Hill function to represent the regulatory proteins. Using the Hill function causes an increase in the number of parameters of the kinetic model as (thresholds and Hill coefficients). As an example, we discuss one of the regulatory proteins of the network, for the other regulatory proteins the formulation is similar.

1- *Carbon catabolite repression*

In this case, C_1 in the extracellular medium is considered the prioritized carbon source which activates the regulatory protein RPc_1. The activated RPc_1 inhibits the transcription of the gene which encodes a protein for transport of C_2 into the cell via tTc_2. The regulatory proteins RPc_1 and tTc_2 are expressed using the Hill functions as follows

$$\begin{aligned} RPc_1 &= F_h^+(C_1, \zeta) = \frac{C_1^h}{\zeta^h + C_1^h}, \\ tTc_2 &= F_h^-(C_1, \zeta) = \frac{\zeta^h}{\zeta^h + C_1^h}, \end{aligned}$$

where ζ is a threshold and h is the Hill coefficient. The sign of $+$ and $-$ in the function F indicates the activation and inhibition Hill function, respectively. The curve of the inhibition Hill function is shown in Figure 2.2. The rest of the regulatory proteins of the network are represented in the same way.

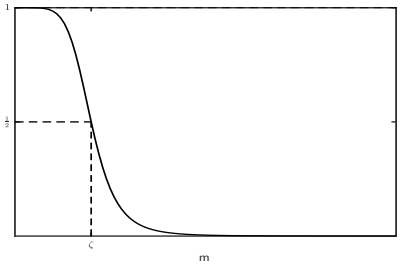


Figure 2.2: The inhibition Hill function $\hat{\Gamma}(m) = \dfrac{\zeta^h}{\zeta^h + m^h}$, where m is a substrate, $h = 2$, and $\zeta = 0.2$.

These Hill kinetics yields the following relations for the regulatory rules defined in Table 2.2:

$$
\begin{aligned}
tTc2 &= \frac{\zeta^h}{\zeta^h + C_1^h}, \\
tR2a = tR7 &= \frac{B^h}{\gamma^h + B^h}, \\
tR5a = tRres &= \frac{O_{ext}^h}{\beta^h + O_{ext}^h}, \\
tR5b &= \frac{\beta^h}{\beta^h + O_{ext}^h}, \\
tR8a &= \frac{\alpha^h}{\alpha^h + H^h},
\end{aligned}
$$

where ζ,γ, β, and α are the thresholds and B, H are the internal metabolites. In our study, we fix the value of the Hill coefficient to $h = 2$.

There are many kinetic rates that can be used to describe the behavior of metabolic networks. In our kinetic model, we use the Michaelis-Menten kinetics for the external metabolites because their reactions are considered enzymatic reactions such that the enzyme active sites become saturated. We use mass action kinetics for internal metabolites for the sake of simplicity and to reduce the number of kinetic parameters. Altogether, the ordinary differential equations to model the dynamical behavior of the simplified metabolic-genetic network in Figure 2.1 are given by

$$
\begin{aligned}
\frac{dC_1}{dt} &= -v_{Tc1} \cdot X \\
\frac{dC_2}{dt} &= -v_{Tc2} \cdot tTc_2 \cdot X \\
\frac{dF_{ext}}{dt} &= -v_{Tf} \cdot X \\
\frac{dO_{ext}}{dt} &= -v_{To2} \cdot X \\
\frac{dD_{ext}}{dt} &= v_{Td} \cdot X \\
\frac{dE_{ext}}{dt} &= v_{Te} \cdot X \\
\frac{dH_{ext}}{dt} &= -v_{Th} \cdot X \\
\frac{dA}{dt} &= v_{Tc1} + v_{Tc2} \cdot tTc_2 - v_{R1} \\
\frac{dB}{dt} &= v_{R1} - v_{R2a} \cdot tR2a - v_{R3} + v_{R2b} \\
\frac{dC}{dt} &= v_{R2a} \cdot tR2a - v_{R2b} - v_{R4} - v_{R7} \cdot tR7 - v_{R6} + 0.8 \cdot v_{R5a} \cdot tR5a + 0.8 \cdot v_{R5b} \cdot tR5b - v_{bio} \\
\frac{dD}{dt} &= 3 \cdot v_{R6} - v_{Td} \\
\frac{dE}{dt} &= 3 \cdot v_{R7} \cdot tR7 - v_{Te} \\
\frac{dF}{dt} &= v_{Tf} + v_{R3} - v_{bio} \\
\frac{dG}{dt} &= v_{R4} - v_{R8a} \cdot tR8a + v_{R8b} - v_{R5a} \cdot tR5a - v_{R5b} \cdot tR5b \\
\frac{dH}{dt} &= v_{Th} + v_{R8a} \cdot tR8a - v_{R8b} - v_{bio} \\
\frac{dATP}{dt} &= -v_{R1} + 2 \cdot v_{R2a} \cdot tR2a - 2 \cdot v_{R2b} + 2 \cdot v_{R6} + v_{R8b} - v_{R8a} \cdot tR8a + v_{Rres} \cdot tRres - 10 \cdot v_{bio} \\
\frac{dNADH}{dt} &= 2 \cdot v_{R2a} \cdot tR2a - 2 \cdot v_{R2b} - 4 \cdot v_{R7} \cdot tR7 + 2 \cdot v_{R5a} \cdot tR5a + 2 \cdot v_{R5b} \cdot tR5b - \\
&\quad 2 \cdot v_{R8a} \cdot tR8a + 2 \cdot v_{R8b} - v_{Rres} \cdot tRres \\
\frac{dO_2}{dt} &= v_{To2} - v_{Rres} \cdot tRres \\
\frac{dX}{dt} &= \omega \cdot v_{bio} \cdot X
\end{aligned}
\tag{2.3}
$$

We use Michaelis-Menten kinetics to model the reaction rates for external metabolites yielding:

$$
\begin{aligned}
v_{Tc1} &= v_{maxc1} \cdot C_1/(k_{M1} + C_1) \\
v_{Tc2} &= v_{maxc2} \cdot C_2/(k_{M2} + C_2) \\
v_{Tf} &= v_{maxf} \cdot F_{ext}/(k_{M3} + F_{ext}) \\
v_{To2} &= v_{maxo2} \cdot O_{ext}/(k_{M4} + O_{ext}) \\
v_{Td} &= v_{maxd} \cdot D/(k_{M5} + D) \\
v_{Te} &= v_{maxe} \cdot E/(k_{M6} + E) \\
v_{Th} &= v_{maxh} \cdot H_{ext}/(k_{M7} + H_{ext})
\end{aligned}
$$

and mass action kinetics to model the reaction rates for internal metabolites yielding:

$$
\begin{aligned}
v_{bio} &= k_8 \cdot F \cdot H \cdot C \cdot ATP \\
v_{R1} &= k_9 \cdot A \cdot ATP \\
v_{R2a} &= k_{10} \cdot B \\
v_{R2b} &= k_{11} \cdot ATP \cdot NADH \cdot C \\
v_{R3} &= k_{12} \cdot B \\
v_{R4} &= k_{13} \cdot C \\
v_{R5a} &= k_{14} \cdot G \\
v_{R5b} &= k_{15} \cdot G \\
v_{R6} &= k_{16} \cdot C \\
v_{R7} &= k_{17} \cdot C \cdot NADH \\
v_{R8a} &= k_{18} \cdot G \cdot ATP \cdot NADH \\
v_{R8b} &= k_{19} \cdot H \\
v_{Rres} &= k_{20} \cdot O_2 \cdot NADH
\end{aligned}
$$

The kinetic model depends on the parameters $k_{M1}, \ldots, k_{M7}$ and $k_8, \ldots, k_{20}$, as well as the thresholds for Hill equations α, β, γ, ζ.

The unit of quantities in the kinetic model should match well. Now, we will clarify the units that are used in our kinetic model. The external metabolite concentration M_{ext} has unit $[mM]$ or $[mmol/L]$ because it is based on the volume of the medium. The internal metabolite concentration M_{int} in unit $[mmol/gDW]$, it is based on cell dry weight DW. The unit of Biomass X is g/L, yield coefficient ω unit is $[gDW/g]$, and the growth rate μ unit is $1/hr$. The unit of v_{max} is $[mmol/(gDW.hr)]$, the turnover constant unit is $1/hr$, and K_M and thresholds ζ, γ, β, and γ have the same units of the corresponding metabolites in every formula of them.

2.4 Model fitting

Notice that our goal is to build a kinetic model that mimics the rFBA model. Since our kinetic model (2.2) contains unknown parameters, we aim to obtain the parameter values. Finding the value of the parameters from the literature will be difficult because the network 2.1 is mimicking core metabolism and it is difficult to know exactly what the metabolites A, B, ..., etc., represent in the real system. Therefore, we aim to perform the parameter estimation technique using the D2D toolbox, that is based on the maximum likelihood estimation which is mentioned in Chapter 1.
In order to perform the parameter estimation technique, we need a dynamical model and data sets. The data sets we use are the simulation results of the rFBA model. We perform the simulation of the rFBA model using the COBRA toolbox [63] by creating a network model in Matlab using the reaction rates and corresponding regulatory rules in Table 2.1. Additionally, we provide the initial conditions of external metabolites and Biomass, and we define the time interval and time step size. By running our model file using the COBRA toolbox, we obtain a data set for the Biomass and concentration of external metabolites over discrete-time steps.

We assume that the kinetic model equations (2.2) are given by

$$\left.\begin{array}{ll} \frac{dM_{ext}}{dt} & = S_{ext} vX \\ \frac{dM_{int}}{dt} & = S_{int} v \\ \frac{dX}{dt} & = \mu X \end{array}\right\} \quad \Longrightarrow \quad \frac{dx(t)}{dt} = f(x(t), \theta), \tag{2.4}$$

where $x(t)$ denotes the concentration vector and $\theta \in \mathbb{R}^{n_\theta}$ is a vector of parameters i.e. the kinetic parameters and the thresholds in the Hill function. The function f is given by the rate equations, like the mass action or the Michaelis-Menten rate laws.

In order to calibrate the kinetic model, defined by equation (2.4) to a given data set (simulation result of rFBA), the dynamic states x are linked to measurements via

$$y(t_i) = g(x(t_i), \theta) + \varepsilon(t_i),$$

where y is the simulation result of rFBA of external metabolites and Biomass at some time points t_i; g is the observation function involving the parameter θ. In our model, the observation function is the dynamical state $x(t)$ of external metabolites and Biomass.
We aim to fit our model to two different data sets for the diauxic switch and aerobic/anaerobic-diauxie scenarios of the cell, which are mentioned in [33]. The data sets of the different scenarios of the rFBA model are obtained using the Cobra toolbox [63]. Using the D2D toolbox [119; 120] we can fit the kinetic model to the data sets.
In the following, we discuss the two different scenarios for the metabolic-genetic network model that have been used for parameter estimation.

Diauxic switch scenario

At first, we consider the diauxic switch scenario [14; 132]. In this scenario, there are two sources of carbon that are introduced to the cell in culture. During the first phase, cells prefer to metabolize the sugar (e.g., C_1) where the cells can grow faster. Only after the first sugar has been exhausted, the cells switch to the second source of carbon, see [26; 83; 110; 111]. Most of the initial concentration for the external metabolites are taken from [33] in all the scenarios discussed here. The initial concentrations of extracellular metabolites and Biomass are $C_1 = 10$, $C_2 = 10$, $O_{ext} = 50$, $D_{ext} = 0$, $E_{ext} = 0$, $H_{ext} = 0$, $F_{ext} = 0$, all in unit [mM], and $X = 0.003$ g/L. We use $v_{maxc1} = v_{maxc2} = 10.5$, $v_{maxf} = v_{maxh} = 5$, $v_{maxd} = v_{maxe} = 12$ and $v_{maxo2} = 15$, all in unit $[mmol/(gDWhr)]$, cf. [33]. We simulate the rFBA model using the COBRA toolbox [63] over discrete experimental time $[0,5]$ with 20 step number and step size 0.25. Using the COBRA toolbox [63; 148], we predict the time profile of external metabolites and Biomass concentration at the diauxic switch scenario, see Figure 2.3. Moreover, we assume the initial condition of most of internal metabolites have zero concentration as $A = 0$, $B = 1, C = 0$, $D = 0$, $E = 0$, $F = 0.2$, $G = 0$, $H = 0.03$, $ATP = 6$, $NADH = 5$, $O_2 = 0$, the unit of the internal metabolites is $[mmol/gDW]$. By using the D2D toolbox, we fit the kinetic model to the simulation results of the rFBA model, see Figure 2.4.

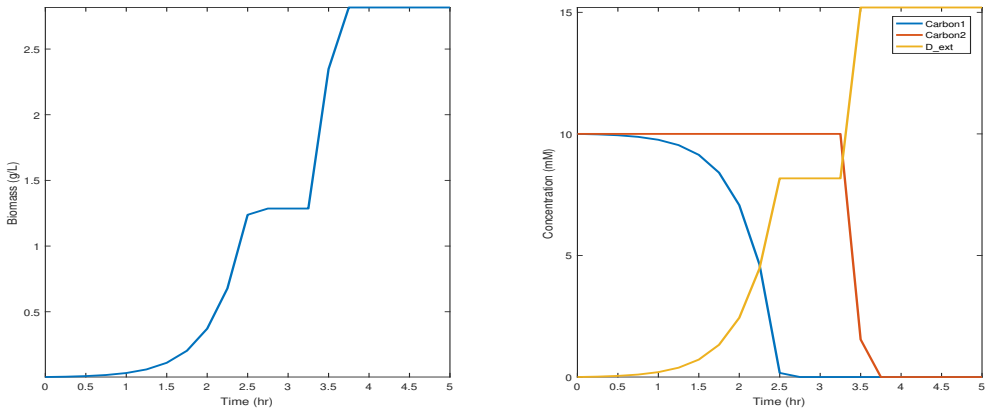
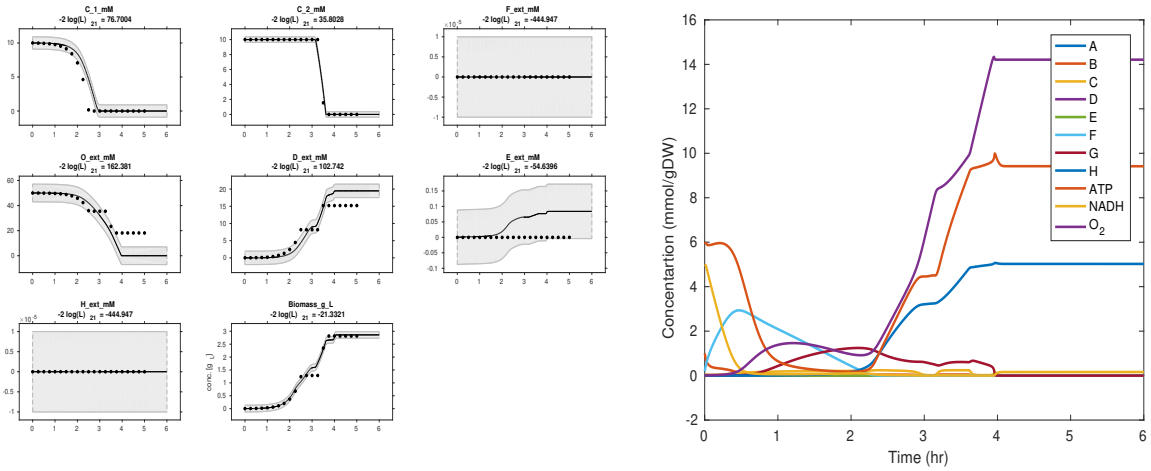


Figure 2.3: Simulation results of the rFBA model generated by the COBRA toolbox: Left figure shows the behavior of Biomass and right side figure depict the curves of the external metabolites. The initial concentrations of Biomass and extracellular metabolites are $C_1 = 10$, $C_2 = 10$, $O_{ext} = 50$, $D_{ext} = 0$, $E_{ext} = 0$, $H_{ext} = 0$, $F_{ext} = 0$, and $X = 0.003$ g/L.

The results of the model fitting have been generated by the D2D toolbox, using the functions `arInit`, `arPlot` to run and plot the data of the ODEs simulation and `arFitLHS` for parameter values generation. Resulting trajectories of model fitting are below:



(a) **(b)**

Figure 2.4: a) Model calibration of external metabolites: $(*)$ indicates the data of Biomass and external metabolites obtained by simulation of the rFBA, $(-)$ indicates the solution of the *ODEs* model. **b)** The trajectory of the internal metabolites for the diauxic-switch scenario.

In Figure 2.4a, we observe that the simulation result of the rFBA model and the solution of the kinetic model have similar qualitative behaviors for most of the external metabolites. Only the data points of the metabolite E_{ext} look different from the solution of the kinetic model but are similar in the magnitude. In Figure 2.4b, the trajectory of all the internal metabolites converges to their steady-state at a time interval of 4hr. Specifically, the trajectory of the internal metabolite O_2 increased continuously with time and reaches to the maximum value of 14 [mmol/gDW]. One possible explanation for this increasing trend could be the higher availability of the O_{ext}. In addition, the cell consumes ATP and NADH for the reactions. Then, it seems like the cell produces energy as the concentration of ATP is getting higher until the value 9 [mmol/gDW]. The trajectory of the metabolites H starts increasing until 5 [mmol/gDW], the reason could be due to the presence of O_{ext}.

Aerobic/Anaerobic-diauxie

In the aerobic/anaerobic-diauxie scenario of transcriptional regulatory modeling, there is only one source of carbon (C_2) and oxygen supplied to the culture. The oxygen is removed after two hours. The initial conditions of extracellular metabolites and Biomass are given as $C_1 = 0$, $C_2 = 10$, $O_{ext} = 2$, $D_{ext} = 0$, $E_{ext} = 0$, $H_{ext} = 0$, $F_{ext} = 0$ mM, and $X = 0.0008$ g/L. We simulate the model using the COBRA toolbox [63] by discrete experimental time $[0,5]$ to 20 step number with size 0.25. The results are given in Figure 2.5 and model fitting in Figure 2.6.

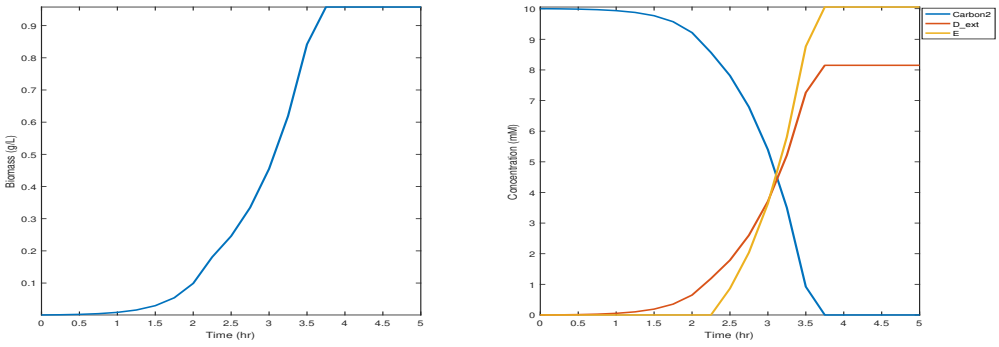
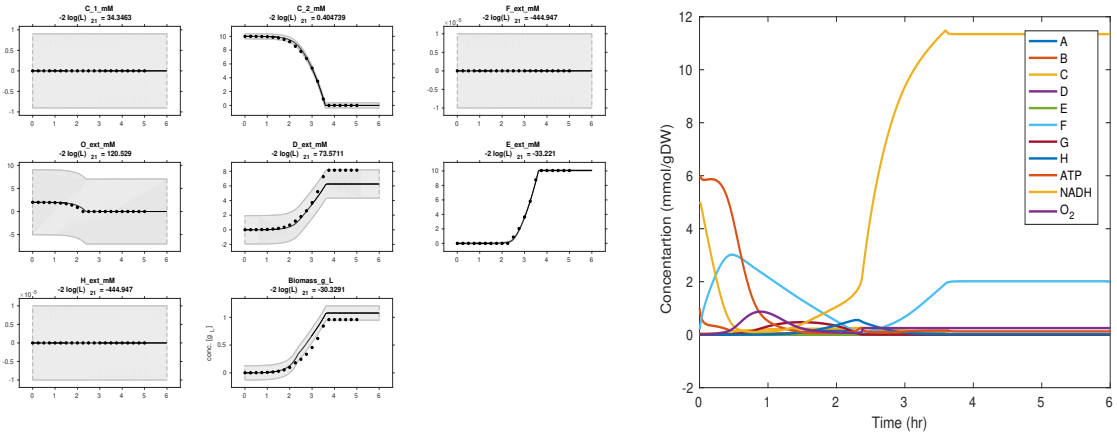


Figure 2.5: The simulation of rFBA model in the aerobic/anaerobic-diauxie scenario using COBRA toolbox. The figure shows the behavior of Biomass, C_2, D_{ext}, and E_{ext}.

From Figure 2.6, we observe that the simulation results of the rFBA model fit well with the solution of the kinetic model. Figure 2.4 shows that the cell consumes ATP and NADH for the reactions. Then, it seems like the cell produces energy as the concentration of NADH is getting higher until the value 11 [mmol/gDW]. The concentration of the metabolite F is

increasing and attains 2 [mmol/gDW]. All the metabolites, except NADH and F, almost converge to zero value.



(a) (b)

Figure 2.6: a) Model calibration of external metabolites: $(*)$ indicates the data of Biomass and external metabolites obtained by simulation of rFBA, $(-)$ indicates the solution of the *ODEs* model generated by the D2D toolbox. **b)** The trajectory of internal metabolites predicted from the solution of the *ODEs* model.

2.4.1 Model verification

We have built the kinetic model and estimated the parameter values, hence we aim to validate our kinetic model. In order to validate our kinetic model, we compared the numerical solution of our model with a different independent data set that is not used before in the parameter estimation process in Sec 2.4. The data sets is generated for a scenario that considers the growth on one source of carbon (C_2) with the presence of amino acid H_{ext} in the culture using the initial conditions $C_1 = 0$, $C_2 = 10$, $O_{ext} = 50$, $D_{ext} = 0$, $E_{ext} = 0$, $H_{ext} = 2$, $F_{ext} = 0$ mM, and $X = 0.0008$ g/L. The results depicted in Figure 2.7 show that the simulation data obtained from the rFBA model using the COBRA toolbox and simulated dynamics of our kinetic model have a qualitatively similar dynamic behavior. The results have been generated using Matlab16b by the function `ode23s` with `TolFunc=1e-6`.

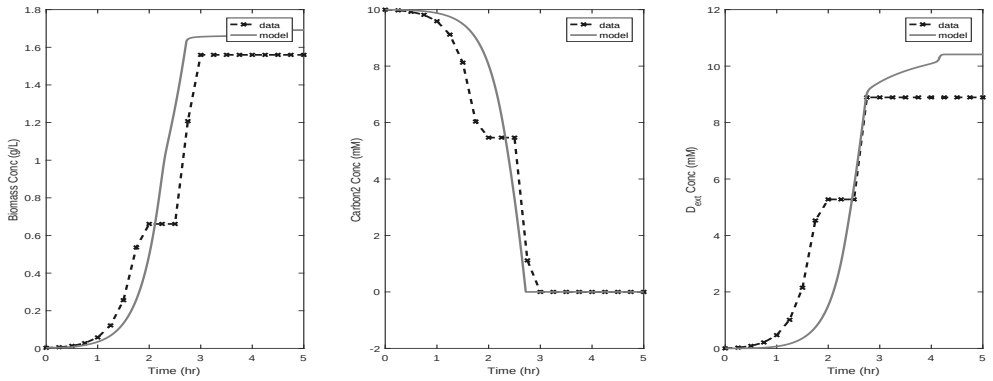


Figure 2.7: Model validation: The behaviors of the simulation of the metabolites and Biomass $(-)$ are qualitatively the same of the data$(-\times)$.

In the previous sections, we have built a kinetic model that mimics the rFBA model, performed the parameter estimation technique, and validated the kinetic model.
In the kinetic model (2.3), we expressed the regulatory protein $tR2a$ using the Hill function which depends on the internal metabolite B. In the following section, we will discuss the possibility of expressing $tR2a$ in the kinetic model as a function of the metabolite C or the reaction rate v_{R2b}. Every expression gives us different kinetic models. Our goal is to discuss the different kinetic models and study how far these models are able to generate diauxic growth behavior.

2.5 Different model structures for regulatory proteins

In this section, we will discuss different model structures for regulatory mechanisms $tR2a$ that are used to convert metabolite B into metabolite C. We discuss three cases for describing the regulatory protein $tR2a$. First, the metabolite B acts as a regulatory metabolite and activates the regulatory protein $tR2a$. Second, the metabolite C acts as a regulatory metabolite and inhibits the regulatory protein $tR2a$. Finally, the reaction rate v_{R2b} acts as a regulatory reaction rate and inhibits the regulatory protein $tR2a$. The three study cases are shown in Figure 2.8.

We are interested in the diauxic switch scenario. Thus, we will focus on the behavior of C_1, C_2, and Biomass. We will explain that every reaction scheme for the kinetic model with regulations generates the diauxic growth behavior for chosen parameter values. In general, we will use the estimated parameter values, see Appendix, Table A.1 and some of the parameter values will be tuned, as we will discuss later.

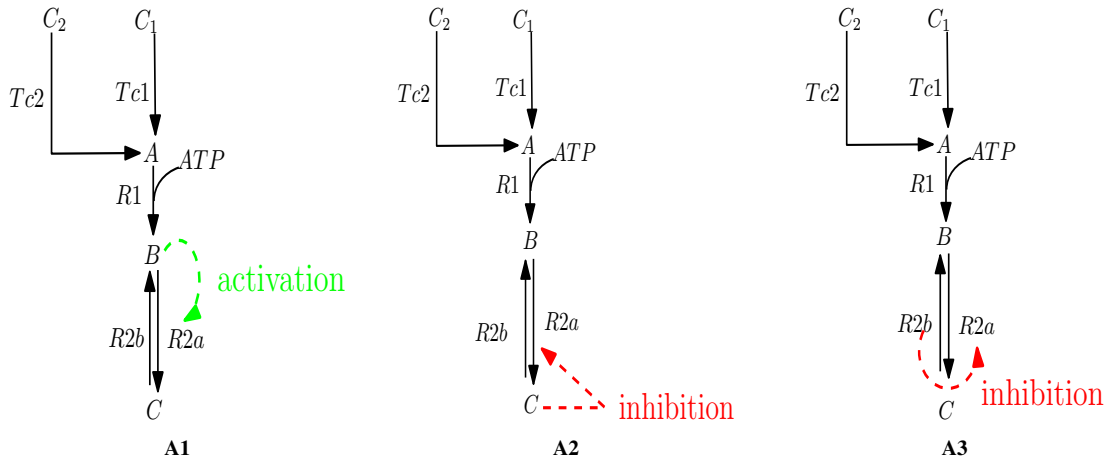


Figure 2.8: Reaction schemes for kinetic models with regulation. **A1**) The metabolite B activates the regulatory protein $tR2a$. **A2**) The presence of metabolite C inhibits $tR2a$. **A3**) In case of the value of reaction $v_{R2b} > 0$, it inhibits $tR2a$.

Case A1: the metabolite B activates the regulatory protein $tR2a$. In the kinetic model (2.3), we use the Hill function to express the relation between B and $tR2a$. Then $tR2a$ is given by

$$tR2a = \frac{B^h}{\gamma^h + B^h}.$$

Case A2: the metabolite C inhibits the regulatory protein $tR2a$. Then t$tR2a$ is given by

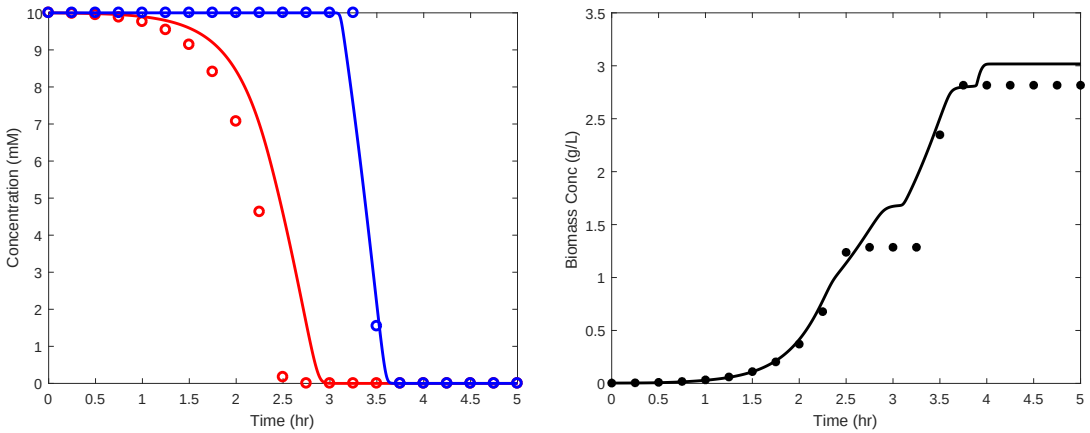
$$tR2a = \frac{\gamma^2}{\gamma^2 + C^2}.$$

In this case, we use the same parameters in Appendix, Table A.1 but we adjust the value of the threshold γ to be 30 [mmol/gDw] to obtain a good fit. In Figure 2.10 (a,b), we observe that the kinetic model generates the diauxic-switch behavior.

Case A3: the regulatory protein $tR2a$ is given by

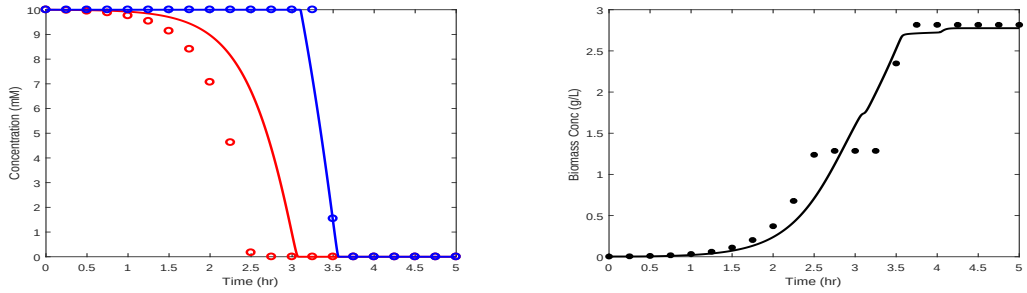
$$tR2a = \frac{\gamma^2}{\gamma^2 + v_{R2b}^2}.$$

We adjust the value of the Michaelis-Menten constants of C_1, C_2 to be $k_{m1} = k_{m2} = 0.08$ [mmol/l], the constant rate of Biomass reaction is $k_8 = 2$ [mmol³/(gdw³.hr)], and the threshold γ is 30 [mmol/gDw], see Figure 2.10 (c,d).
By solving the system of differential equations (2.3) numerically using Matlab function `ode45` and comparing with the simulation of rFBA model, we obtain the following results:

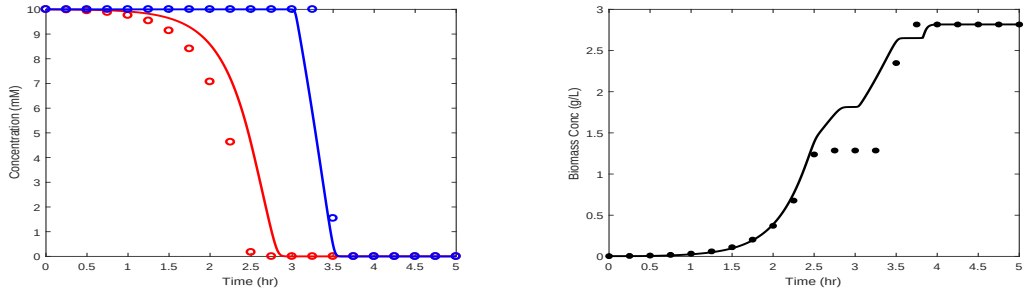


Case A1 Case A1

Figure 2.9: Comparison of the simulation of kinetic model (solid lines) with the simulation of rFBA model (circles). Substrates C_1 and C_2 are shown in red and blue, respectively and Biomass is shown in black.



(a) Case A2 (b) Case A2



(c) Case A3 (d) Case A3

Figure 2.10: Comparison of the simulation of kinetic model (solid lines) with the simulation of rFBA model (circles). Substrates C_1 and C_2 are shown in red and blue, respectively and Biomass is shown in black.

In this section, we have discussed different model structures for regulatory rule $tR2a$ that generate the diauxic growth behavior for chosen parameters.

2.6 Conclusion

In this chapter, we built a kinetic model that mimics the regulatory flux balance analysis model. While Covert and Palsson [33] have used the boolean functions to express the regulatory proteins, in our kinetic model we have used a continuous representation by Hill functions. We have presented a mathematical kinetic model of a simplified core carbon metabolic network describing the extracellular and intracellular metabolites, as well as the growth rates. We have discussed the diauxic switch and aerobic/anaerobic-diauxie scenarios. These scenarios were used to estimate the parameter values in our kinetic model. The parameter estimation process is done by using the D2D toolbox which is based mainly on the Chi-square test, log-Likelihood, and using Latin hypercube sampling for the initial guesses of the parameters. A numerical simulation using the *ODEs* model can be used to study the behavior of the metabolites inside the cell. The *ODEs* model has a great advantage because

the regulatory flux balance analysis fails in studying the dynamics of internal metabolites because it is based on a steady-state assumption. In addition, we have studied different model structures for gene regulations and show that every model generates diauxic growth behavior.

Chapter 3

A continuous version of a hybrid system

In this chapter, we introduce a continuous model that mimics the hybrid system of a metabolic-regulatory network studied in [95]. Our continuous model aimed to overcome the difficulty of the hybrid model. The hybrid model becomes complicated if the number of regulatory proteins is large since the modeling of regulatory rules and detecting of events becomes difficult. In other words, the increase in the number of regulatory rules leads to an increase in the number of modes and events by 2^n where $n = 1, 2, ...,$ is the number of regulatory rules. We show that by using our continuous model, we can obtain the same results as the results produced by the hybrid model for a metabolic regulatory network more easily. This is because we use the Hill functions to express the regulatory rules in our continuous model instead of the boolean functions in the hybrid model. In general, the main challenge in continuous models is to find the kinetic parameter values. Here, our continuous model is of great advantage because it does not need any more parameters than the hybrid model.

3.1 Hybrid automaton of metabolic-regulatory network

An integrated model of metabolism and gene regulation is considered as an example of a hybrid system. Metabolites are described by the continuous variables, while a gene regulatory mechanism is described by discrete states, which are expressed in logic functions. In [95], the hybrid automaton system is applied to the carbon catabolite repression model as an application of hybrid automaton systems. The hybrid model integrates metabolism with transcriptional regulation, macromolecule production, and enzyme resources. A hybrid automaton system enables us to study the dynamic interplay between these different cellular processes.

In our study, we will introduce a continuous model of the hybrid model of a metabolic-regulatory network (MRN) of the diauxic shift, see Figure 3.1. The network has two carbon sources C_1, C_2, introduced to the cell in the medium, and converted into the precursor M. The precursor M is a substance that can be converted into another substance. Two regulated proteins RP and T_2 are described by boolean variables $\overline{RP}$ and $\overline{T_2}$. The regulatory rules of the network are as follows:

- If the molar amount of C_1 is greater than the threshold γ, the gene encoding for RP is activated.
- If the molar amount of RP inside the cell is greater than the threshold α, the gene encoding for T_2 is repressed.

The regulatory rules are denoted in Figure 3.1 with green and dark orange arrows.

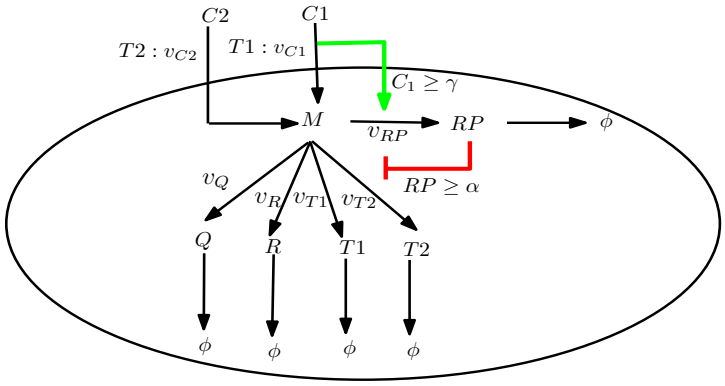


Figure 3.1: A metabolic network with two regulatory rules. C_1, C_2 are the two carbon sources. T_1, T_2 are the enzymes for converting carbon sources into precursor metabolite M. Q represents non-catalytic macromolecule. RP is a regulatory protein, R represents the ribosome catalyzing the protein production.

The components of the hybrid system of MRN (Figure 3.1) are described as follows: the continuous variable Y is $C_1, C_2, M, Q, R, T_1, T_2$, and RP. The continuous variables are expressed by the differential equations. The states of two regulatory proteins are expressed by four modes, see Figure 3.2. These four modes describe the hybrid automaton. The modes are (on, on), (on, off), (off, off), and (off, on). The *Events* are the regulatory proteins RP, T_2 depending on the value of thresholds γ and α, respectively. Each node in Figure 3.2 represents the differential equations of continuous variables and some invariants (constraints).

In the hybrid model, the Michaelis-Menten kinetics are used for most reaction rates and mass action laws are used for degradation terms. The differential equations of C_1, C_2, and precursor M in the hybrid model are given by

$$\begin{aligned}\frac{dC_1}{dt} &= -v_{C1},\\ \frac{dC_2}{dt} &= -v_{C2},\\ \frac{dM}{dt} &= v_{C1} + v_{C2} - v_M,\end{aligned}$$

with reaction rates

$$\begin{aligned}v_{C1} &= k_{cat1} \cdot T_1 \cdot C_1/(K_T + C_1),\\ v_{C2} &= k_{cat2} \cdot T_2 \cdot C_2/(K_T + C_2),\\ v_M &= k_r \cdot R \cdot M/(K_r + M),\end{aligned}$$

where k_r, k_{cati}, $i = 1,2$ are a turnover constants, K_T is the Michaelis-Menten constant, and K_r denotes the half-saturation constant, see Table 3.1.
The reaction rates v_p of macromolecular $p \in \{Q, R, T_1, T_2, RP\}$ are assumed to be the same as in [47]. For instance, the differential equation of macromolecular Q is

$$\frac{dQ}{dt} = v_p - kd_e \cdot Q,$$

where kd_e is degradation constant. The reaction rates v_p are given by

$$v_p = \frac{\beta_p}{n_p} \cdot v_M, \;\; v_M = \frac{k_r \cdot M \cdot R}{K_r + M}, \quad p \in \{Q, R, T_1, T_2, RP\},$$

where the weights β_p represent the fraction of cellular resources allocated to protein p, such that $\sum_p \beta_p = 1$. The constant n_p denotes the length of proteins p, for more detail, see [47]. We notice that in every node in Figure 3.2, the weights change such that the sum is one. In node (on, on), there are 5 proteins where the synthesis term exists, so the weight of every protein is $1/5$. We observe that these changes in weights are based on the regulatory rules in every node, they change from $1/5$ to $1/4$ to $1/3$ to $1/4$. The Biomass is defined as the sum of the molecular masses as follows

$$Biomass(t) = w \cdot M(t) + \sum_{p \in \{Q,R,T_1,T_2,RP\}} w \cdot n_p \cdot p(t),$$

where w is average molar weight of precursor M.

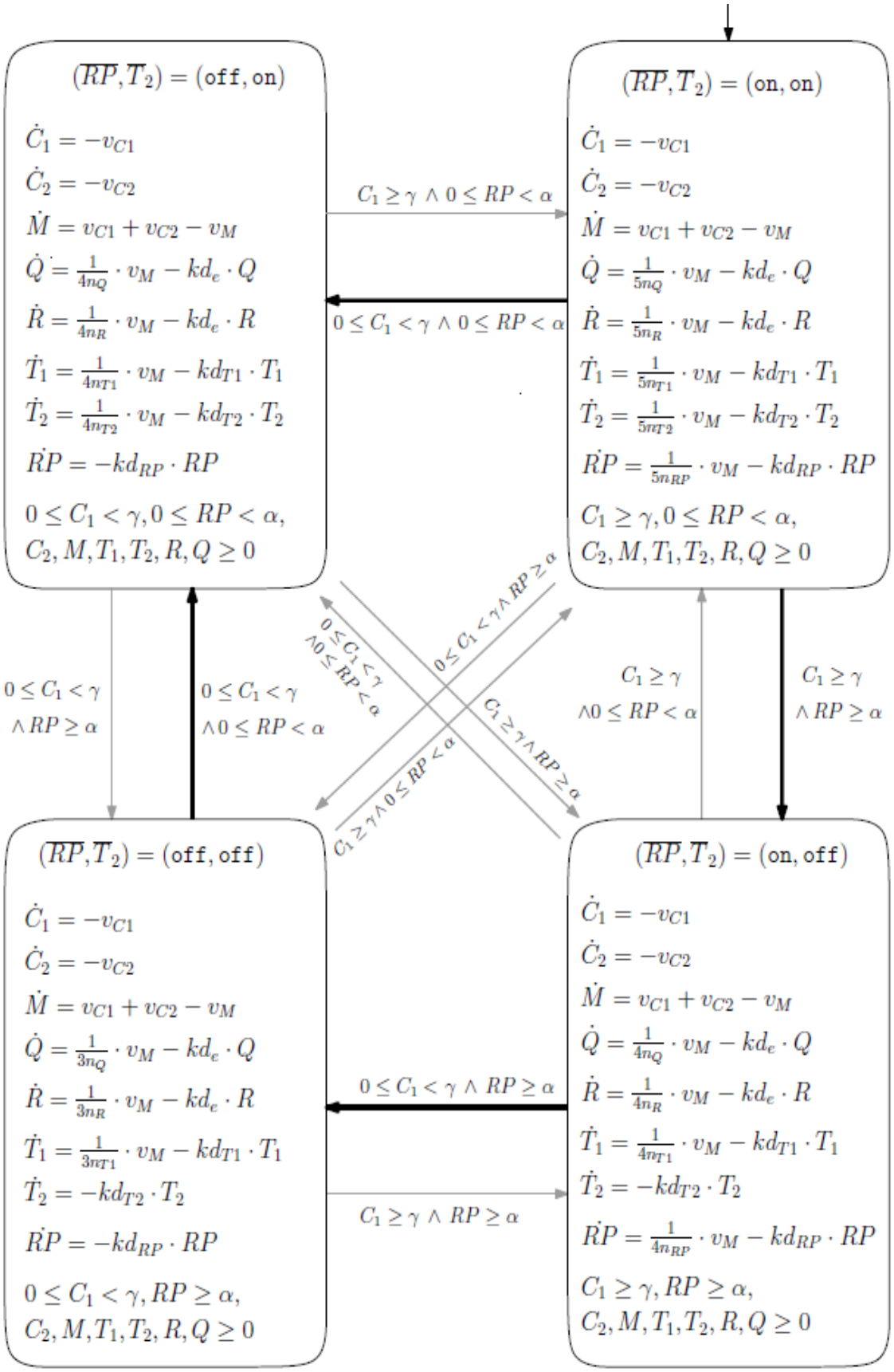


Figure 3.2: The hybrid automaton of metabolic-regulatory network. The figure is taken from [95], License Number: 4781951163564.

Reaction rates	Value	unit
k_{cat1}	3000	min^{-1}
k_{cat2}	2000	min^{-1}
k_r	7	min^{-1}
Michaels-Menten constants		
K_T	1000	$mmol$
K_r	1260	$mmol$
Degradation constants		
kd_e	0.01	min^{-1}
kd_{RP}	0.2	min^{-1}
kd_{T1}	0.05	min^{-1}
kd_{T2}	0.05	min^{-1}
Length of enzyme, ribosome		
n_Q, n_{RP}	300	
n_{T1}	400	
n_{T2}	1500	
n_R	7459	
Thresholds		
γ	20	$mmol$
α	0.03	$mmol$
Average molar weight of precursor		
w	100	$mg\ mmol^{-1}$

Table 3.1: Table of kinetic parameters, enzymes length, and threshold values.

We aim to build a continuous model that mimics the hybrid model of the metabolic-regulatory network Figure 3.1, in which the metabolite concentrations and regulatory proteins are expressed by continuous variables. In addition, we compare the results of the hybrid and continuous models.

3.2 A continuous version of the hybrid automata model

In the continuous model, we use the Michaelis-Menten kinetics for production and consumption reactions and mass action kinetics for degradation reactions as in the hybrid model. In the hybrid model, the regulatory proteins are described by the boolean function with value

0 or 1, while in our continuous model, the regulatory proteins are expressed by the Hill functions. The system of ordinary differential equations of the network is given by

$$\frac{dx}{dt} = f(x(t), \theta) = S \cdot v(x(t), \theta), \quad x(t_0) = x_0, \tag{3.1}$$

where $x \in \mathbb{R}^{n_m}$ is the vector of concentration of n_m species and $\theta \in \mathbb{R}^{n_\theta}$ is a vector of parameters, e.g., kinetics parameters. The function f can be written into the stoichiometric matrix $S \in \mathbb{R}^{n_m \times n_r}$ and reaction rate $v \in \mathbb{R}^{n_r}$, n_r the number of reactions, v can be expressed by the mass action, Michaelis-Menten kinetics, or Hill functions.
In the metabolic regulatory network, two regulatory proteins are RP and T_2 with thresholds γ, α, respectively. The regulatory protein RP has a boolean value 1 if and only if $C_1 \geq \gamma$, and the regulatory protein T_2 has a boolean value 0 if and only if $RP \geq \alpha$. We represent these regulatory proteins by two Hill functions, RPC_1 and RP_{RP}, defined as follows:

- RPC_1 is the Hill function representing the regulatory protein RP, defined as follows

  $RPC_1 = C_1^h/(\gamma^h + C_1^h)$,
  where h is the Hill coefficient which is assumed to be $h = 2$. The function RPC_1 has a value of almost 1 if enough amount of C_1 exists, while RPC_1 has a small value gradually when the concentration of C_1 is decreasing. The function RPC_1 has a value in range $[0, 1]$.

- RP_{RP} is the Hill function representing the regulatory protein T_2, defined as follows

  $RP_{RP} = \alpha^h/(\alpha^h + RP^h)$.

  The function RP_{RP} has a value of almost 0 if enough amount of RP exists, while RP_{RP} has a high value gradually when the value of RP is decreasing.

Now, we will express the regulatory proteins by the Hill functions in the differential equations of RP and T_2, see Table 3.2. For instance, the differential equation of T_2 is given by

$$\frac{dT_2}{dt} = v_M \cdot RP_{RP} - kd_{T2} \cdot T_2,$$

where kd_{T2} is the degradation constant and RP_{RP} is the Hill function.
We used the same parameters and initial conditions as in [95]. The values of constant reaction rates of Michaels-Menten, mass action, and Hill kinetics are given in Table 3.1. For the sake of simplicity, we use the same notations of parameters to make it easier for the reader to compare the different models. Notice that the units of (Q, R, T_1, T_2, RP) are in *mmol* that convert into *mg* unit using the formula: $w \cdot n_p \cdot p(t)$.

Ordinary differential equations of metabolic-regulatory network (MRN) model

$$\frac{dC_1}{dt} = -v_{C1}$$

$$\frac{dC_2}{dt} = -v_{C2}$$

$$\frac{dM}{dt} = v_{C1} + v_{C2} - v_M$$

$$\frac{dQ}{dt} = \frac{\beta_p}{n_Q} \cdot v_M - kd_e \cdot Q$$

$$\frac{dR}{dt} = \frac{\beta_p}{n_R} \cdot v_M - kd_e \cdot R$$

$$\frac{dT_1}{dt} = \frac{\beta_p}{n_{T1}} \cdot v_M - kd_{T1} \cdot T_1$$

$$\frac{dT_2}{dt} = \frac{\beta_p}{n_{T2}} \cdot v_M \cdot RP_{RP} - kd_{T2} \cdot T_2$$

$$\frac{dRP}{dt} = \frac{\beta_p}{n_{RP}} \cdot v_M \cdot RPC_1 - kd_{RP} \cdot RP$$

Reaction rates

$$v_{C1} = k_{cat1} \cdot T_1 \cdot C_1/(K_T + C_1)$$

$$v_{C2} = k_{cat2} \cdot T_2 \cdot C_2/(K_T + C_2)$$

$$v_M = k_r \cdot R \cdot M/(K_r + M)$$

$$RPC_1 = C_1^2/(\gamma^2 + C_1^2)$$

$$RP_{RP} = \alpha^2/(\alpha^2 + RP^2)$$

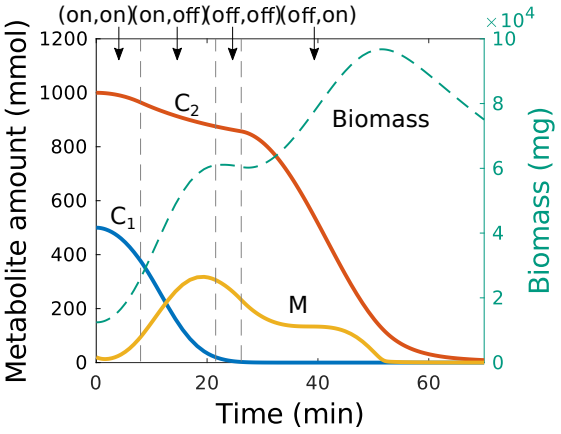
Biomass equation

$$Biomass(t) = w \cdot M(t) + \sum_{p \in \{Q,R,T_1,T_2,RP\}} w \cdot n_p \cdot p(t)$$

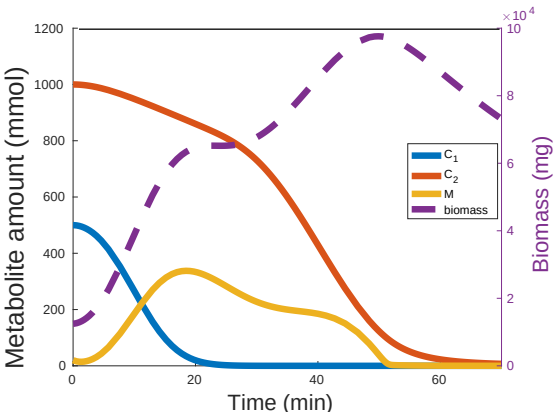
Table 3.2: System of ordinary differential equations of the simplified metabolic regulatory network.

3.3 Comparison of the continuous and hybrid model

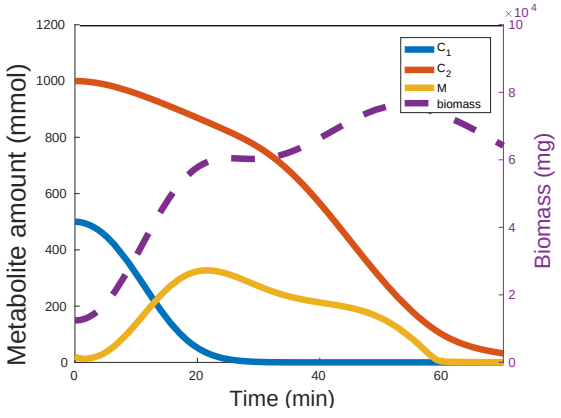
In the hybrid model, the weight β_p change in each node from $\beta_p = 1/5$ in mode (on, on) to $\beta_p = 1/4$ in mode (on, off) to $\beta_p = 1/3$ in mode (off, off) to $\beta_p = 1/4$ in mode (off, on), see Figure 3.2. In order to find the right expression for these values in the continuous model, we will substitute the weight value of the regulatory proteins (β_p) in the continuous model by 1/5, 1/4, or 1/3. Then we discuss the simulation results of our continous model and the hybrid model as follows. Notice that the results of the hybrid model are taken from [95] with License Number: 4781951163564.



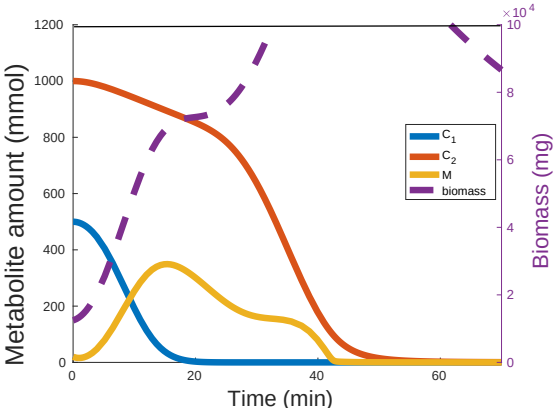
(a) Hybrid model, figure is taken from [95].



(b) The value of the weight β_p is $\frac{1}{4}$

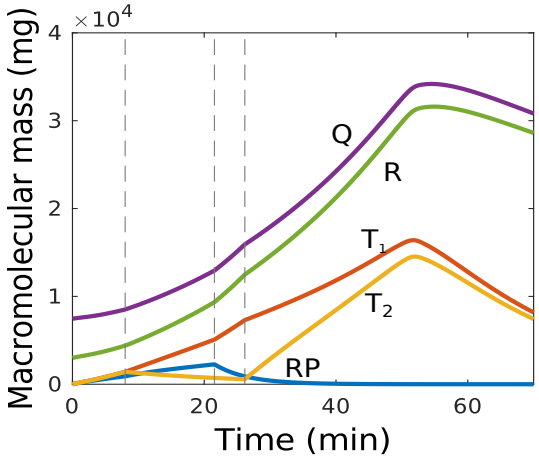


(c) The value of the weight β_p is $\frac{1}{5}$

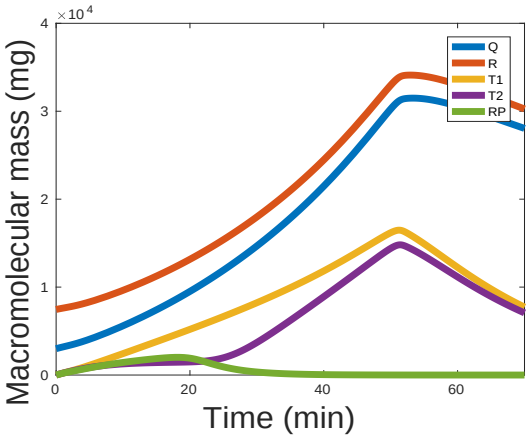


(d) The value of the weight β_p is $\frac{1}{3}$

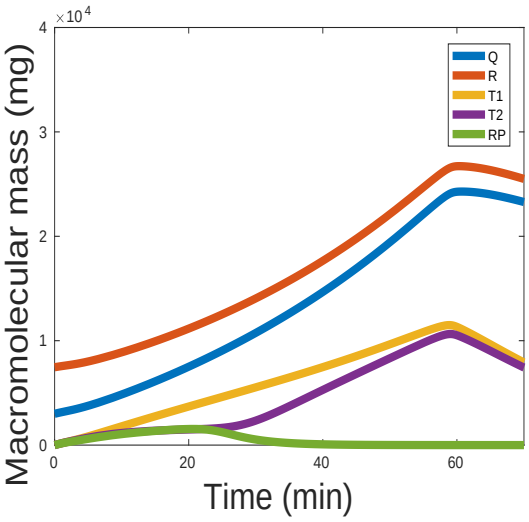
Figure 3.3: a Simulation results of metabolites C_1, C_2, M (left axis) and Biomass (right axis) in the hybrid model. **b, c, d** The trajectory of metabolites and Biomass in continuous model with different values of the weight β_p.



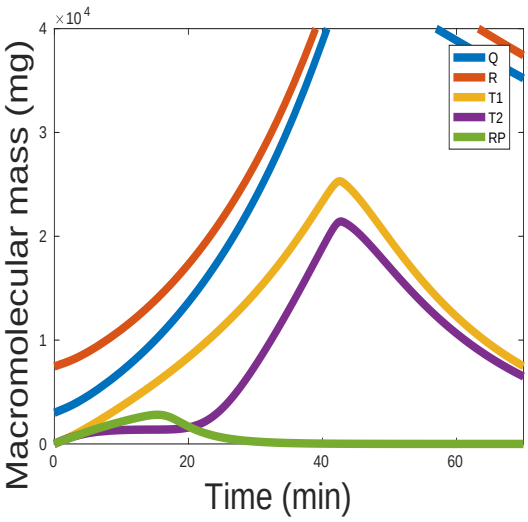
(a) Hybrid model, figure is taken from [95].



(b) The value of the weight β_p is $\frac{1}{4}$



(c) The value of the weight β_p is $\frac{1}{5}$



(c) The value of the weight β_p is $\frac{1}{3}$

Figure 3.4: a Simulation results of macromolecular Q, R, T_1, T_2, and RP in the hybrid model. **b, c , d** Simulation results of macromolecular in the continuous model with different values of the weight β_p.

The simulations of the continuous system were generated in Matlab2016b using the Matlab function `ode45` with tolerances $RTOL = ATOL = 10^{-6}$, and with initial conditions $(C_1, C_2, M, Q, R, T_1, T_2, RP) = (500, 1000, 20, 0.1, 0.01, 0, 0, 0)$ in *mmol* unit. Figures 3.3 and 3.4 depict the results of three different continuous models as well as the result of the hybrid model. In Figures 3.3 and 3.4, we observe that the trajectories of metabolites, macromolecules, and Biomass of the two continuous models **c, d** are a bit different from the hybrid model (case **a**), while the case **b** almost matches the result of the hybrid model. In

case **b**, in the beginning of the simulation, the regulatory proteins *RP* and T_2 increase, while the metabolites C_1 and C_2 consume slowly. Then after some time (0-20 min), it can be seen that the protein *RP* is still increasing while T_2 is almost constant (not produced). At that time, C_2 is not being used, and the cell grows on C_1 which is consumed exponentially. Later, after 20 min, we notice that C_1 is almost exhausted, and the cell starts to utilize the source of carbon C_2. In addition, at about 25 min, the regulatory protein *RP* is not produced anymore and its value becomes zero, while T_2 is increasing, see Figure 3.4.

In the other cases, **c, d**, it is observed that the behavior of Biomass is quite different from the hybrid model. For instance, in case **c**, the trajectory of Biomass looks qualitatively similar but the height of trajectory is lower than the hybrid model. In case **d**, the increment in Biomass is more at the start and gets higher between 40 and 60 minutes compared to the hybrid model. The trajectory of C_2 in case **c** takes time to exhaust until 60 min, while in case **d**, it exhausted a bit faster than the case of the hybrid model.

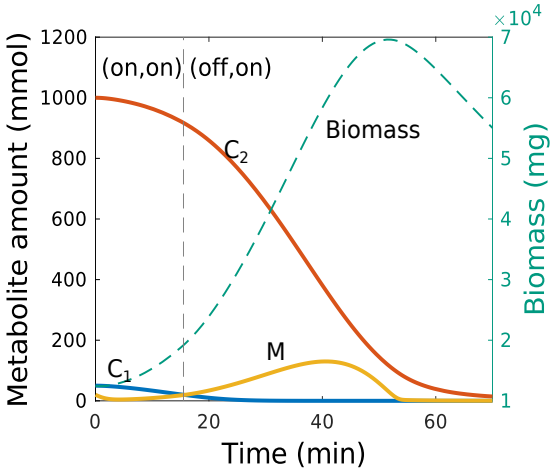
Figure 3.4 depicts the trajectories of macromolecular in different cases. It is observed that in case **b**, the trajectories of T_1, T_2 are qualitatively similar to the hybrid model. In case **c** the trajectories of T_1, T_2 appear lower and a bit later (56 min) as compared to the hybrid model. In contrast, in case **d**, the trajectories of T_1, T_2 appear higher and a bit earlier than the hybrid model.

The above results show that case **b** is considered the best representative on the change of weights in the continuous model, and this seems to contradict with the condition $\sum_p \beta_p = 1$. However, we have another interpretation that case **b** does not contradict with the condition $\sum_p \beta_p = 1$. The interpretation is that at the beginning of the time, the weight of T_2 is not there because T_2 is not active in the presence of C_1. It means that we have four macromolecules present and the sum of weight is 1. During the time, C_1 is consumed, then the regulatory protein *RP* does not exist and T_2 is active, i.e., the sum of weight is 1. Thus, we will represent the weight in the continuous model by the value of $\beta_p = 1/4$.

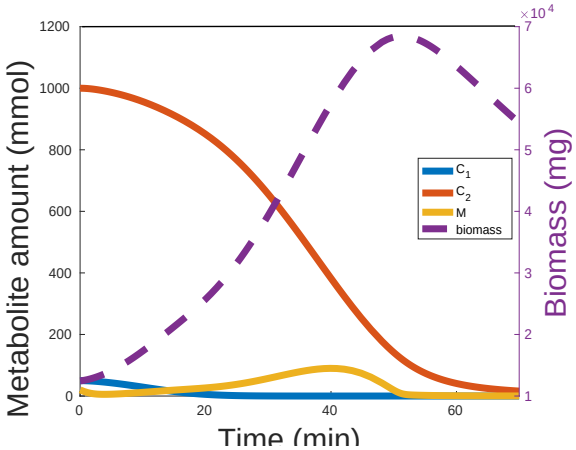
3.3.1 A diauxic shift for different initial conditions

In this scenario, we discuss the continuous model with a different initial condition of C_1. Assume that the initial condition is 50 instead of 500 *mmol*.
In Figures 3.5, 3.6, we observe that the proteins *RP* and T_2 are increasing at the beginning while two sources of carbon C_1 and C_2 are consuming. After a short time, the regulatory protein *RP* starts to decrease based on the behavior of C_1, which decreases linearly while T_2

is still increasing. At 20 minutes, we notice the C_1 is exhausted, which leads to RP not being produced anymore, in contrast with T_2, which is increasing. Figure 3.5 depicts the Biomass trajectory, which is increasing exponentially.

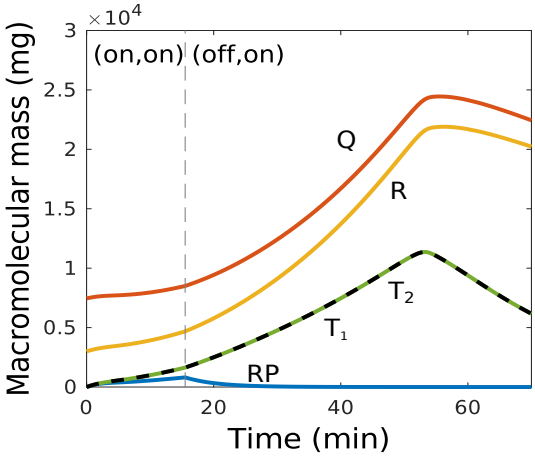


(a) The hybrid model, figure is taken from [95].

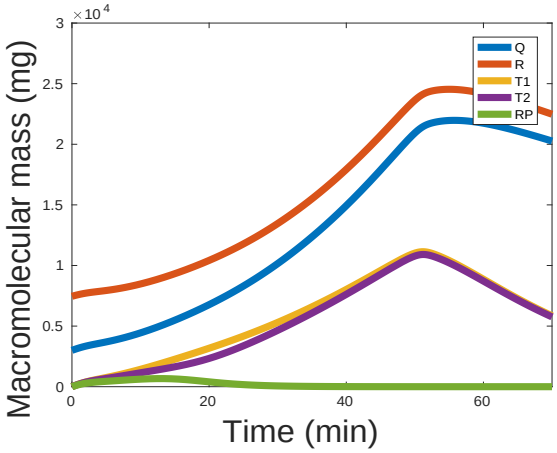


(b) The value of the weight β_p is $\frac{1}{4}$.

Figure 3.5: **a** Simulation results of metabolites C_1, C_2, M (left axis) and Biomass (right axis) with the same of initial conditions as in case one, except $C_1 = 50$ instead of 500 in the hybrid. **b** the trajectory fo metabolites and biomass in the continuous model



(c) Hybrid model, figure is taken from [95].

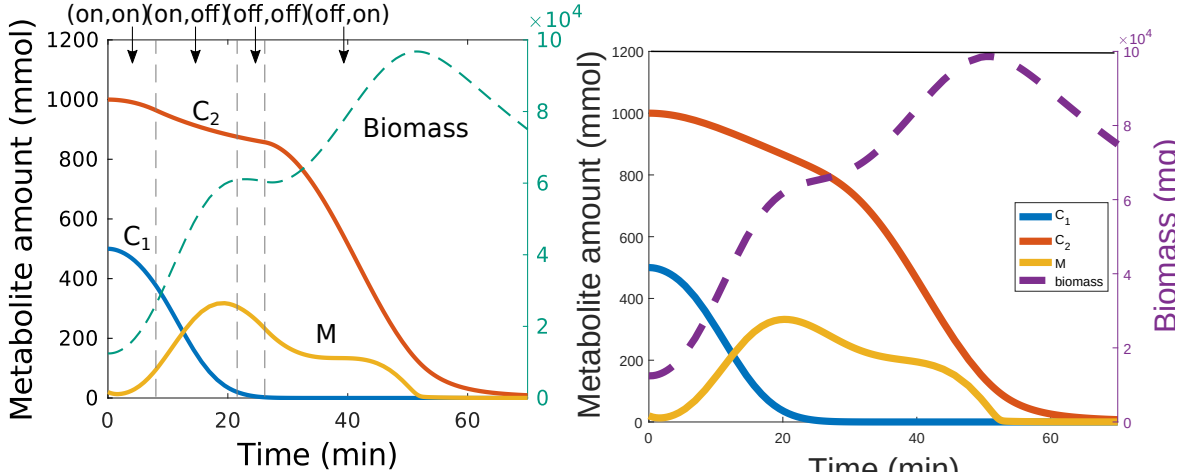


(d) The value of the weight β_p is $\frac{1}{4}$.

Figure 3.6: The trajectory of macromolecular masses Q, R, T_1, T_2, and RP. **c** The hybrid model, **d** The continuous model.

3.3.2 The generalization of a continuous model of a hybrid model for metabolic-regulatory networks

In this section, our goal is to derive a general formula of a continuous model for a hybrid model of the metabolic-genetic network. As we discussed in the previous section, the regulatory proteins are described by the Hill functions. We notice that the weight change in every mode depends on the regulatory protein states. In order to generalize the form of the weight, we formulate the weight as a function of regulatory proteins. In the case study, we have 5 macromolecules, two are regulatory proteins RP and T_2 that are expressed by the Hill functions RP_{Rp}, RPC_1, respectively. Then, we can describe the weights as $\beta_p = 1/((5-2)+RP_{Rp}+RPC_1)$. By substituting this term in the ODE system and comparing the simulation results of the continuous and hybrid models, we obtain the results in Figures 3.7, 3.8.



(a) Hybrid model, figure is taken from [95].

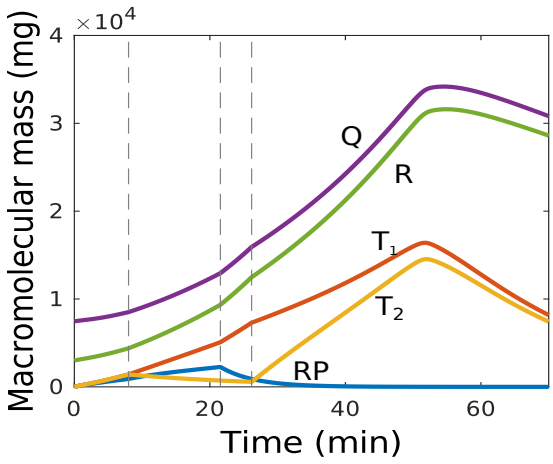
(b) The weight β_p is $1/(3+RP_{Rp}+RPC_1)$

Figure 3.7: Simulation results of metabolites C_1, C_2, M (left axis), and Biomass (right axis). **a** The hybrid model. **b** The continuous model.

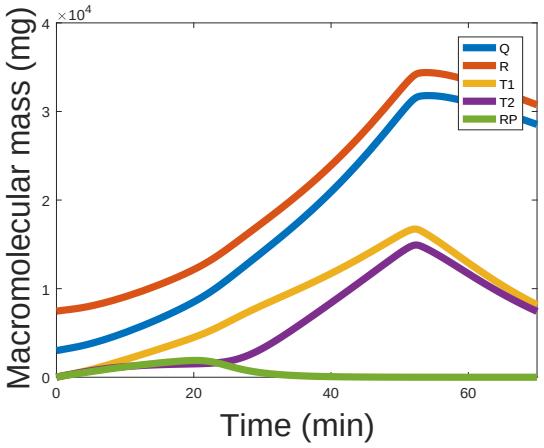
We observe that the trajectories of macromolecules in the continuous and hybrid models are qualitatively similar. The general formula of the weight is given by

$$\beta_p = \frac{1}{(n-ng)+Reg_{ng}},$$

where $n = 1, 2, \ldots.$ is the number of all macromolecules and $ng = 1, 2, \ldots.$ is the number of regulatory proteins, and Reg is the Hill functions of the regulatory proteins.



(c) Hybrid model, figure is taken from [95]



(d) The weight is β_p is $1/(3+RP_{Rp}+RPC_1)$

Figure 3.8: The trajectory of the macromolecules in the hybrid and continuous models. **c** The hybrid model. **d** The continuous model.

3.4 Conclusion

In this chapter, we have introduced a continuous model of the hybrid system of a metabolic-regulatory network model studied in [95]. We have discussed two different scenarios with different initial conditions of C_1. We studied different differential equation systems to describe the weight's changes of the macromolecules (β_p) in the metabolic-regulatory network. We compared the simulation results of the continuous model of the different values of β_p with the hybrid model. We conclude that the continuous model of the weight $\beta_p = \frac{1}{4}$ is considered the best choice to represent the weight changes of regulatory proteins because the simulation of continuous model fits well with the hybrid model. Our continuous model is of great advantage in cases with large number of regulatory rules. Whereas, the hybrid model becomes more complicated if the number of regulatory rules is large because the modeling of regulatory proteins and detecting of events becomes difficult. In the study example, we noticed four (2^2) transition graphs for describing two regulatory proteins. This means that the transition graph numbers will increase by 2^n where $n = 1, 2, ...,$ is the number of regulatory rules, which makes the studying of the hybrid model a bit difficult. Moreover, we generalized the form of the continuous model for the hybrid model of metabolic-regulatory networks.
In the previous discussions, we proposed kinetic models of the rFBA and hybrid models of a metabolic-genetic network. In the following chapters, we aim to discuss a mathematical method called model order reduction. We aim to apply the method to kinetic models of different metabolic-genetic networks.

Chapter 4

Model reduction by time scale separation technique

In this chapter, we study a model order reduction method for metabolic-genetic networks by time scale separation technique. Model order reduction is a mathematical concept to find a low-dimensional approximation for a system of high-dimension. A successful reduction method can simplify a model while preserving its relevant features. In metabolic-genetic networks, metabolic reactions occur at rates in seconds or less (fast variables), while gene expression usually takes between minutes and hours to complete (slow variables). The discrepancy in the time of metabolic and regulation reactions leads to a model reduction topic known as time scale separation technique. However, in some cases, the differences in the time scales do not suffice for obtaining a reduced model that behaves in the same way as the original model [50]. Thus, we will discuss a theory called Tikhonov's theorem [62], which guarantees that the reduced and original models have similar behavior if its conditions are satisfied. In the available literature, it was not clear enough how to satisfy the required conditions of this technique to biological systems. By following the work of [62], we clearly described the boundary-layer system which is needed to show the asymptotic stability of the fast variables. We apply the time scale separation technique to different examples of a metabolic-genetic network. In our study, we check numerically that the required conditions are satisfied.

4.1 Time-scale properties

In this section, we will define a dynamical system that describes the metabolites and enzymes in a metabolic-genetic network. Metabolites variables m move faster in time than enzymes

variables e. We can represent this phenomenon using the expression of a singular perturbation model.

The singular perturbation model of a dynamical system is a state-space model in which the derivatives of some of the states are multiplied by a small positive parameter ε, see [62]. Consider the standard singular perturbation model of a metabolic-genetic network

$$\begin{aligned} \varepsilon \dot{m}(t) &= f(m(t), e(t)), \quad m(t_0) = m_0, \\ \dot{e}(t) &= g(m(t), e(t)), \quad e(t_0) = e_0, \end{aligned} \tag{4.1}$$

for all $t \in [t_0, T]$, where $T \in \mathbb{R}_{\geq 0}$, the functions $f : \mathbb{R}^{n_m} \times \mathbb{R}^{n_e} \to \mathbb{R}^{n_m}$, $g : \mathbb{R}^{n_m} \times \mathbb{R}^{n_e} \to \mathbb{R}^{n_e}$ are sufficiently smooth, n_m, n_e are the numbers of metabolites and enzyme variables, respectively. The constant $\varepsilon \in \mathbb{R}^+$ is a small positive real number. The idea behind time scale separation technique for metabolic-genetic networks is that the metabolite variables $\dot{m} = f(m,e)/\varepsilon$ are very fast compared to the slow variable e. The fast variables m evolve much faster, reach to quasi-steady-state. When we set $\varepsilon = 0$ in (4.1), the dimension of the state equations reduces from $n_e + n_m$ to n_e, such that $\varepsilon \dot{m} = f$ convert into the algebraic equation:

$$f(m(t), e(t)) = 0. \tag{4.2}$$

Suppose that there exist q functions r_i, $i = 1, 2, \ldots, q$, such that (4.2) can be solved for m as

$$m = r_i(e), \;\; i = 1, 2, ..., q. \tag{4.3}$$

Then, we substitute (4.3) into (4.1) at $\varepsilon = 0$. The reduced model is given as follows:

$$\begin{aligned} \dot{\bar{e}}(t) &= g(r(\bar{e}(t)), \bar{e}(t)), \quad \bar{e}(t_0) = e_0, \\ \bar{m}(t) &= r(\bar{e}(t)), \end{aligned} \tag{4.4}$$

where we have dropped the subscript i from r and denoted the solution of (4.4) by $\bar{e}(t)$. Notice that the original variable m is starting at t_0, but the quasi steady state $\bar{m}$ is not free to start from t_0, there is a discrepancy between the initial values [62]. However, we assume that

$$m(t) - \bar{m}(t) = O(\varepsilon),$$

holds on an interval excluding t_0, that is $t \in [t^*, T]$, where $t^* > t_0$. If the error $m(t) - \bar{m}(t)$ is $O(\varepsilon)$ over $[t^*, T]$, then it must be true that the variable m approaches $\bar{m}$ during the initial

(boundary-layer) interval $[t_0, t^*]$. In order to m converges to its quasi-steady-state $\bar{m}$, stability conditions should be satisfied.

Here, we will study the dynamical system (4.1) at the boundary-layer interval $[t_0, t^*]$. We introduce the new variable $\hat{m}$ such that

$$\hat{m} = m - r(e),$$

then, the system (4.1) is written in the new variables $(e, \hat{m})$ as

$$\varepsilon \dot{\hat{m}} = f(\hat{m} + r(e), e) - \varepsilon \frac{\partial r(e)}{\partial e} g(\hat{m} + r(e), e), \quad \hat{m}(t_0) = m(t_0) - r(e(t_0)), \tag{4.5}$$

$$\dot{e} = g(\hat{m} + r(e), e), \quad e(t_0) = e_0, \tag{4.6}$$

where $\hat{m} = 0$ is the equilibrium point of the system (4.5). We define the new variable $\tau = t/\varepsilon$ and set $\varepsilon \dfrac{d\hat{m}}{dt} = \dfrac{d\hat{m}}{d\tau}$. Then the system (4.5) is represented by

$$\frac{d\hat{m}}{d\tau} = f(\hat{m} + r(e(\varepsilon\tau)), e(\varepsilon\tau)) - \varepsilon \frac{\partial r(e)}{\partial e} g(\hat{m} + r(e), e), \tag{4.7}$$

by setting $\varepsilon = 0$, then the system (4.7) becomes

$$\frac{d\hat{m}}{d\tau} = f(\hat{m} + r(e(0)), e(0)). \tag{4.8}$$

The system (4.8) is called the boundary-layer system. The stability property that should be satisfied for the boundary-layer system is the exponential stability of its equilibrium point. The following definition is mentioned in [62] (Definition 9.1).

Definition 1. *The equilibrium $\hat{m} = 0$ of the boundary-layer system* (4.8) *is exponentially stable, uniformly in e_0, if there exist positive constants κ, ψ, and ρ such that the solution of* (4.8) *satisfies*

$$\|\hat{m}(\tau)\| \leq \kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|, \quad \forall \|\hat{m}(\tau_0)\| \leq \rho, \qquad \forall \tau \geq 0. \tag{4.9}$$

Theorem 1. *Consider the singular perturbation problem* (4.1) *for all $t \in [t_0, T]$ where $T \in \mathbb{R}_{\geq 0}$ and assume the system* (4.1) *has the unique solutions $m(t)$, $e(t)$. Consider the following conditions:*

- *There exists an isolated root* $m = r(e)$ *of* (4.2) *such that* $f(\bar{e}(t), r(\bar{e}(t))) = 0$ *for all* $t \in [t_0, T]$*, where* $\bar{e}(t)$ *denotes the unique solution over* $[t_0, T]$ *of the reduced system* $\dot{\bar{e}} = g(r(\bar{e}), \bar{e}), \quad e(t_0) = e_0$.

- *The origin of the boundary-layer model* (4.8) *is exponentially stable, uniformly in e, i.e., the solutions of* (4.8) *satisfy the inequality* (4.9).

Then, the relations (4.10)*,* (4.12) *hold for all* $t \in [t_0, T]$ *and there exist a time* $t^* \geq 0$*, such that* (4.11) *holds for all* $t \in [t^*, T]$.

$$e(t) = \bar{e}(t) + O(\varepsilon), \tag{4.10}$$

$$m(t) = \bar{m}(t) + O(\varepsilon). \tag{4.11}$$

$$m(t) = \bar{m}(t) + \hat{m}(t) + O(\varepsilon), \tag{4.12}$$

The theorem 1 is known as Tikhonov's theorem, see [62; 89]. In the following sections, we will discuss two examples for unimolecular and bimolecular reactions as applications of a time scale separation technique for model reduction of a metabolic-genetic network. Another example of bimolecular reactions is studied, see Appendix A.2.

4.2 Applications to metabolic-genetic networks

In the following examples, we will follow the method of applying the time scale separation technique as it is introduced in section 4.1, while trying to satisfy the conditions of Tikhonov's theorem.

4.2.1 Unimolecular reactions network

At first, we will discuss the unimolecular reaction network. The network describes a pathway that converts the metabolites 1,2 into metabolite 4. Suppose that there is a plasmid coding for enzyme e, whose expression is activated by high concentration of metabolite 4. The enzyme e catalyses the reaction that converts metabolite 3 into metabolite 5. The network is shown in Figure 4.1.

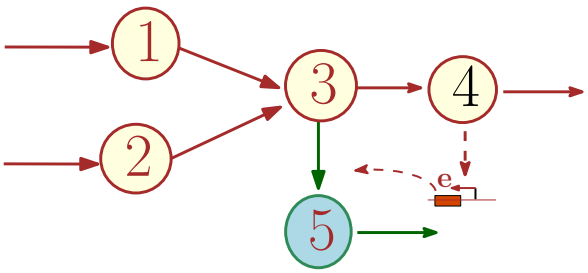


Figure 4.1: Metabolic-genetic network. The pathway converts metabolite 1 and metabolite 2 into metabolite 4. The enzyme e catalyses the reaction that converts metabolite 3 into metabolite 5.

To form a mathematical model of enzyme and metabolite components, we use the Michaelis-Menten kinetics to describe the reaction of metabolites. The mass action kinetics are used for degradation terms of the enzyme, and the Hill function for the synthesis term of the enzyme, see [89]. The dynamical system of the network is given as follows:

$$\begin{aligned}
\frac{dm_1}{dt} &= I_1 - \frac{k_{cat1}m_1}{K_{M1}+m_1}e_{1\to3},\\
\frac{dm_2}{dt} &= I_2 - \frac{k_{cat1}m_2}{K_{M1}+m_2}e_{2\to3},\\
\frac{dm_3}{dt} &= \frac{k_{cat1}m_1}{K_{M1}+m_1}e_{1\to3} + \frac{k_{cat2}m_2}{K_{M2}+m_2}e_{2\to3} - \frac{k_{cat3}m_3}{K_{M3}+m_3}e_{3\to4} - \frac{k_{cat4}m_3}{K_{M4}+m_3}e_{3\to5},\\
\frac{dm_4}{dt} &= \frac{k_{cat3}m_3}{K_{M3}+m_3}e_{3\to4} - \frac{E_4m_4}{K_{O4}+m_4},\\
\frac{dm_5}{dt} &= \frac{k_{cat4}m_3}{K_{M4}+m_3}e_{3\to5} - \frac{E_5m_5}{K_{O5}+m_5},\\
\frac{de}{dt} &= k_0 + k_1\Gamma(m_4) - k_d e,
\end{aligned} \tag{4.13}$$

where $\Gamma(m_4) = \dfrac{m_4^h}{\gamma^h + m_4^h}$ is the Hill function, $k_0 + k_1\Gamma(m_4)$ is the synthesis term of enzyme e that describes transcription and translation, and $k_d e$ is the degradation term of enzyme.
The values of constant rates and initial conditions are taken from [89]. The constant $k_{cati} = k_{cat} = 32s^{-1}$, $K_{Mi} = K_M = 4.7\mu M$ for $i = 1,2,3,4$ and $e_{1\to3} = e_{2\to3} = e_{3\to4} = e_N = 200\ nM$. We assume that $E_4 = E_5 = k_{cat}e_N$, $K_{O4} = K_{O5} = K_M$ and $e = e_{3\to5}$. The values of $k_0 = 5e-7\ nM$, $k_1 = 5e-5\ nM$, $k_d = 2e-4s^{-1}$ and $\gamma = 0.2\ \mu M$.

To apply the time scale separation technique, we should represent the dynamical system (4.13) in the same form as (4.1). Thus, we will simplify the model by the following process.

Non-dimensionalization The non-dimensionalization process simplifies the model by reducing the number of parameters used. The dynamical system (4.13) of the network is written in the formula of (4.1) by substituting the variables of the model (4.13) with the following components:

$$y_i = \frac{m_i}{K_{M_i}},\ x = \frac{e}{\hat{e}},\quad \hat{t} = k_d t,\ \ \hat{e} = \frac{k_0 + k_1}{k_d},\ \ i = 1, ..., 5.$$

We obtain a dynamical system in new variables (x, y) that have no physical units. The non-dimensionalization system is given as follows:

$$\begin{aligned}
\varepsilon \frac{dy_1}{d\hat{t}} &= \tilde{I}_1 - \frac{y_1}{1+y_1},\\
\varepsilon \frac{dy_2}{d\hat{t}} &= \tilde{I}_2 - \frac{y_2}{1+y_2},\\
\varepsilon \frac{dy_3}{d\hat{t}} &= \frac{y_1}{1+y_1} + \frac{y_2}{1+y_2} - \frac{y_3}{1+y_3} - \frac{\hat{e}}{e_N}\frac{y_3 x}{1+y_3},\\
\varepsilon \frac{dy_4}{d\hat{t}} &= \frac{y_3}{1+y_3} - \frac{y_4}{1+y_4},\\
\varepsilon \frac{dy_5}{d\hat{t}} &= \frac{\hat{e}}{e_N}\frac{y_3 x}{1+y_3} - \frac{y_5}{1+y_5},
\end{aligned} \tag{4.14}$$

$$\frac{dx}{d\hat{t}} = \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(y_4) - x, \tag{4.15}$$

where $\tilde{I}_i = \dfrac{I_i}{k_{cat} e_N}$, $i = 1, 2$, $\Gamma^*(y_4) = \Gamma(K_M y_4)$, and $\varepsilon = \dfrac{K_M k_d}{k_{cat} e_N} \approx 1.5 \times 10^{-4}$. By setting $\varepsilon = 0$, the right-hand side of metabolite equations (4.14) converges to zero. We obtain a system of differential-algebraic equations with five algebraic equations and one differential equation. The algebraic equations have unique roots that are given as follows:

$$r_1(x) = \frac{\tilde{I}_1}{1-\tilde{I}_1},\quad r_2(x) = \frac{\tilde{I}_2}{1-\tilde{I}_2},\quad r_3(x) = r_4(x) = \frac{\tilde{I}_1 + \tilde{I}_2}{\dfrac{\hat{e}}{e_N}x + 1 - \tilde{I}_1 - \tilde{I}_2},\quad r_5(x) = \frac{\tilde{I}_1 + \tilde{I}_2}{\dfrac{\hat{e}}{e_N}\dfrac{1}{x} + 1 - \tilde{I}_1 - \tilde{I}_2}. \tag{4.16}$$

The system (4.14),(4.15) with $\varepsilon = 0$, is a *DAEs* system of index 1 [88], which can be seen from the fact that we can write down the unique solution of the algebraic part directly in

(4.16) and obtain an underlying ODE in (4.17). By substituting the value of $r_4(x)$ in the enzyme equation (4.15), the reduced model is given as follows

$$\begin{aligned} \dot{\bar{x}} &= \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(r_4(\bar{x})) - \bar{x}, \quad \bar{x}(0) = x_0, \\ \bar{y} &= r(\bar{x}). \end{aligned} \tag{4.17}$$

The simulations of the original and reduced model were generated in Matlab (R2016b), using the Matlab function `ode45` and `ode15s` with the default setting to solve the corresponding systems of *ODEs* and *DAEs*. The initial conditions are taken from [89], $(y_1, y_2, y_3, y_4, y_5, x) = (0.255, 0.45, 2.4255, 2.617, 1.148, 1.7640)$. We assume the value of import rates $\tilde{I}1$, $\tilde{I}2$ are $1/5$, $1/6$, respectively. The value of import rates is less than one because the root of algebraic equations (4.16) (concentrations) should be positive values.

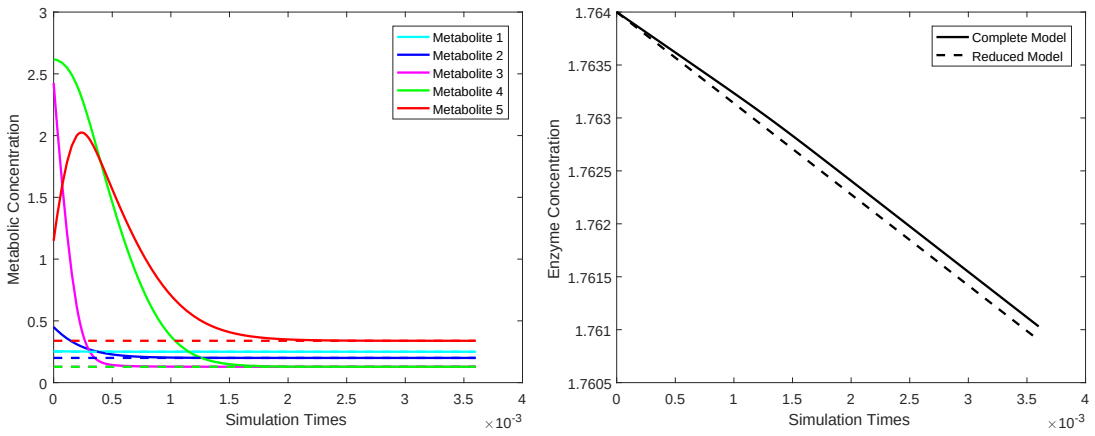


Figure 4.2: a The trajectory of the metabolites in the original model with a solid line and reduced model with a dashed line. **b** The enzyme trajectory of the complete model remains a fraction of a nano molar away from that of the reduced model.

In Figure 4.2, we observe that the fast variables (metabolites) in the original model converge to their steady state values after some time. Then the trajectory of metabolites in the reduced and original model be applicable. The trajectory of the enzyme of the reduced model remains close to that of the original model. We notice that the dynamical system of 6 differential equations is reduced to 1 differential equation system, such that we obtain the main information from the lower-dimensional reduced model.
In order to confirm that the variable y converges to its steady state $r(x)$ and the error $y - r(x)$ is indeed $O(\varepsilon)$, we should satisfy the stability condition of the boundary-layer system.

To discuss the boundary-layer system, we define the new variable $\hat{y} = y - r(x)$, then the system (4.19) can be written as

$$\begin{aligned}
\varepsilon\frac{d\hat{y}_1}{d\hat{t}} &= \bar{I}_1 - \frac{\hat{y}_1 + r_1(x)}{1+\hat{y}_1 + r_1(x)},\\
\varepsilon\frac{d\hat{y}_2}{d\hat{t}} &= \bar{I}_2 - \frac{\hat{y}_2 + r_2(x)}{1+\hat{y}_2 + r_2(x)},\\
\varepsilon\frac{d\hat{y}_3}{d\hat{t}} &= \frac{\hat{y}_1 + r_1(x)}{1+\hat{y}_1 + r_1(x)} + \frac{\hat{y}_2 + r_2(x)}{1+\hat{y}_2 + r_2(x)} - \frac{\hat{y}_3 + r_3(x)}{1+\hat{y}_3 + r_3(x)} - \frac{\hat{e}}{e_N}\,\frac{(\hat{y}_3 + r_3(x))x}{1+\hat{y}_3 + r_3(x)},\\
\varepsilon\frac{d\hat{y}_4}{d\hat{t}} &= \frac{\hat{y}_3 + r_3(x)}{1+\hat{y}_3 + r_3(x)} - \frac{\hat{y}_4 + r_4(x)}{1+\hat{y}_4 + r_4(x)},\\
\varepsilon\frac{d\hat{y}_5}{d\hat{t}} &= \frac{\hat{e}}{e_N}\,\frac{(\hat{y}_3 + r_3(x))x}{1+\hat{y}_3 + r_3(x)} - \frac{\hat{y}_5 + r_5(x)}{1+\hat{y}_5 + r_5(x)}.
\end{aligned}$$

By setting $\varepsilon\dfrac{d\hat{y}}{d\hat{t}} = \dfrac{d\hat{y}}{d\tau}$, then the boundary-layer system is given by

$$\begin{aligned}
\frac{d\hat{y}_1}{d\tau} &= \bar{I}_1 - \frac{\hat{y}_1 + r_1(x_0)}{1+\hat{y}_1 + r_1(x_0)},\\
\frac{d\hat{y}_2}{d\tau} &= \bar{I}_2 - \frac{\hat{y}_2 + r_2(x_0)}{1+\hat{y}_2 + r_2(x_0)},\\
\frac{d\hat{y}_3}{d\tau} &= \frac{\hat{y}_1 + r_1(x_0)}{1+\hat{y}_1 + r_1(x_0)} + \frac{\hat{y}_2 + r_2(x_0)}{1+\hat{y}_2 + r_2(x_0)} - \frac{\hat{y}_3 + r_3(x_0)}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{e}}{e_N}\,\frac{(\hat{y}_3 + r_3(x_0))x_0}{1+\hat{y}_3 + r_3(x_0)},\\
\frac{d\hat{y}_4}{d\tau} &= \frac{\hat{y}_3 + r_3(x_0)}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{y}_4 + r_4(x_0)}{1+\hat{y}_4 + r_4(x_0)},\\
\frac{d\hat{y}_5}{d\tau} &= \frac{\hat{e}}{e_N}\,\frac{(\hat{y}_3 + r_3(x_0))x_0}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{y}_5 + r_5(x_0)}{1+\hat{y}_5 + r_5(x_0)}.
\end{aligned} \tag{4.18}$$

The simulations of the boundary-layer system (4.18) were generated in Matlab (R2016b) using the Matlab function `ode45` with the default setting to solve the corresponding systems of *ODEs* with initial conditions $\hat{y}(0) = y(0) - r(x(0))$. In Figure 4.3a, we observe that the trajectories of the metabolites of the boundary-layer system converge to the equilibrium point $\hat{y} = 0$. We check the exponential stability of equilibrium point of boundary-layer system by satisfying the inequality in (4.9). We choose the values of constants $\kappa = 2, \psi = 0.3$ that satisfy the inequality (4.9). Thus, we find that $\|\hat{m}(\tau)\| \leq \kappa e^{(-\psi\tau)}\,\|\hat{m}(\tau_0)\|$, see Figure 4.3b. The equilibrium point $\hat{y} = 0$ of the system (4.18) is exponentially stable because the inequality (4.9) is satisfied. By applying Tikhonov's theorem, the error of the reduced and complete system will be of order $O(\varepsilon)$.

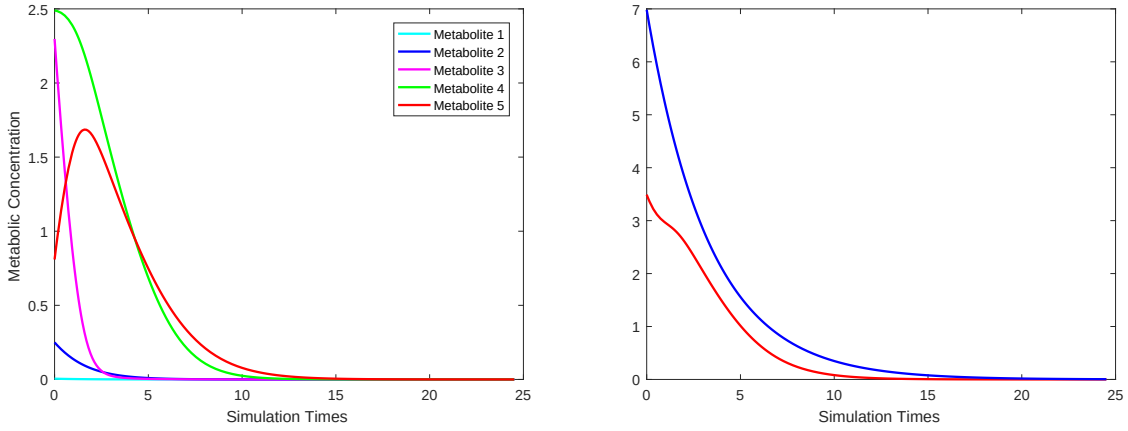


Figure 4.3: a The trajectory of metabolites of the boundary-layer system converges to $\hat{y} = 0$. **b** The red trajectory is the norm of $\|\hat{m}(\tau)\|$ and the blue trajectory is the value of $\kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|$.

4.2.2 Bimolecular reactions network

In this subsection, we discuss a bimolecular reactions of a metabolic-genetic network. The pathway converts a metabolite 1 into metabolite 4. Suppose that there is a plasmid coding for enzyme e, whose expression is activated by high concentration of metabolite 4. The enzyme e catalyses the reaction that converts metabolite 2 into metabolite 5. The network is shown in the figure below.

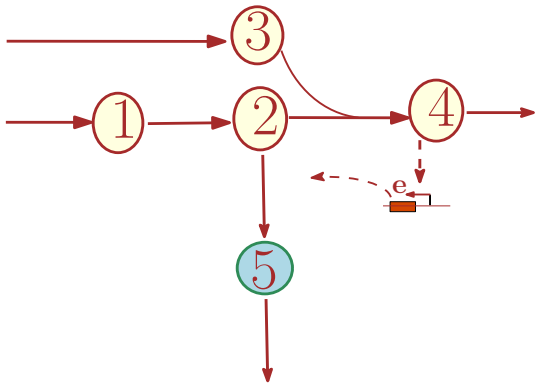


Figure 4.4: Metabolic-genetic network. The pathway converts metabolite 1 into metabolite 4. The enzyme e catalysis the reaction that converts metabolite 2 into metabolite 5.

The mathematical model of the network is expressed by a system of ordinary differential equations. Using kinetic laws as the Michaelis-Menten kinetics for metabolites reaction, the Hill function for enzyme synthesis, and mass action kinetics for enzyme degradation. The

dynamical system is given as follows:

$$\begin{aligned}
\frac{dm_1}{dt} &= I_1 - \frac{k_{cat1} m_1}{K_{M1}+m_1} e_{1\to 2},\\
\frac{dm_2}{dt} &= \frac{k_{cat1} m_1}{K_{M1}+m_1} e_{1\to 2} - k_{cat2,3} \frac{m_2}{K_{M2}+m_2} \frac{m_3}{K_{M3}+m_3} e_{2,3\to 4} - \frac{k_{cat2} m_2}{K_{M2}+m_2} e_{2\to 5},\\
\frac{dm_3}{dt} &= I_2 - k_{cat2,3} \frac{m_2}{K_{M2}+m_2} \frac{m_3}{K_{M3}+m_3} e_{2,3\to 4},\\
\frac{dm_4}{dt} &= k_{cat2,3} \frac{m_2}{K_{M2}+m_2} \frac{m_3}{K_{M3}+m_3} e_{2,3\to 4} - \frac{E_4 m_4}{K_{O4}+m_4},\\
\frac{dm_5}{dt} &= \frac{k_{cat2} m_2}{K_{M2}+m_2} e_{2\to 5} - \frac{E_5 m_5}{K_{O5}+m_5},\\
\frac{de}{dt} &= k_0 + k_1 \Gamma(m_4) - k_d e,
\end{aligned}$$

where, $\Gamma(m_4) = \dfrac{m_4^h}{\gamma^h + m_4^h}$ is the Hill function, $k_0 + k_1\Gamma(m_4)$ is the synthesis term of enzyme e that describes transcription and translation, and $k_d e$ is the degradation term of enzyme [89]. The constant rates and initial conditions are taken from [89]. The constant $k_{cati} = k_{cat} = 32s^{-1}$, $K_{Mi} = K_M = 4.7\mu M$ for $i = 1,2,3,4$ and $e_{1\to 3} = e_{2,3\to 4} = e_N = 200\, nM$. We assume that $E_4 = E_5 = k_{cat} e_N$, $K_{O4} = K_{O5} = K_M$ and $e = e_{2\to 5}$. The values of $k_0 = 5e-7\ nM$, $k_1 = 5e-5\ nM$, $k_d = 2e-4s^{-1}$ and $\gamma = 0.2\ \mu M$.

For the sake of simplicity and applying the time scale separation technique, we will study the non-dimensionalization system in the new variables (x,y) as follows:

$$\begin{aligned}
\varepsilon \frac{dy_1}{d\hat{t}} &= \bar{I}_1 - \frac{y_1}{1+y_1},\\
\varepsilon \frac{dy_2}{d\hat{t}} &= \frac{y_1}{1+y_1} - \frac{y_2}{1+y_2}\frac{y_3}{1+y_3} - \frac{\hat{e}}{e_N}\frac{y_2 x}{1+y_2},\\
\varepsilon \frac{dy_3}{d\hat{t}} &= \bar{I}_2 - \frac{y_2}{1+y_2}\frac{y_3}{1+y_3},\\
\varepsilon \frac{dy_4}{d\hat{t}} &= \frac{y_2}{1+y_2}\frac{y_3}{1+y_3} - \frac{y_4}{1+y_4},\\
\varepsilon \frac{dy_5}{d\hat{t}} &= \frac{\hat{e}}{e_N}\frac{y_2 x}{1+y_2} - \frac{y_5}{1+y_5},
\end{aligned} \tag{4.19}$$

$$\frac{dx}{d\hat{t}} = \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(y_4) - x. \tag{4.20}$$

Where $\Gamma^*(y_4) := \Gamma(K_M y_4)$. At $\varepsilon = 0$, the fast variables (4.19) converge to their steady-state and we obtain system of *DAEs*. The roots of the algebraic equations are given as follows:

$$r_1(x) = \frac{\tilde{I}_1}{1-\tilde{I}_1}, \quad r_2(x) = a, \quad r_3(x) = \frac{-\tilde{I}_2(1+a)}{a+\tilde{I}_2+a\tilde{I}_2},$$
$$r_4(x) = \frac{\tilde{I}_2}{1-\tilde{I}_2}, \quad r_5(x) = \frac{\tilde{I}_1-\tilde{I}_2}{1-\tilde{I}_1+\tilde{I}_2},$$

where $a = \dfrac{e_N(\tilde{I}_1-\tilde{I}_2)}{\hat{e}x-\tilde{I}_1+\tilde{I}_2}$. To obtain the reduced model, we substitute the value of $r_4(x)$ into (4.20), at $\varepsilon = 0$. Then the reduced model is defined as follows:

$$\begin{aligned} \dot{\bar{x}} &= \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(r_4(\bar{x})) - \bar{x}, \quad \bar{x}(0) = x_0, \\ \bar{y} &= r(\bar{x}). \end{aligned} \tag{4.21}$$

The simulations of *ODEs* and *DAEs* were generated in Matlab (R2016b), using the Matlab function `ode45`, `ode15s` functions with absolute tolerance 10^{-4} and relative tolerance 10^{-6} for *DAEs* solver. The initial conditions are inspired from [89], $(y_1, y_2, y_3, y_4, y_5, x) = (0.255, 0.45, 2.4255, 2.617, 1.148, 1.7640)$. We assume the value of import rates $\tilde{I}1 = 1/2$ and $\tilde{I}2 = 0.03$. We choose these values to avoid inconsistent initial conditions while solving the differential-algebraic equations system.

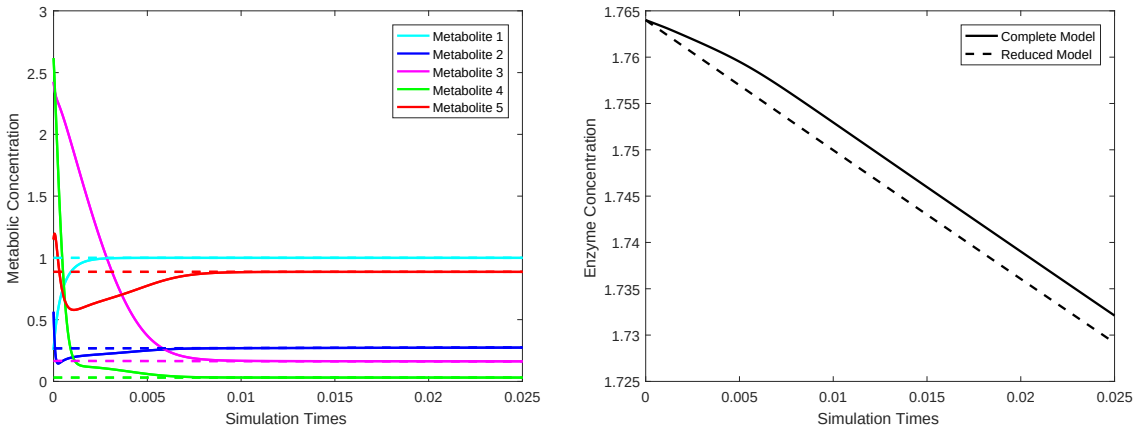


Figure 4.5: **a** The metabolite trajectories of the complete model (solid lines) converge rapidly to that of the reduced model (dashed lines). **b** The trajectory of the enzyme of the complete model remains close to that of the reduced model.

In Figure 4.5, we observe that the fast variables (metabolites) in the original model (4.19) converge to their steady state values after some time. The trajectory of the enzyme of the reduced model remains a nano molar away from that of the original model. Notice that we reduced a 6-dimensional model to a 1-dimensional model.

To discuss the boundary-layer system, we introduce the new variable $\hat{y} = y - r(x)$ and by setting $\varepsilon \frac{d\hat{y}}{d\hat{t}} = \frac{d\hat{y}}{d\tau}$, then the system (4.19) becomes:

$$
\begin{aligned}
\frac{d\hat{y}_1}{d\tau} &= \bar{I}_1 - \frac{\hat{y}_1 + r_1(x_0)}{1 + \hat{y}_1 + r_1(x_0)}, \\
\frac{d\hat{y}_2}{d\tau} &= \frac{\hat{y}_1 + r_1(x_0)}{1 + \hat{y}_1 + r_1(x_0)} - \frac{\hat{y}_2 + r_2(x_0)}{1 + \hat{y}_2 + r_2(x_0)} \frac{\hat{y}_3 + r_3(x_0)}{1 + \hat{y}_3 + r_3(x_0)} - \frac{\hat{e}}{e_N} \frac{(\hat{y}_2 + r_2(x_0))x_0}{1 + \hat{y}_2 + r_2(x_0)}, \\
\frac{d\hat{y}_3}{d\tau} &= \bar{I}_2 - \frac{\hat{y}_2 + r_2(x_0)}{1 + \hat{y}_2 + r_2(x_0)} \frac{\hat{y}_3 + r_3(x_0)}{1 + \hat{y}_3 + r_3(x_0)}, \\
\frac{d\hat{y}_4}{d\tau} &= \frac{\hat{y}_2 + r_2(x_0)}{1 + \hat{y}_2 + r_2(x_0)} \frac{\hat{y}_3 + r_3(x_0)}{1 + \hat{y}_3 + r_3(x_0)} - \frac{\hat{y}_4 + r_4(x_0)}{1 + \hat{y}_4 + r_4(x_0)}, \\
\frac{d\hat{y}_5}{d\tau} &= \frac{\hat{e}}{e_N} \frac{(\hat{y}_2 + r_2(x_0))x_0}{1 + \hat{y}_2 + r_2(x_0)} - \frac{\hat{y}_5 + r_5(x_0)}{1 + \hat{y}_5 + r_5(x_0)}.
\end{aligned}
\tag{4.22}
$$

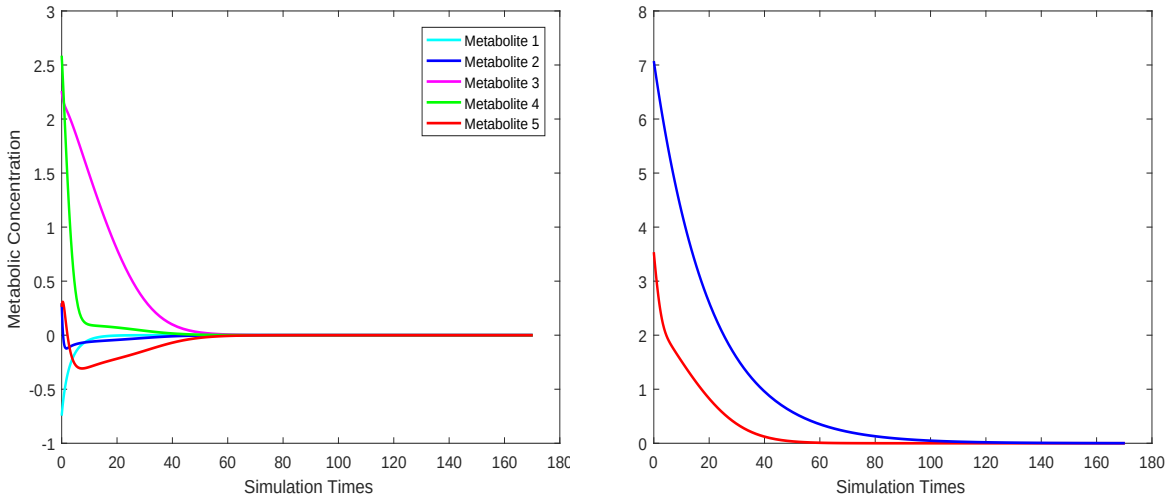


Figure 4.6: a The metabolites trajectory of the boundary-layer system. **b** The red trajectory is the norm of $\|\hat{m}(\tau)\|$ and the blue trajectory is $\kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|$.

In Figure 4.6a, we observe that all the trajectories of the metabolites of the boundary-layer system (4.22) converge to the equilibrium point $\hat{y} = 0$. The exponential stability of the equilibrium point of the system (4.22) have proved by satisfying the inequality in (4.9). By choosing with $\kappa = 2, \psi = 0.05$, we notice that $\|\hat{m}(\tau)\| \leq \kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|$, see Figure 4.6b, thus the equilibrium point $\hat{y} = 0$ is exponential stable.

Uniqueness solution of the reduced system

In the reduced system (4.21), the function $\Gamma^*(r_4(\bar{x}))$ is a constant value because the function $r_4(\bar{x})$ is a constant. Then the system has a unique solution by solving the (4.21) as follows

$$\frac{d\bar{x}}{d\hat{t}} = \hat{h} - \bar{x}, \tag{4.23}$$

where $\hat{h}$ is a constant and $\hat{h} = \dfrac{k_0}{k_0 + k_1} + \dfrac{k_1}{k_0 + k_1}\Gamma^*(cons)$. By integrating the system (4.23),

$$\int \frac{d\bar{x}}{\hat{h} - \bar{x}} = \int d\hat{t},$$

then the reduced system has a unique solution

$$\bar{x} = \hat{h} - ce^{-\hat{t}},$$

where c is the integral constant. Since there is a unique solution of the reduced model and the equilibrium point of the boundary-layer system (4.22) is exponentially stable, the conditions of Tikhonov's theorem are satisfied. Then the difference between the trajectory of the original model and reduced will be of order $O(\varepsilon)$.

4.3 Conclusion

In this chapter, we have discussed the time scale separation technique for model order reduction of a metabolic-genetic network model. We have clarified the required conditions for applying this technique. We applied the technique to unimolecular and bimolecular reaction network models. We have satisfied the conditions of Tikhonov's theorem, that need for the application of the technique. These conditions are the uniqueness of a solution of the reduced system and the exponential stability of an equilibrium point of the boundary-layer system. We have reduced the dimension of a complete system to a lower-dimension system.

The advantage of the time scale separation technique is that it preserves the dynamics of the original system. However, it is difficult to apply it to a large-scale system because the Tikhonov's conditions are difficult to satisfy. Additionally, the separation of slow and fast components in large-scale biological systems is difficult.

Chapter 5

Model reduction by proper orthogonal decomposition

In this chapter, we discuss the model order reduction technique by proper orthogonal decomposition (POD) for kinetic models of biological systems. The POD method works by projecting a high-dimensional system onto a lower-dimensional space while preserving the most important information of the full order system. From a practical point of view, the POD technique is more advantageous for a large-scale model than the time scale separation technique, since the POD technique does not have to take care of satisfying the conditions required, e.g., for Tikhonov's theorem. We apply the POD technique to different biological systems from the BioModels database. Additionally, we use the POD method to compute a reduced-order model for different initial conditions of the dynamical system. Using different initial conditions in the time scale separation technique will usually fail, since the algebraic equations of the fast variables may not be solvable if inconsistent initial values are prescribed.

5.1 Proper orthogonal decomposition

POD is a method of data analysis aimed at obtaining low-dimensional approximate descriptions of high-dimensional processes while retaining the essential features of the full model. POD is known as Karhunen–Loève Expansion (KLE) [75; 97], which was presented by Kari Karhunen. The POD method is based on projecting the full order system onto a lower-dimensional subspace that captures the main characteristics of the full order model. The great advantage of the POD method is that one can use a *Singular Value Decomposition (SVD)* [23; 57; 142] of the snapshot matrix (data set) to calculate the optimal basis that represents the low-dimensional subspace.

The selection of a data set plays an important role, and can be obtained by the *method of snapshots* [133], where the optimal basis is computed based on a set of state solutions.
We consider a system of ordinary differential equations are given by

$$\frac{dx(t)}{dt} = Ax(t) + F(x(t)) \quad x(0) = x_0, \quad \text{for } t \in [0,T], \tag{5.1}$$

where $x(t)$ is the n_m-dimensional state vector of our model, F is a nonlinear function, and A is a constant matrix. We compute the snapshot matrix $\mathbb{X} = [x(t_1), x(t_2), \ldots, x(t_{n_t})] \in \mathbb{R}^{n_m \times n_t}$ by solving a system of *ODEs*. Then, by applying the singular value decomposition (*SVD*) [23; 68], we obtain

$$\mathbb{X} = U \begin{bmatrix} \Sigma_{\hat{r}} & 0 \\ 0 & 0 \end{bmatrix} V^{\mathsf{T}}$$

where $U \in \mathbb{R}^{n_m \times n_m}, V^{\mathsf{T}} \in \mathbb{R}^{n_t \times n_t}$ are unitary matrices with orthonormal columns called singular vectors and $\Sigma_{\hat{r}} = diag(\hat{\sigma}_1, \hat{\sigma}_2, ..., \hat{\sigma}_{\hat{r}}) \in \mathbb{R}^{\hat{r} \times \hat{r}}$ is a matrix with a real, non negative entries on the diagonal and zeros off the diagonal called singular values and $\hat{r} = rank(\mathbb{X})$.

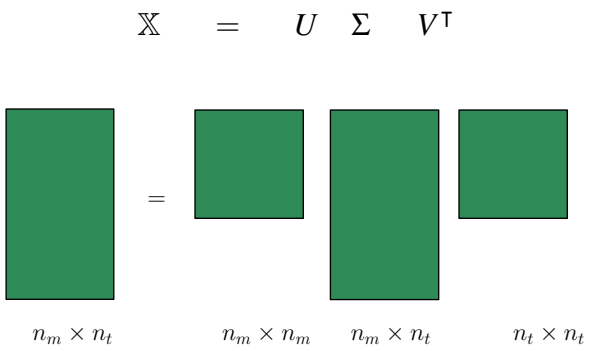


Figure 5.1: Scheme clarifies the singular value decomposition.

Remark:

The SVD is considered to be one of the most important matrix factorizations in data science, as it exists for any matrix and can be used for approximation high-dimensional data by low-dimensional data in terms of dominant patterns. In many natural systems, it is noticed that data exhibit dominant patterns, which may be characterized by a low-dimensional manifold [23; 68]. The SVD is a unique matrix decomposition, it is used for approximation low-dimensional to high-dimensional data in terms of dominant patterns.

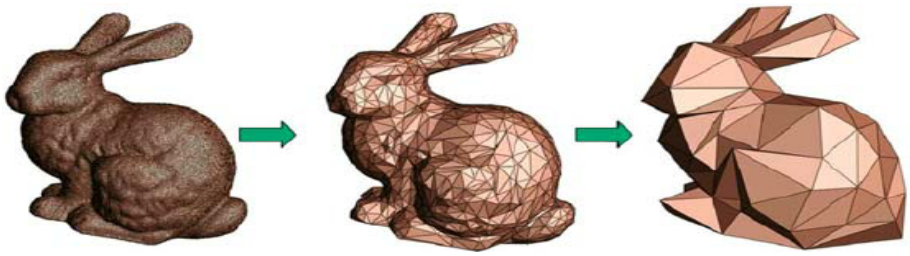


Figure 5.2: Graphical illustration of model order reduction, the figure is taken from [129], License Number: 4782000887372.

Figure 5.2 illustrates the concept in a graphical easy-to-understand way, which means that sometimes little information is needed to describe a model. This figure shows that even with only a few facets, the rabbit can be recognized [129].

Low rank truncation

The most useful property of the *SVD* is that it provides a hierarchy of low-rank approximations. In this step, we truncate the most important features (dominant), which capture the more information about data.
The general philosophy in model order reduction is to cut off singular values with $\hat{\sigma}_i < \varepsilon$, where ε is chosen such that a (much) smaller number of basis elements $n_k \ll n_m$ is sufficient to capture the main features of the solution of (5.1). The goal is to choose n_k small enough while the *relative information content* [2] of the basis for the n_k-dimensional subspace, defined by

$$I(n_k) = \frac{\sum_{i=1}^{n_k} \hat{\sigma}_i^2}{\sum_{i=1}^{\hat{r}} \hat{\sigma}_i^2},$$

is near to one. If the n_k-dimensional subspace should contain a percentage p of the information contained in the full dimensional space $\mathbb{R}^{n_m}$, then one should choose n_k such that

$$n_k = \operatorname{argmin} \left\{ I(n_k) \,|\, I(n_k) \geq \frac{p}{100} \right\}.$$

By selection for U_{n_k} the first n_k columns from U, i.e., $U_{n_k} = [u_1, \ldots, u_{n_k}] \in \mathbb{R}^{n_m \times n_k}$ and $\hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_{n_k}$ as well as $\hat{\sigma}_{n_k} > \hat{\sigma}_{n_k+1} \geq \ldots \geq \hat{\sigma}_{\hat{r}}$ with $\hat{\sigma}_{n_k+1}, \ldots, \hat{\sigma}_{\hat{r}}$ are sufficiently small. That means the first n_k singular values are the (hopefully) few dominant patterns that explain the high-dimensional data. Columns of U like a combination of my measurement that show up in our data and the rows of V is the time history of dominant measurement.

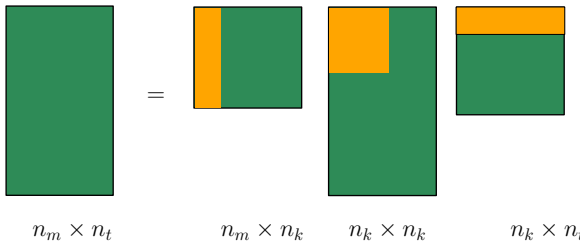


Figure 5.3: Scheme illustrates the truncation of basis vectors

Galerkin Projection

A set of orthonormal basis vectors $\{u_1, \ldots, u_{n_k}\}$ is used to project the system of equations (5.1) onto a lower-dimensional subspace. This set is computed by means of the (POD) method. For the state samples, we search for a subspace $\mathscr{U} = span\{u_1, \ldots, u_{n_k}\} \subset \mathbb{R}^{n_m}$ in which the samples can be optimally described. In other words, the error between samples and their projection $\mathscr{U}$ is minimized in a L_2 sense, the approximation error

$$\sum_{j=1}^{n_t} \left\| x_j - U_{n_k} U_{n_k}^{\mathsf{T}} x_j \right\|_2^2 = \sum_{i=n_k+1}^{\hat{r}} \hat{\sigma}_i^2,$$

see [25]. The Galerkin projection onto the low dimension subspace $\mathscr{U}$ that is used to obtain a reduced order model for (5.1), it takes the form $x(t) = U_{n_k}\tilde{x}(t)$, where $\tilde{x}(t) \in \mathbb{R}^{n_k}$ and dimension $n_k \ll n_m$. The reduced order system is of the form

$$\frac{d\tilde{x}(t)}{dt} = U_{n_k}^{\mathsf{T}} A U_{n_k} \tilde{x}(t) + U_{n_k}^{\mathsf{T}} F(U_{n_k}\tilde{x}(t)).$$

By solving this system of much smaller dimension, the solution of a high-dimensional dynamical system can be approximated. In the following section, we will apply POD model reduction method to some of the biological models from the BioModels database.

5.2 Application of the POD method to kinetic model examples

In the following, we apply the POD model reduction method introduced in the previous section to kinetic models for different kinetic model examples. We consider the simple metabolic-genetic network example of [33], a kinetic model of Lactococcus lactis metabolism introduced in [30], a large-scale kinetic model of yeast metabolic network [138], and a large-scale kinetic model of E. coli metabolic network [134].

5.2.1 Kinetic model of metabolic-genetic network

The dynamical system of the metabolic-genetic network contains *nineteenth* state vector. In Chapter 2, we have built *ODEs* model (2.3) that describes the network. Here, the snapshot matrix is considered to be the solution of *ODEs* then we apply SVD and choose the POD modes, the most dominant basis that describes the data (solution of *ODEs*).

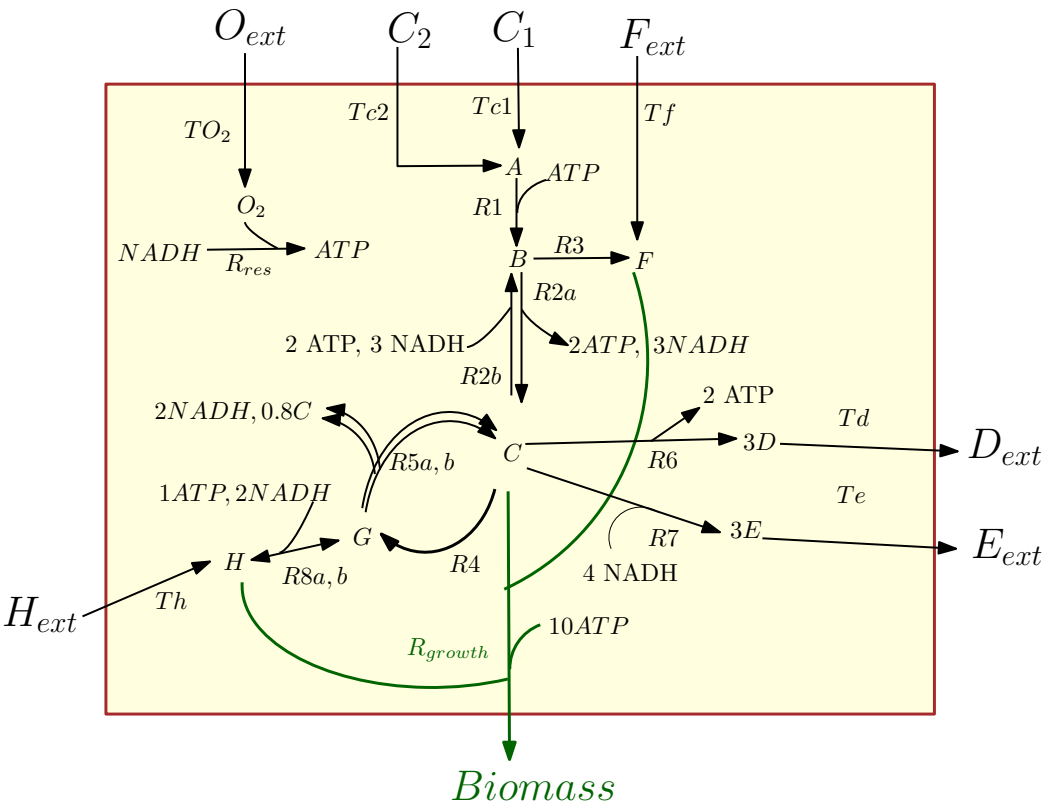


Figure 5.4: A simplified core carbon metabolic network.

We consider the model of the simple metabolic-genetic network from [33] at diauxie-switch scenario that has been introduced in Section 2.5, Chapter 2. We use the Matlab function `ode23s` with tolerances $RTOL = ATOL = 10e^{-6}$ and initial conditions

$$x_0 = [10, 10, 0, 50, 0, 0, 0, 0.003, 0, 1, 0, 0, 0, 0.2, 0, 0.03, 6, 5, 0]$$

to compute a numerical solution of the ODE system (2.3) with the parameter values Table A.1 in the time interval $[0,5]$ using a non-equidistant output time-grid. This numerical solution yields the snapshot matrix $\mathbb{X}$. We use the Matlab function `svd` to calculate the singular value decompositions of $\mathbb{X}$. The singular values are depicted in Figure 5.5.

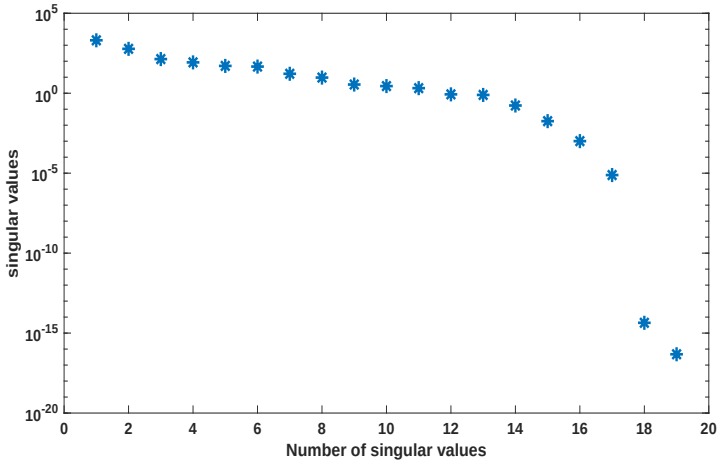


Figure 5.5: Singular values of snapshot matrix $\mathbb{X}$ for the simple metabolic-genetic network example of [33].

The behavior of the model order reduction method strongly depends on the decay of the singular values of the snapshot matrix. We can observe gradually decaying singular values with a strong decay for the smallest singular values indicating that neglection of these will not result in any considerable loss of information in the reduced-order model.
The original model has a dimension of $n_m = 19$. We compute a reduced-order model of dimension $n_k = 16$ using the POD approach. The results of a simulation of the reduced-order system in comparison with the simulation results of the original system for some selected components of the state vector are depicted in Figure 5.6. We can observe that the main features of the dynamical behavior of the original model states are preserved in reduced-order models.

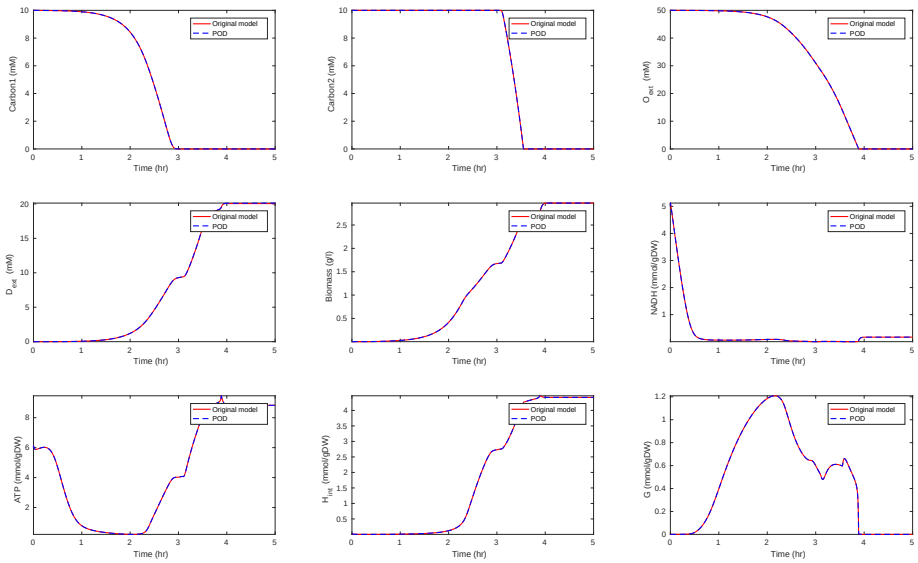


Figure 5.6: Comparison of some of the model components in the original model and the reduced order model (POD) of dimension $n_k = 16$.

In a second scenario we reduce the dimension to $n_k = 15$. The results are depicted in Figure 5.7.

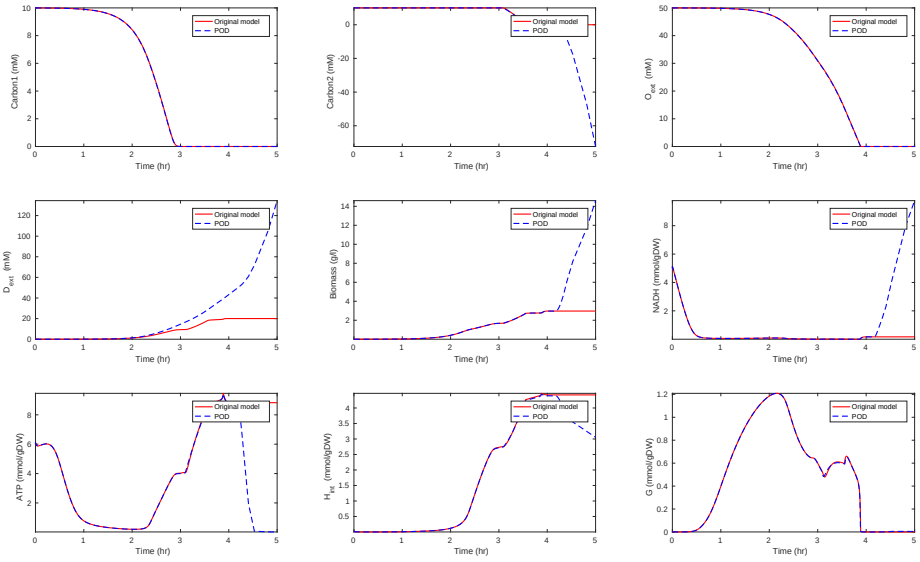


Figure 5.7: Comparison of some of the model components in the original model and the reduced order model (POD) of dimension $n_k = 15$.

In a third scenario we reduce the dimension to $n_k = 14$. The results are depicted in Figure 5.8.

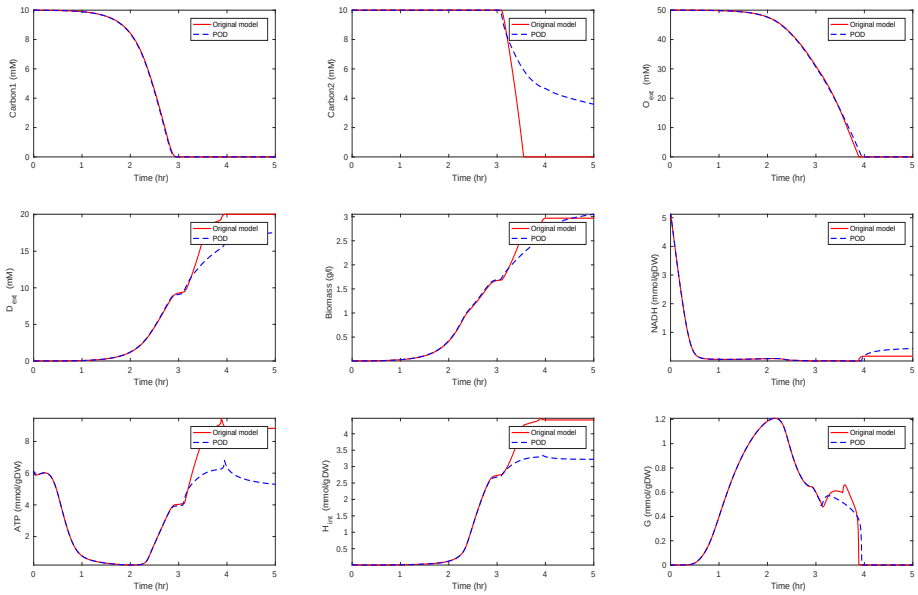


Figure 5.8: Comparison of some of the model components in the original model and the reduced order model (POD) of dimension $n_k = 14$.

We observe that some curves still match good with the original model, while the solutions move off after $t = 4$ in the second scenario and after $t = 3$ in the third scenario. The behavior of C_1, O_{ext} and G still fits well for all reduced models. In general, we can say that the external components still show the same behavior. Since biomass expresses the growth of the cell, it is an important component of the network that should also be preserved. It is observed that the reduced model preserves the biomass behavior quite well, such that we get approximately the same growth. We compare the time costs for the simulations, see Table 5.1. For evaluating computing times, we use the Matlab function `timeit`.

	original model	POD ROM
Scenario 1 ($n_k = 16$)	1.2s	6.04s
Scenario 2 ($n_k = 15$)	1.13s	25.3s
Scenario 3 ($n_k = 14$)	1.14s	4.8s

Table 5.1: Comparison of computing times for the three different scenarios.

We also study the effects of reducing snapshot matrix columns dimension. As we know, the ode solver function in Matlab uses a random time steps for the numerical solution of the differential equations. In most cases, the dimension of snapshot column big, so we tried to reduce the number of snapshot columns by using the solution in specific time points and apply the model reduction method on this snapshot matrix that has fewer columns comparing to the original solution. We notice that there are no significant changes in singular value decomposition behavior and the trajectory of the metabolites in the reduced model, see appendix A.3.

5.2.2 Kinetic model of Lactococcus lactis metabolism

In [30], the existing kinetic model of L. lacctis central metabolism is extended to include industrially relevant production pathways such as mannitol and 2,3-butanediol. The Michaelis-Menten kinetics are used for reaction rates, where the parameters were estimated from multivariate time series metabolite concentrations obtained by them through in vivo Nuclear Magnetic Resonance (NMR).

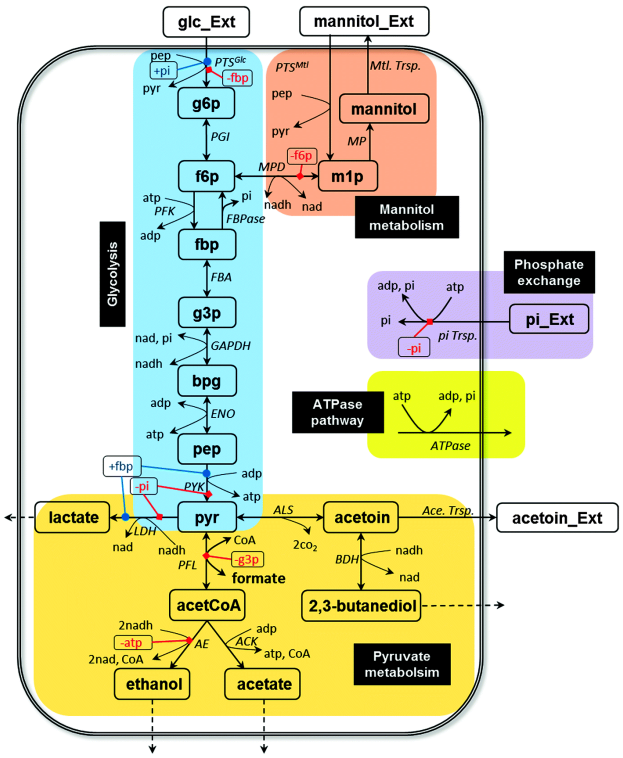


Figure 5.9: Schematic network representation of the central metabolism of L. lactis. The figure is taken from [30], License ID: 1022592-1.

variabel	initial value
glc_{ext}	40
$g6p$	0
$f6p$	0
fbp	15.3
$g3p$	0
bpg	1.26
pep	2.48
pyr	0
$acetCoA$	0
$acetoin$	0
$acetoin_{ext}$	0
$butanediol$	0
$m1p$	0

variabel	initial value
$mannitol$	0
$mannitol_{ext}$	0
$lactate$	0
$ethanol$	0
$acetate$	0
pi	38.26
pi_{ext}	50
atp	4.89
nad	4.67
CoA	1
$formate$	0
adp	20.39
$nadh$	$2.3 \cdot 10^{-6}$

Table 5.2: Initial values for the kinetic model of L. lactis metabolism.

The ODE model equations are taken from the BioModels Database [1]. Since the authors interested in glc_{Ext} concentration, lactate, and ATP, we apply the model order reduction method to the kinetic model and by taking care of the behavior of these metabolites to be the same after and before the reduction. We reduce the dimension of the dynamical system from 26 to 12. The snapshot matrices $\mathbb{X}$ is again obtained from a simulation of the original model equations over the time interval $[0, 150]$ using the Matlab solver `ode23s` with default settings with initial values given in Table 5.2. The singular values are depicted in Figure 5.10.

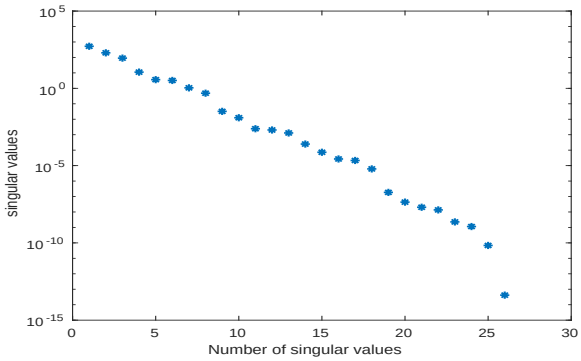


Figure 5.10: The singular values of snapshot matrix $\mathbb{X}$ for the kinetic model of Lactococcus lactis metabolism [30].

[1] http://identifiers.org/biomodels.db/BIOMD0000000572

In a first scenario we reduce the dimension of the dynamical system from $n_m = 26$ to $n_k = 12$. In Figure 5.11, the results of the simulation of the POD (reduced model) in comparison with the original full-order model are illustrated. We can see that the curves for the reduced-order models well match the curves for the original model.

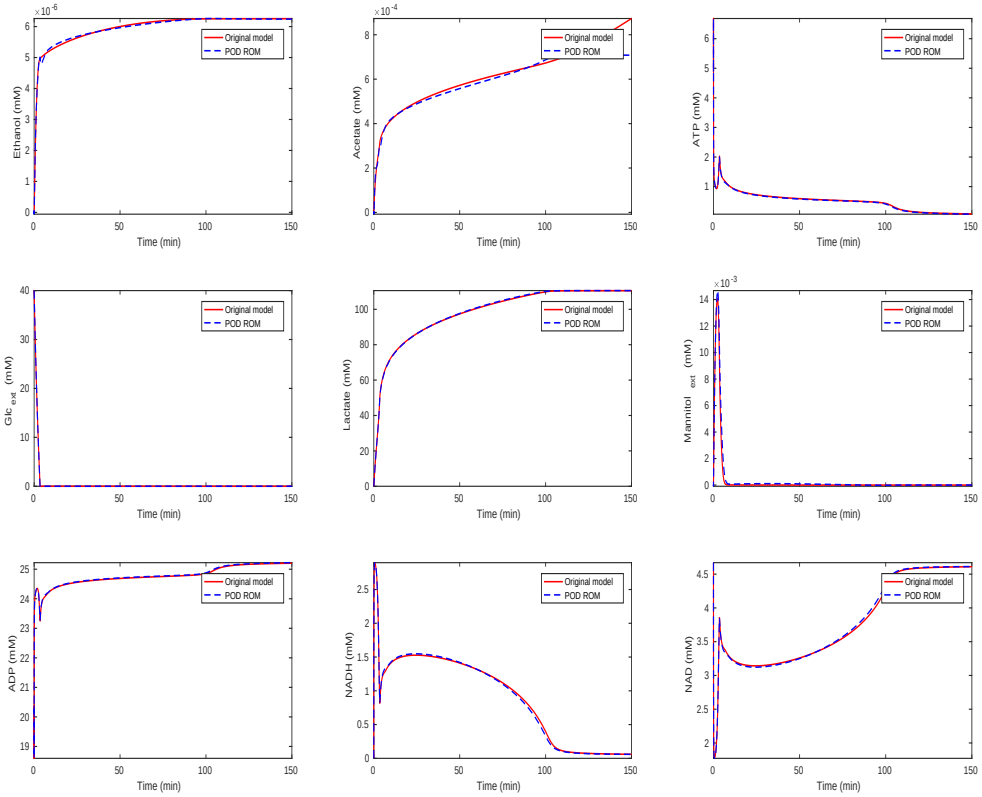


Figure 5.11: Comparison of the behavior of some metabolites from the model of L. lactis metabolism for the reduced-order and original model (Scenario 1).

In the second scenario, we reduce the system to size $n_k = 11$. The results are depicted in Figure 5.12. In this case, the POD model do not capture the dynamical behavior of the original model well. In particular, the concentration of external glycolysis, lactate, and ATP matches the behavior of the original model. A comparison of computation times is given in Table 5.3. For both scenarios the relative information content is $I(12) \approx I(11) \approx 1$. Again, we see the additional computational effort in the computation of the reduced-order model.

	original model	POD ROM
Scenario 1 ($n_k = 12$)	0.3	1.5
Scenario 2 ($n_k = 11$)	0.3	0.9

Table 5.3: Comparison of computing times for the different scenarios.

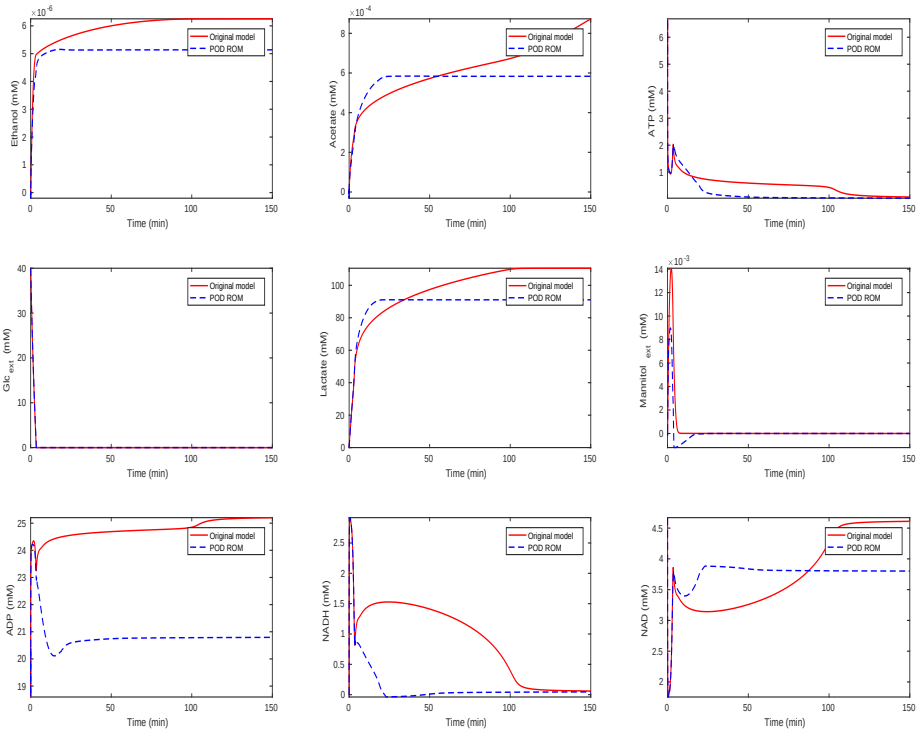


Figure 5.12: Comparison of the behavior of some metabolites from the model of L. lactis metabolism for the reduced order and original model (Scenario 2).

5.2.3 Kinetic model of yeast metabolic network

In [138], a workflow for converting metabolic reconstructions into large-scale kinetic models of yeast metabolism has been developed. Its purpose is to take available data sets, perform a thorough analysis of the parameter constraints, and then to produce the kinetic model using large data integration. Here, we apply the order reduction method to the large-scale yeast model as described in [138]. The ODE model equations are taken from the BioModels database [2].

[2] http://identifiers.org/biomodels.db/BIOMD0000000496

The snapshot matrices $\mathbb{X}$ is again obtained from a simulation of the original model equations over the time interval $[0,3000]$ using the Matlab solver `ode23s` with default settings with initial values given in Table 5.2. The singular values are depicted in Figure 5.13.

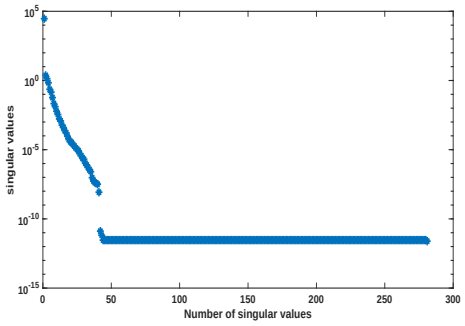


Figure 5.13: Singular values of snapshot matrix $\mathbb{X}$ for kinetic model of yeast metabolic network of [138].

The original ODE system is of dimension $n_m = 281$ and we reduce it to $n_k = 45$ (Scenario 1) and $n_k = 40$ (Scenario 2). The results of the simulations are given in Figure 5.14 and 5.15.

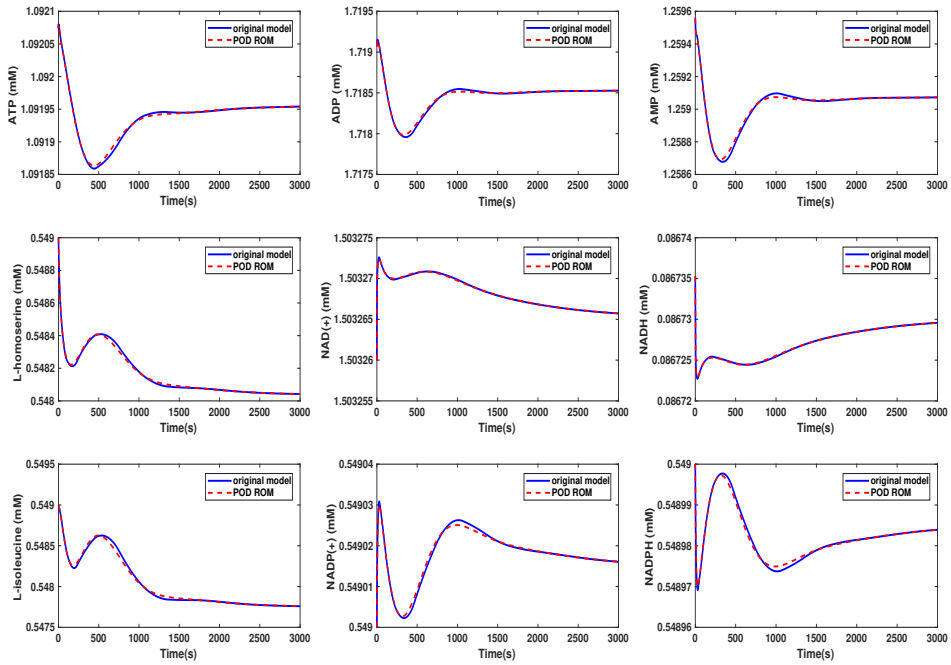


Figure 5.14: Comparison of the behavior of some metabolites in the yeast model for the original and the POD reduced model (Scenario 1).

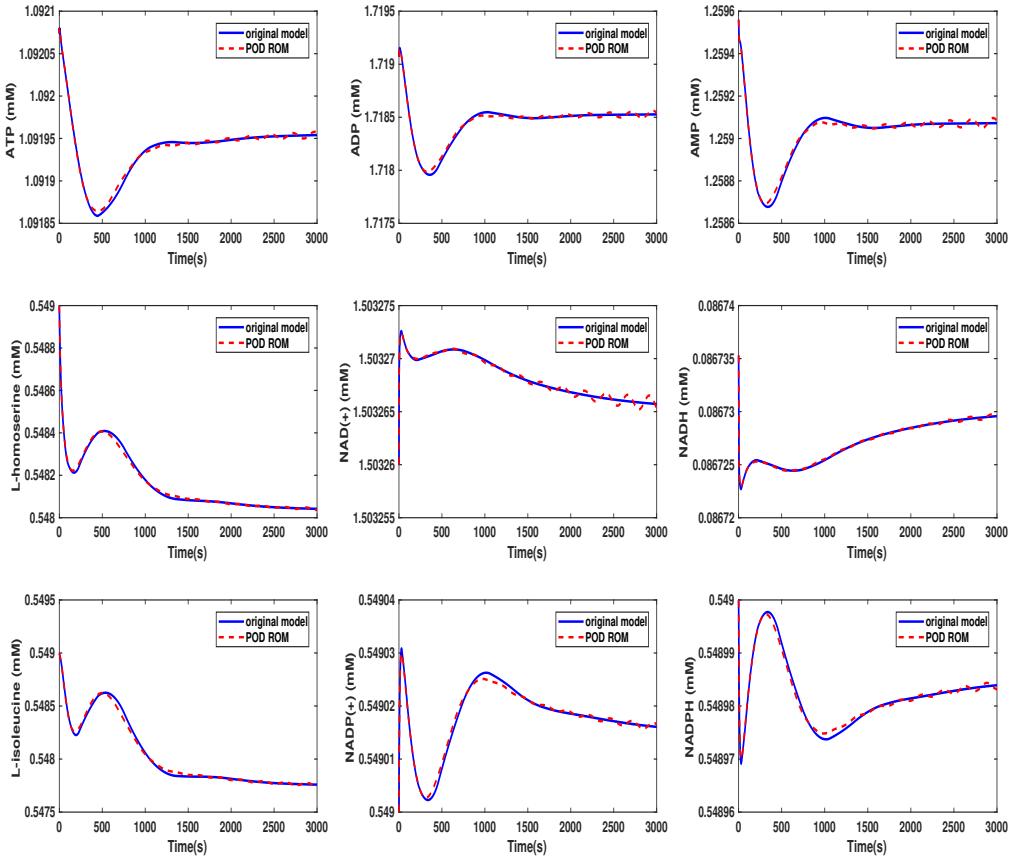


Figure 5.15: Comparison of the behavior of some metabolites in the yeast model for the original and the POD reduced model (Scenario 2).

A comparison of computing times is given in Table 5.5

	original model	POD ROM
Scenario 1 ($n_k = 45$)	0.28s	0.20s
Scenario 2 ($n_k = 40$)	0.28s	0.46s

Table 5.4: Comparison of computing times for the two different scenarios.

We can see that in Scenario 1, the metabolites show the same behavior in both models. We select here the same metabolites as were presented in the supplementary information of [138]. In Scenario 2, the behavior of the metabolites is still good preserved, although we

can see that some of the curves start to oscillate. These oscillations increase if we further decrease the dimension of the reduced-order model until the numerical simulation becomes unstable. The time cost of the simulation could be lowered by a factor of by approximately 3/4 in Scenario 1.

5.2.4 Kinetic model of E. coli metabolic network

A method for the generation of genome-scale kinetic models of E. coli organism from reconstruction data has been proposed in [134]. Building a kinetic model requires kinetic parameters, fluxes, and rate laws. In [134] a small kinetic model has been used with the presence of experimental data, which has then been extended using typical estimates in cases where experimental data is not available. We apply the POD method to the generated kinetic model of E. coli as presented in [134] containing $m = 402$ state variables and we compare the original and reduced model for some of the metabolites behavior. The ODE model equations are taken from the BioModels database [3].

The snapshot matrix $\mathbb{X}$ is obtained from a simulation of the ODE model over the time interval $[0, 300]$ using the Matlab solver `ode23tb` with setting $ATOL = 10^{-3}$ and initial values are taken from the database model. The behavior of the singular values of the snapshot matrix is depicted in Figure 5.16. We can observe a fast decay in the singular values with a large number of values indicating that neglecting these values will not result in any considerable loss of information in the reduced-order model.

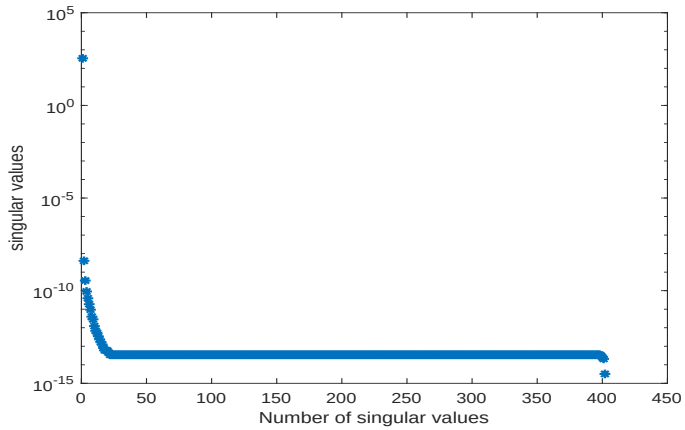


Figure 5.16: Singular values of snapshot matrix $\mathbb{X}$ of E. coli metabolic network of [134].

The original ODE system is of dimension $n_m = 402$ and we reduce it to $n_k = 40$. The results of the simulations are given in Figure 5.17 and 5.18.

[3] http://identifiers.org/biomodels.db/BIOMD0000000469

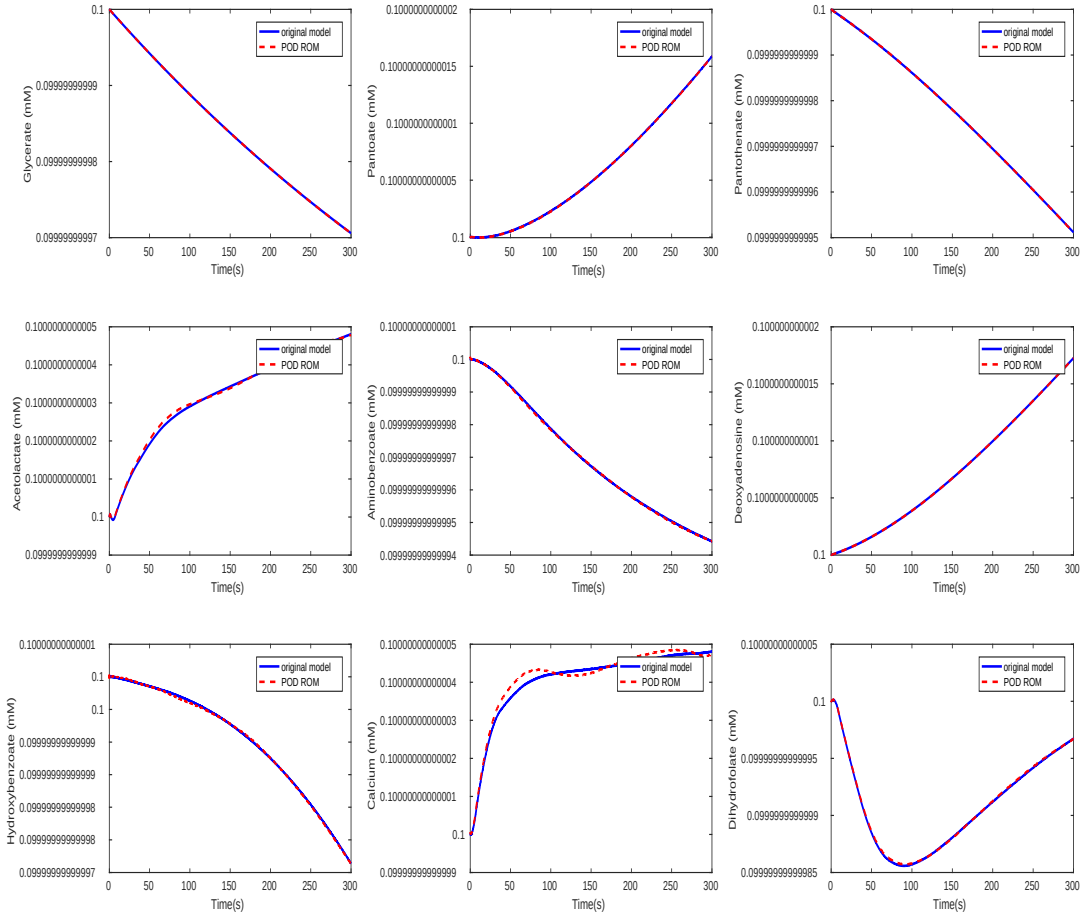


Figure 5.17: Comparison of the behavior of some metabolites in the E. coli model for the original (in blue) and the POD reduced model (in red).

We select here random metabolites for comparing. We can see that the metabolites show the same behavior in both models. The time cost of the simulation in reduced model could be lower by a factor of by approximately 1/6 compared to the time cost of the original model.

	original model	POD ROM
Scenario ($k = 45$)	0.10s	0.014s

Table 5.5: Computing times for the POD method.

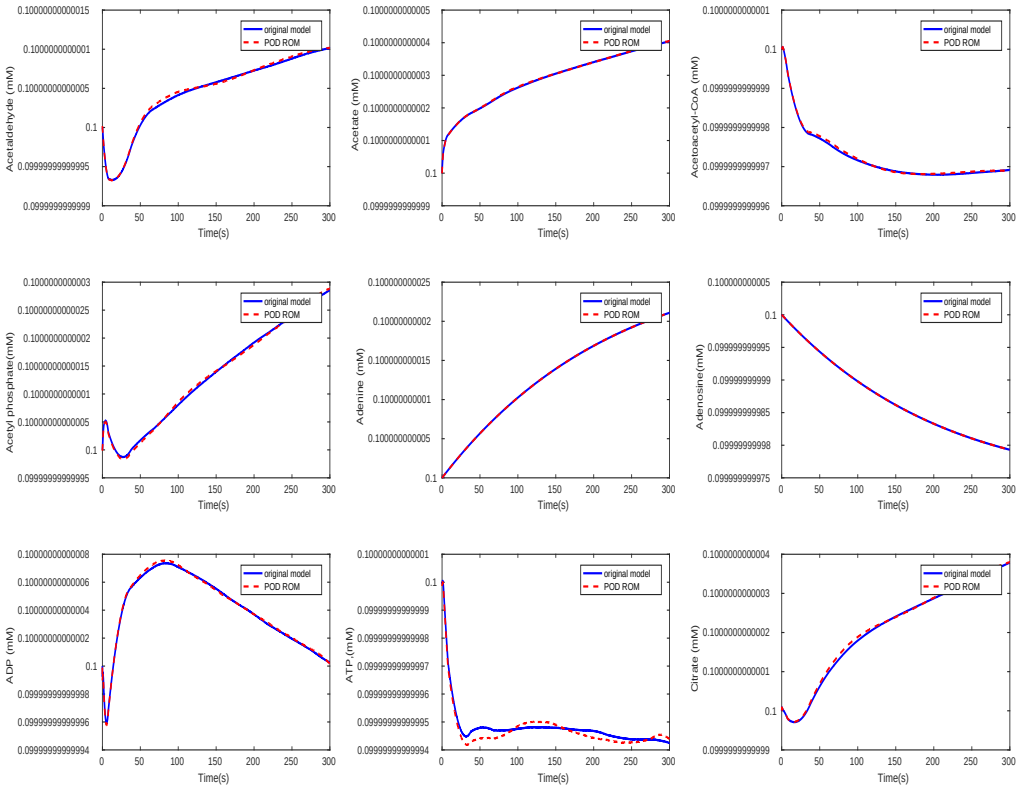


Figure 5.18: Comparison of the behavior of some metabolites in the E.coli model for the original (in blue) and the POD reduced model (in red).

We conclude from the previous discussions that the time cost of the POD method is small when we applied to large models, as in examples 3, 4, compared to examples 1, 2. Thus, the POD method is more effective in large models. The computational efficiency in the simulation can be increased using the POD-reduced model if we have large-scale models as examples presented in Sec 5.2.3 and Sec 5.2.4, while for small-scale examples as in Sec 5.2.1 and Sec 5.2.2 the dynamical behavior is preserved in the POD reduced models, however, there is no pay off when it comes to computational times.

5.3 POD for kinetic model with different initial conditions

In this section, we study the POD method for a dynamical system with different initial conditions. In a kinetic system model, different initial conditions can be used to describe different scenarios or different modes of a biological system. We follow the same

steps of the POD approach as in Section 5.1. At first, we compute the snapshot matrix $\mathbb{X} = [x(t_1), x(t_2), \ldots, x(t_{n_t})] \in \mathbb{R}^{n_m \times n_t}$ by solving the system of *ODEs* (5.1) with the initial conditions x_0. Then we compute another snapshot matrix $\mathbb{Y} = [x(t_1), x(t_2), \ldots, x(t_{n_t})] \in \mathbb{R}^{n_m \times n_t}$ with the different initial conditions y_0. We combine the two snapshot matrices $\mathbb{X}$, $\mathbb{Y}$ in a snapshot matrix $\mathbb{Z}$. The snapshot matrix $\mathbb{Z}$ is given in the following form

$$\mathbb{Z} = [\mathbb{X} \;\; \mathbb{Y}] \in \mathbb{R}^{n_m \times n_l}, \;\; n_l = 2n_t.$$

Then, by applying the singular value decomposition (*SVD*), we obtain

$$\mathbb{Z} = U \begin{bmatrix} \Sigma_{\hat{r}} & 0 \\ 0 & 0 \end{bmatrix} V^{\mathsf{T}},$$

where $U \in \mathbb{R}^{n_m \times n_m}$ and $V^{\mathsf{T}} \in \mathbb{R}^{n_l \times n_l}$ are unitary matrices with orthonormal columns. The term $\Sigma_{\hat{r}} = diag(\hat{\sigma}_1, \hat{\sigma}_2, ..., \hat{\sigma}_{\hat{r}}) \in \mathbb{R}^{\hat{r} \times \hat{r}}$ is a matrix with a real, non negative entries on the diagonal and zeros off the diagonal and $\hat{r} = rank(\mathbb{Z})$. Then by truncating the most dominant singular vectors and projecting the original system onto the lower-dimensional subspace, we obtain a reduced order model.

5.3.1 Application to kinetic model

In the following, we apply the idea to the kinetic model of the metabolic-genetic network from [33] for different scenarios. The scenarios are diauxic switch and aerobic/anaerobic-diauxie and both scenarios can be predicted using the kinetic model by prescribing specific initial values.

Diauxic switch scenario

We consider the model of the simple metabolic-genetic network from [33] at diauxie-switch scenario that has been introduced in Section 2.5, Chapter 2. We use the Matlab function `ode23s` with tolerances $RTOL = ATOL = 10^{-6}$ and initial conditions

$$x_0 = [10, 10, 0, 50, 0, 0, 0, 0.003, 0, 1, 0, 0, 0, 0.2, 0, 0.03, 6, 5, 0]$$

to compute a numerical solution of the ODE system (5.1) with parameters in the time interval $[0, 5]$ using a non-equidistant output time-grid. This numerical solution yields the snapshot matrix $\mathbb{X}$.

Aerobic/Anaerobic-diauxie scenario

We consider aerobic/anaerobic-diauxie scenario that has been introduced in Section 3.5. We use the Matlab function `ode23s` with tolerances $RTOL = ATOL = 10^{-6}$ and initial conditions

$$y_0 = [0, 10, 0, 2, 0, 0, 0, 0.0008, 0, 1, 0, 0, 0, 0.2, 0, 0.03, 6, 5, 0]$$

to compute a numerical solution of the ODE system (5.1) with parameters in the time interval $[0, 5]$ using a non-equidistant output time-grid. This numerical solution yields the snapshot matrix $\mathbb{Y}$.

We use the Matlab function `svd` to calculate the singular value decompositions of $\mathbb{Z} = [\mathbb{X}\ \mathbb{Y}]$. The singular values are depicted in Figure 5.19.

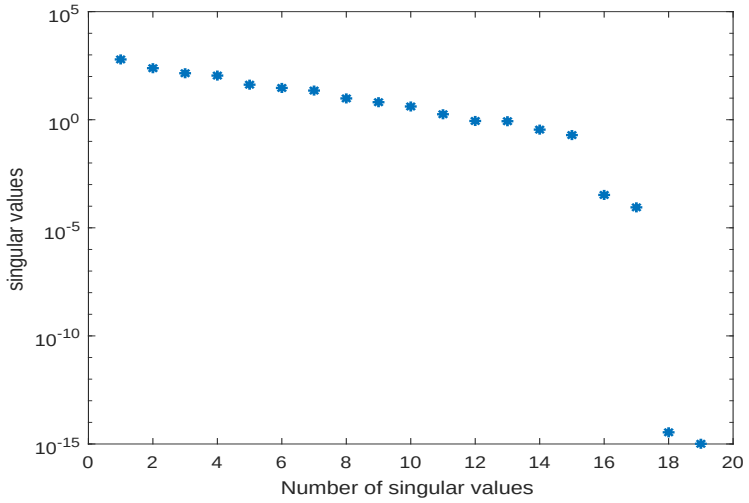


Figure 5.19: Singular values of snapshot matrix $\mathbb{Z}$ for the simple metabolic-genetic network example of [33].

The original model has a dimension of $n_m = 19$. We compute a reduced-order model of dimension $n_k = 17$ using the POD method. The results of a simulation of the reduced-order system in comparison with the simulation results of the original system are depicted in Figures 5.20, 5.21 for the two different scenarios diauxic switch and aerobic/anaerobic-diauxie, respectively.

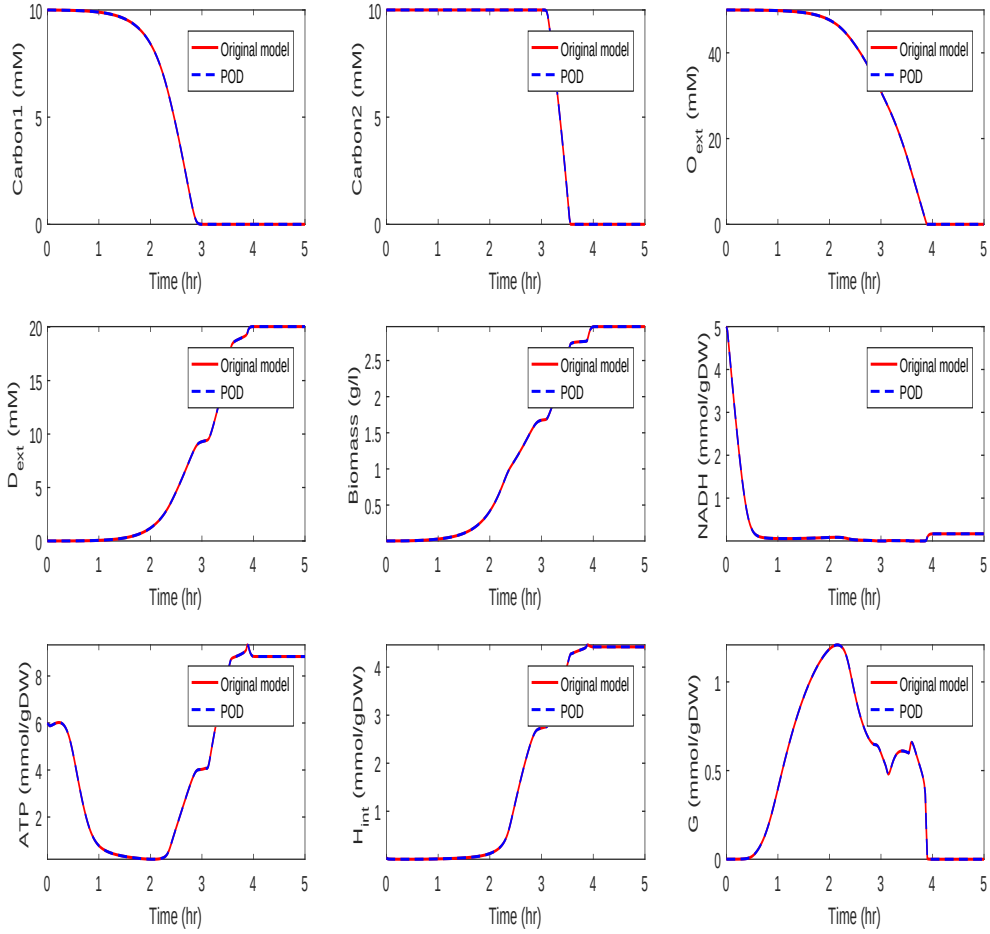


Figure 5.20: Comparison of the behavior of some metabolites in the diauxic switch scenario for the original and the POD reduced model.

We can observe that the main features of the dynamical behavior of the original model states are preserved for both scenarios using the same reduced order model. This is an important result, as it shows that as long as the snapshot matrix provided the essential information a reduced order model has to be computed only once, and can then be used to predict the dynamical behavior of the system for different scenarios (for our specific model). The magnitude by which the system can be reduced depends on the decay of the singular values. Singular values of magnitude 10^{-15} can be neglected, while singular values of order 10^{-4} still contain important information. For large-scale network as in Sec 5.2.3 and Sec 5.2.4 the difference in the dimension of the reduced and the original model can be expressed to be much larger.

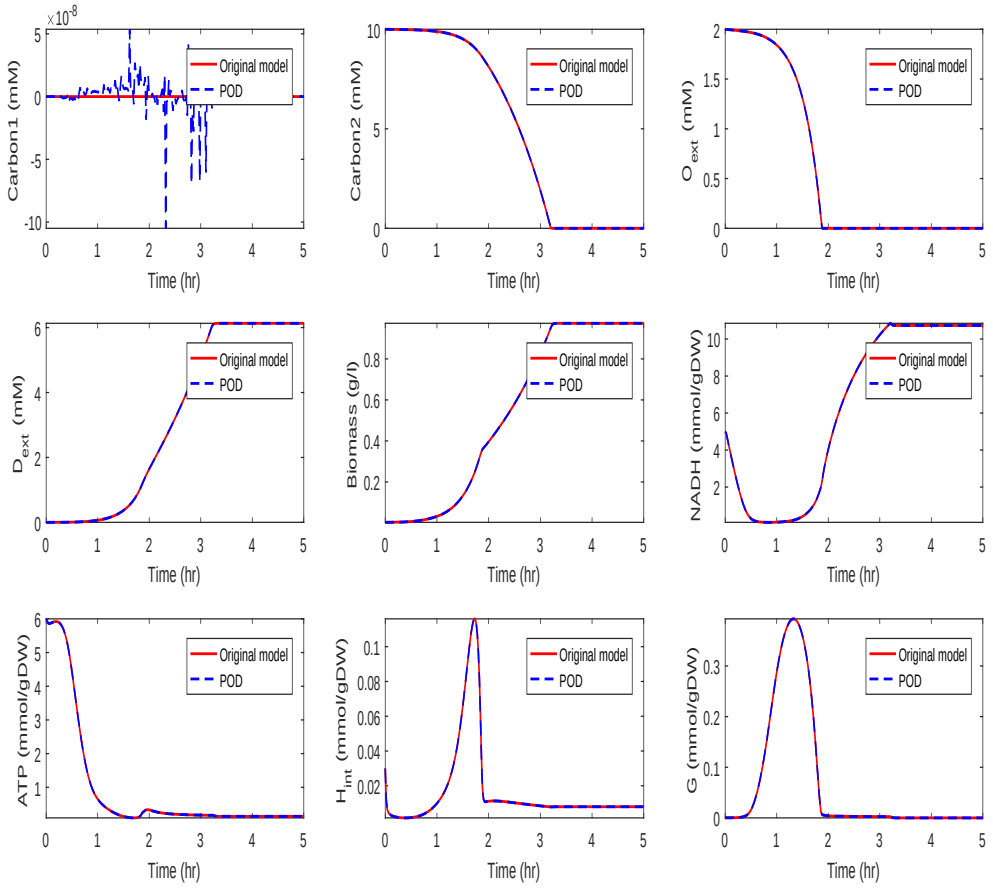


Figure 5.21: Comparison of the behavior of some metabolites in the aerobic/anaerobic-diauxie scenario for the original and the POD reduced model.

5.4 Conclusion

In this chapter, we have discussed the model reduction of a kinetic model using the POD approach. Since kinetic models of chemical networks can be large, the possibility to obtain a reduced-order model that replicates the desired dynamical behavior is of vital importance. We have applied the model reduction techniques to several examples. It has to be noted that the computation of a reduced-order model requires some additional computational effort. However, the computation of the snapshots and the SVD usually has to be computed only once (often this is called the offline phase), while the simulations of the reduced-order models are run several times, e.g., within an optimization process (in the so-called online phase). Thus, the effort for setting up a reduced order model will pay off if we can reduce a large-scale

model to a much smaller dimension and running simulations of the reduced-order model for a long time. Moreover, the efficiency of the model order reduction strongly depends on the decay of the singular values of the snapshot matrix. How far the model can be reduced also depends on the application and on what components of the solution one is interested in. In addition, we have predicted the behavior of the network for different scenarios (diauxic switch and aerobic/anaerobic diauxie) with the same reduced order model.

Chapter 6

Conclusions

This thesis is dedicated to studying the dynamics of metabolic-genetic networks using continuous models. Many approaches successfully studied these networks, e.g., the regulatory flux balance analysis (rFBA). The rFBA model is based on steady-state assumptions. Thus the rFBA model handles only external metabolites. However, in real systems, the cell is not in a steady-state, but is in a dynamic state. Thus, our first contribution was to introduce a kinetic model that mimics the rFBA model of the metabolic-genetic network studied in [33] to overcome the limitations of the rFBA model. The kinetic model gives a full picture of the dynamical system by studying the dynamics for every component (internal and external) over a continuous time interval. The kinetic model we introduced is formed of a system of differential equations with unknown parameters, e.g., kinetic reaction rate constants. So we performed a parameter estimation technique using the *Data2Dynamics* toolbox to obtain the parameter values. By studying the kinetic model, the behavior of the components of the cell becomes clear and easy to analyze. In addition, different theories can be applied to the kinetic model, e.g., model reduction and parameter estimation, etc.

Another approach that can be used to study metabolic-regulatory networks is the hybrid model. The hybrid model of a metabolic-regulatory network has been studied in [95]. The hybrid model becomes more complicated if the number of regulatory proteins is large since the modeling of regulatory rules and detecting of events becomes very difficult. It means that the increase in the number of regulatory rules leads to exponential increase in the number of modes and events by 2^n where n = 1, 2, ..., is the number of regulatory rules. We aimed to introduce a continuous model that mimics the hybrid model studied in [95] because a continuous model is easy to handle. We have successfully shown that by using our continuous model, we can more easily obtain the same results as the results produced by the hybrid model for a metabolic regulatory network. This is because we have used the Hill function to

express the regulatory rules instead of the boolean functions in the hybrid model.

In metabolic-genetic networks, metabolic reactions occur at rates of seconds or less (fast variables), while gene expression usually takes between minutes and hours to complete (slow variables). The discrepancy in the time of metabolic and regulatory reactions leads to a model reduction topic known as time scale separation technique. A good deal of research has discussed the time scale separation technique for model reduction. However, from our point of view, there were no clear guidelines for how to apply it to kinetic models. Our contribution was to explain and clarify the method and the required conditions, e.g., Tikhonov's theorem, which is needed to apply this technique. By following the work of [62], we have applied that technique to different metabolic-genetic networks and obtained a reduced-order model that predicts the same behavior of the full order model. The advantage of the time scale separation method is that it preserves the dynamics of the original system. However, the conditions of Tikhonov's theorem are difficult to satisfy for large scale dynamical systems, so we have suggested to apply another technique for model reduction called the proper orthogonal decomposition (POD).

We have discussed the model reduction using the POD technique for kinetic models of biochemical networks. We illustrated the POD technique on some examples of kinetic network models. We have observed that the reduction of the model order depends on the decay of the singular value decomposition. We have computed the time cost of the full and reduced model. The computation of the reduced-order model requires additional computational effort. However, we have observed that in large scale models, the time cost in the simulation can be significantly reduced using the reduced-order method. In addition, we succeed in using the POD method to compute a reduced-order model for different initial conditions (diauxic switch and aerobic/anaerobic-diauxie scenarios) of the metabolic-genetic network. We can say that the POD method works well for large-scale systems, but the dynamics are given for the surrogate model that is obtained via a projection of the original model. Thus, it is difficult to predict the contribution of a specific single component.
We believe that, in the end, the kinetic or continuous model is the best mathematical form to study cell behavior in case one wants to analyze the full components of the cell. Although the kinetic model provides the full picture of the dynamics of the cell components, the process of finding the value of the parameters is still a challenge in system biology.

	Characteristics	Limitations
rFBA model	It does not need kinetic parameter values	It cannot predict internal metabolite concentrations
	It handles only external metabolites	It cannot use for modeling dynamic behavior
	It relies on stoichiometric characteristics (studied in Covert et al., 2001)	It does not uniquely specify the fluxes
Kinetic model	It used for modeling dynamic behavior	It needs kinetic parameter values
	It handles with external and internal metabolites (studied in the thesis)	
Hybrid model	It consists of ODEs and the boolean functions (studied in Lin et al., 2019)	Hard to apply to large scale systems (i.e., an increase in the number of events and modes)
Continuous model	It consists of ODEs and the Hill functions Easy to handel (studied in the thesis)	-

Model reduction techniques		
	Advantages	Disadvantages
Time scale separation	It preserves the dynamics of the original system	The required conditions are difficult to satisfy (e.g., Tikhonov's theorem)
	It requires two different time scales fast and slow	
POD	It does not need to satisfy Tikhonov's theorem	The dynamics are given for a surrogate model
	It works well for large scale system	
	It is useful in optimization processes	
	It computes a reduced model for different initial conditions	

Figure 6.1: Main characteristics of different mathematical models and model order reduction methods used in the thesis.

Appendix A

A.1 Parameter's values

Constant rates	Value	Unit
k_{M1}	0.38	mM
k_{M2}	0.38	mM
k_{M3}	6.2	mM
k_{M4}	$5.6 \cdot 10^{-5}$	mM
k_{M5}	0.00011	$mmol/gDW$
k_{M6}	10^{-5}	$mmol/gDW$
k_{M7}	41	mM
k_8	980	$mmol^3/(gDW^3 \cdot hr)$
k_9	1000	$mmol^2/(gDW^2 \cdot hr)$
k_{10}	300	$mmol/(gDW \cdot hr)$
k_{11}	140	$mmol^3/(gDW^3 \cdot hr)$
k_{12}	23	$mmol/(gDW \cdot hr)$
k_{13}	25	$mmol/(gDW \cdot hr)$
k_{14}	1.6	$mmol/(gDW \cdot hr)$
k_{15}	1000	$mmol/(gDW \cdot hr)$
k_{16}	13	$mmol/(gDW \cdot hr)$
k_{17}	2.9	$mmol^2/(gDW^2 \cdot hr)$
k_{18}	150	$mmol^3/(gDW^3 \cdot hr)$
k_{19}	$7.9 \cdot 10^{-5}$	$mmol/(gDW \cdot hr)$
k_{20}	170	$mmol^2/(gDW^2 \cdot hr)$
ζ	10^{-5}	mM
γ	0.024	$mmol/gDW$
β	0.11	mM
α	290	$mmol/gDW$
ω	1 (assumed)	$gDW/mmol$

Table A.1: The estimated parameters of kinetic model using two data sets of diauxic-switch and aerobic/anaerobic scenarios.

A.2 Bimolecular reactions network

We study a bimolecular reaction metabolic-genetic network such that the network describes two substrates 1 and 2, which are consumed by one reaction to produce the metabolite 3, see Figure A.1.

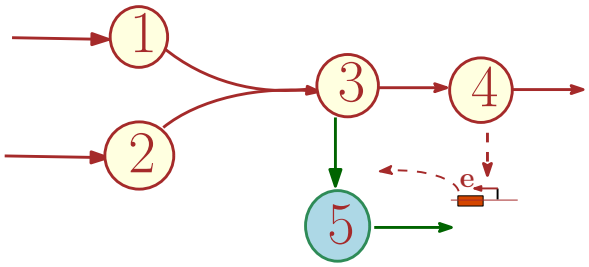


Figure A.1: The pathway converts metabolite 1 and metabolite 2 together into metabolite 3 by one reaction. The enzyme e, catalysis the reaction that converts metabolite 3 into to metabolite 5.

The network is described by a system of *ODEs* using Michaelis-Menten Kinetics expression for metabolites reaction, Hill function for enzyme synthesis, and mass action kinetic for enzyme degradation. The dynamical system is given as follows:

$$
\begin{aligned}
\frac{dm_1}{dt} &= I_1 - \frac{k_{cat1} m_1 m_2}{K_{M1}K_{M2} + K_{M2}m_1 + K_{M1}m_2 + m_1 m_2} e_{1,2\to 3}, \\
\frac{dm_2}{dt} &= I_2 - \frac{k_{cat1} m_1 m_2}{K_{M1}K_{M2} + K_{M2}m_1 + K_{M1}m_2 + m_1 m_2} e_{1,2\to 3}, \\
\frac{dm_3}{dt} &= \frac{k_{cat1} m_1 m_2}{K_{M1}K_{M2} + K_{M2}m_1 + K_{M1}m_2 + m_1 m_2} e_{1,2\to 3} - \frac{k_{cat3} m_3}{K_{M3} + m_3} e_{3\to 4} - \frac{k_{cat4} m_3}{K_{M4} + m_3} e_{3\to 5}, \\
\frac{dm_4}{dt} &= \frac{k_{cat3} m_3}{K_{M3} + m_3} e_{3\to 4} - \frac{\hat{E}_4 m_4}{K_{O4} + m_4}, \\
\frac{dm_5}{dt} &= \frac{k_{cat4} m_3}{K_{M4} + m_3} e_{3\to 5} - \frac{\hat{E}_5 m_5}{K_{O5} + m_5}, \\
\frac{de}{dt} &= k_0 + k_1 \Gamma(m_4) - k_d e,
\end{aligned}
$$

where $e = e_{3\to 5}$, the constant rate values of k_{cat}, K_M, e_N, and export rates have the same value of the example in [89].

We consider the non-dimensionalization system with new variables (x,y) as follows:

$$\begin{aligned}
\varepsilon\frac{dy_1}{d\hat{t}} &= \tilde{I}_1 - \frac{y_1y_2}{1+y_1+y_2+y_1y_2},\\
\varepsilon\frac{dy_2}{d\hat{t}} &= \tilde{I}_2 - \frac{y_1y_2}{1+y_1+y_2+y_1y_2},\\
\varepsilon\frac{dy_3}{d\hat{t}} &= \frac{y_1y_2}{1+y_1+y_2+y_1y_2} - \frac{y_3}{1+y_3} - \frac{\hat{e}}{e_N}\frac{y_3x}{1+y_3},\\
\varepsilon\frac{dy_4}{d\hat{t}} &= \frac{y_3}{1+y_3} - \frac{y_4}{1+y_4},\\
\varepsilon\frac{dy_5}{d\hat{t}} &= \frac{\hat{e}}{e_N}\frac{y_3x}{1+y_3} - \frac{y_5}{1+y_5},\\
\frac{dx}{d\hat{t}} &= \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(y_4) - x,
\end{aligned}$$

where $\Gamma^*(y_4) := \Gamma(K_My_4)$, $\varepsilon = \dfrac{K_Mk_d}{k_{cat}e_N}$, and $\tilde{I}_i = \dfrac{I_i}{k_{cat}e_N}$, $i = 1,2$.

Here, we assume the import rates $\tilde{I}_1, \tilde{I}_2$ have the same value then the first two metabolites have the same differential equation. It means that one metabolite can be written in terms of another one as follows:

$$\frac{dy_1}{d\hat{t}} = \frac{dy_2}{d\hat{t}}, \tag{A.2}$$

by integrating the equation (A.2) over the time $\hat{t}$, we obtain

$$y_1 = y_2 + \delta,$$

where δ is an integral constant, at $\varepsilon = 0$, we express the roots of algebraic equation by $r(x)$. The roots of algebraic equations are given as follows

$$\tilde{I}_1 - \frac{r_1(x)r_2(x)}{1+r_1(x)+r_2(x)+r_1(x)r_2(x)} = 0,$$

$$\frac{\tilde{I}_1 + \tilde{I}_1r_1(x) + \tilde{I}_1r_2(x) + \tilde{I}_1r_1(x)r_2(x) - r_1(x)r_2(x)}{1+r_1(x)+r_2(x)+r_1(x)r_2(x)} = 0,$$

$$\tilde{I}_1 + \tilde{I}_1(r_2(x)+\delta) + \tilde{I}_1r_2(x) + \tilde{I}_1(r_2(x)+\delta)r_2(x) - (r_2(x)+\delta)r_2(x) = 0,$$

$$r_2^2(x)(1-\tilde{I}_1) - r_2(x)(2\tilde{I}_1 + \tilde{I}_1\delta - \delta) - (\tilde{I}_1 + \tilde{I}_1\delta) = 0,$$

$$r_2(x) = \frac{(2\tilde{I}_1 + \tilde{I}_1\delta - \delta) + \sqrt{(2\tilde{I}_1 + \tilde{I}_1\delta - \delta)^2 + 4(1-\tilde{I}_1)(\tilde{I}_1 + \tilde{I}_1\delta)}}{2(1-\tilde{I}_1)},$$

we ignore the negative solution since we study the concentrations that have positive values

$$r_3(x) = r_4(x) = \frac{\tilde{I}_1}{\frac{\hat{e}}{e_N}x + 1 - \tilde{I}_1},$$

$$r_5(x) = \frac{\tilde{I}_1}{\frac{\hat{e}}{e_N}\frac{1}{x} + 1 - \tilde{I}_1},$$

thus the reduced model is given by

$$\begin{aligned} \dot{\bar{x}} &= \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1}\Gamma^*(r_4(\bar{x})) - \bar{x}, \quad \bar{x}(0) = x_0, \\ \bar{y} &= r(\bar{x}). \end{aligned} \tag{A.3}$$

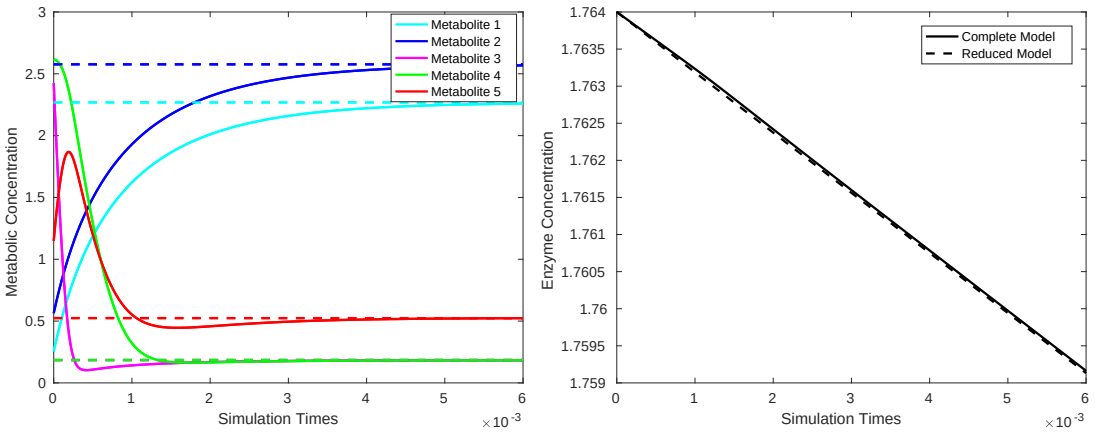


Figure A.2: a The metabolite trajectories of the complete model (solid lines) converge rapidly to that of the reduced model (dashed lines). **b** The trajectory of the enzyme of the complete model remains so close to the reduced model.

We obtain the reduced system of bi-substrates reactions, in particular, in the case of the import rate of the metabolites y_1, y_2 has equal value. The simulation of the complete and the reduced systems were generated in Matlab2016b using function `ode45` with a default setting with initial conditions $(y_1, y_2, y_3, y_4, y_5, x) = (0.255, 0.5638, 2.4255, 2.617, 1.1489, 1.7640)$.

The value of δ is 0.3088 and $\tilde{I}_1, \tilde{I}_2$ have value is $1/2$. It is clear from Figure A.2 that the the trajectory of the original system remains close to the reduced system and the trajectories of the fast variables converge to their quasi steady state after some time.

Now, we will discuss the boundary layer system. The boundary layer system is obtained using the same previous steps of changing the variables to $\hat{y} = y - r(x)$ and substituting $\frac{d\hat{t}}{d\tau} = \varepsilon$. The boundary layer system is given as follows

$$
\begin{aligned}
\frac{d\hat{y}_1}{d\tau} &= \bar{I}_1 - \frac{\hat{y}_1 + r_1(x_0)}{1+\hat{y}_1 + r_1(x_0)} \frac{\hat{y}_2 + r_2(x_0)}{1+\hat{y}_2 + r_2(x_0)}, \\
\frac{d\hat{y}_2}{d\tau} &= \bar{I}_2 - \frac{\hat{y}_1 + r_1(x_0)}{1+\hat{y}_1 + r_1(x_0)} \frac{\hat{y}_2 + r_2(x_0)}{1+\hat{y}_2 + r_2(x_0)}, \\
\frac{d\hat{y}_3}{d\tau} &= \frac{\hat{y}_1 + r_1(x_0)}{1+\hat{y}_1 + r_1(x_0)} \frac{\hat{y}_2 + r_2(x_0)}{1+\hat{y}_2 + r_2(x_0)} - \frac{\hat{y}_3 + r_3(x_0)}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{e}}{e_N} \frac{(\hat{y}_3 + r_3(x_0))x_0}{1+\hat{y}_3 + r_3(x_0)}, \\
\frac{d\hat{y}_4}{d\tau} &= \frac{\hat{y}_3 + r_3(x_0)}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{y}_4 + r_4(x_0)}{1+\hat{y}_4 + r_4(x_0)}, \\
\frac{d\hat{y}_5}{d\tau} &= \frac{\hat{e}}{e_N} \frac{(\hat{y}_3 + r_3(x_0))x_0}{1+\hat{y}_3 + r_3(x_0)} - \frac{\hat{y}_5 + r_5(x_0)}{1+\hat{y}_5 + r_5(x_0)},
\end{aligned}
$$

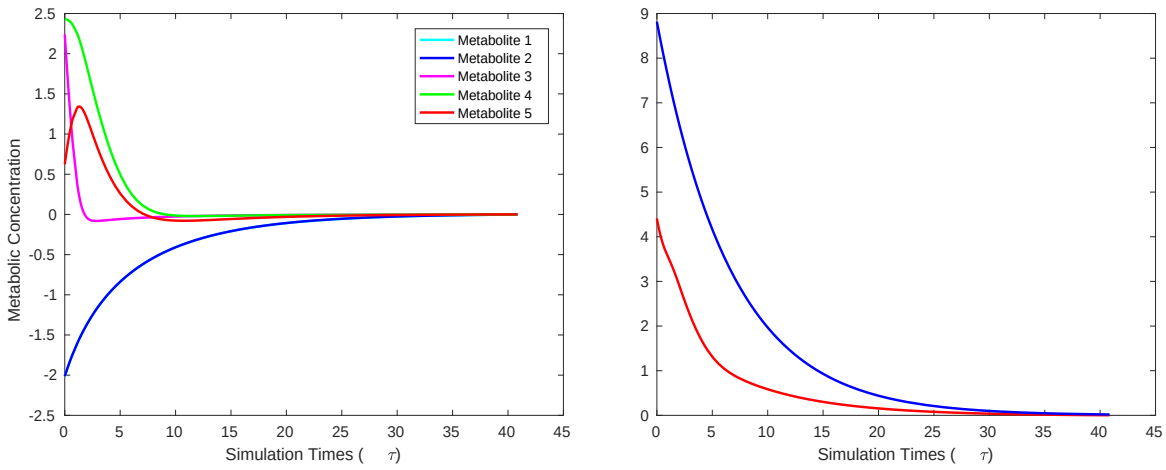


Figure A.3: **a** The metabolites trajectory of the boundary-layer system. **b** The red trajectory is the norm of $\|\hat{m}(\tau)\|$ and the blue trajectory is $\kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|$.

The simulations of the boundary layer system was generated in Matlab (R2016b) using the Matlab function `ode45` with the default setting to solve the corresponding systems of *ODEs* with initial conditions $\hat{y}(0) = y(0) - r(x(0))$. In Figure A.3a, we observe that all the

trajectories of the metabolites of the boundary layer system converge to the equilibrium point $\hat{y} = 0$. The exponential stability of equilibrium point of boundary-layer system by satisfying the inequality in (4.9). By choosing with $\kappa = 2, \psi = 0.15$, we notice that $\|\hat{m}(\tau)\| \leq \kappa e^{(-\psi\tau)} \|\hat{m}(\tau_0)\|$, see Figure A.3b, thus the equilibrium point $\hat{y} = 0$ is exponential stable. Then the error of the reduced and complete system will be of order $O(\varepsilon)$.

A.3 Model reduction by the POD method

Here, we discuss how far the change in the dimension of the snapshot matrix can affect the behavior of the reduced model. The ODEs solver in MATLAB chooses internal time steps to solve *ODE* system. In the kinetic model of the metabolic-genetic network, the snapshot matrix is 19×2313 dimension. The number of 19 indicates to the metabolites and the time steps number is 2313. We will adjust the dimensions of the snapshot matrix and discuss the effect of small-time steps number on the model reduction technique. We reduce snapshot matrix column from 19×2313 to 19×29 dimension. We compute a reduced-order model of dimension $n_k = 16$ using the POD approach. The singular values of the snapshot matrix are shown in Figure A.4.

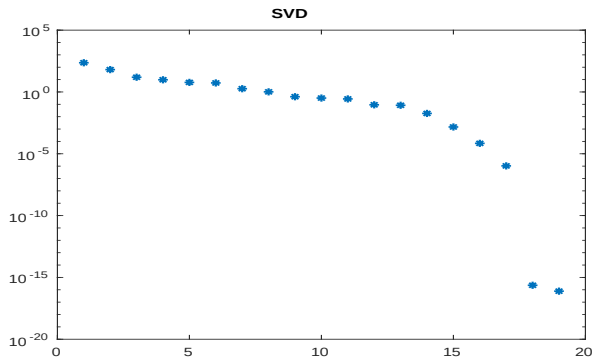


Figure A.4: Singular values of snapshot matrix of dimension 19×29.

The results of a simulation of the reduced-order system in comparison with the simulation results of the original system are depicted in Figure A.5.

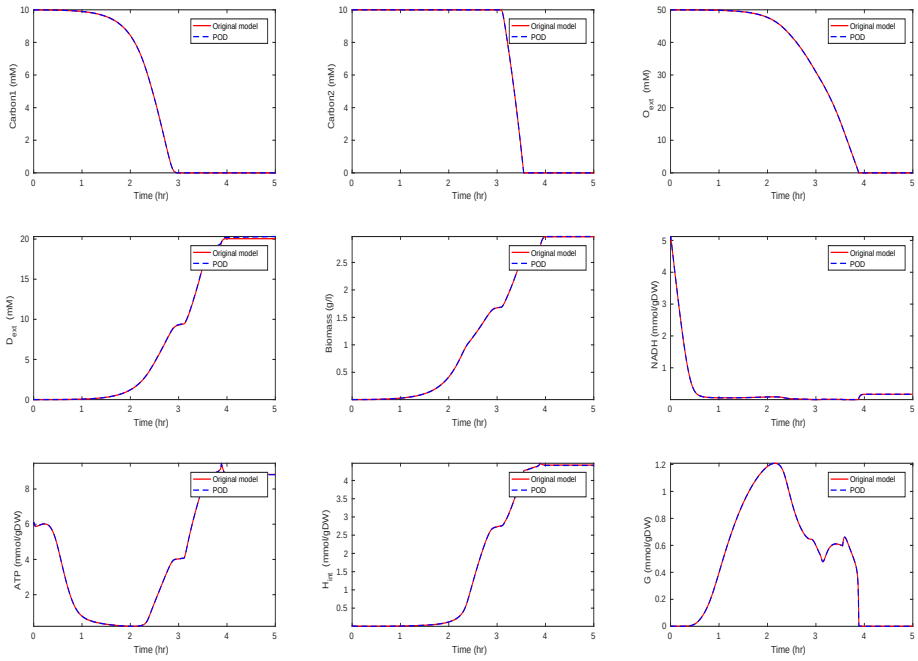


Figure A.5: Comparison of the behavior of some metabolites in the diauxic switch case for the original and the POD reduced model.

Here, we adjust the dimensions of the snapshot matrix 19×2313 to 19×16. We compute a reduced-order model of dimension $n_k = 16$ using the POD approach. The singular values is shown in Figure A.6.

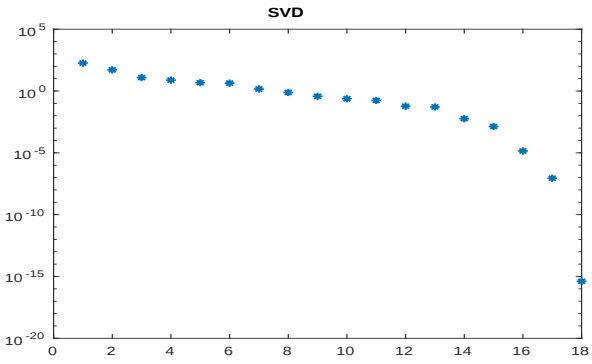


Figure A.6: singular values of snapshot matrix in 19×16 dimension.

The results of a simulation of the reduced-order system in comparison with the simulation results of the original system are depicted in Figure A.7.

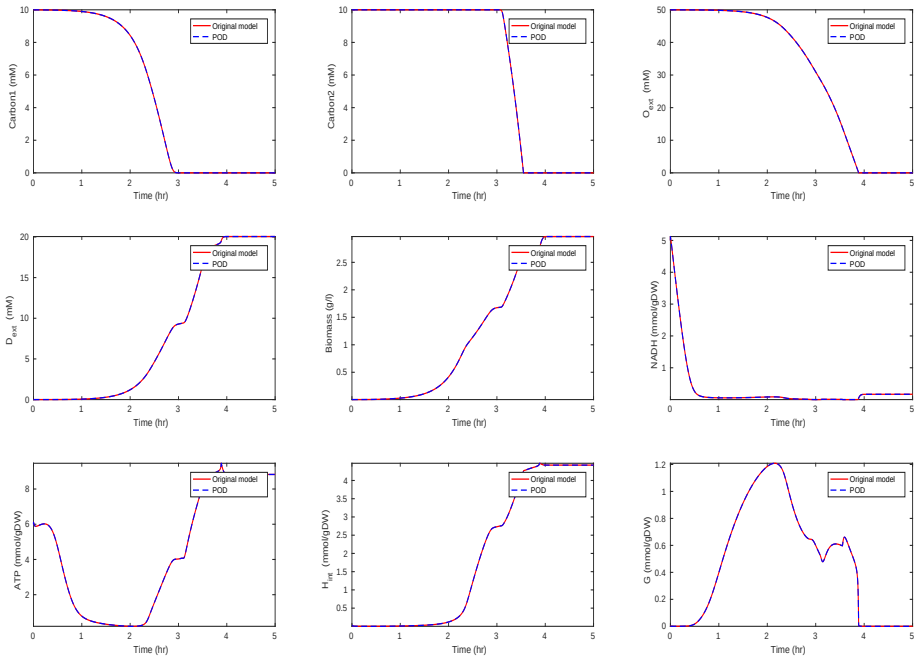


Figure A.7: Comparison of the behavior of some metabolites in the diauxic switch case for the original and the POD reduced model.

From the above results, we can observe that the main features of the dynamical behavior of the original model states are preserved in reduced-order models. There is no significant change in the behavior of the reduced model, and even if we have adjusted the dimension of the snapshot matrix.

Bibliography

[1] Frederick R Adler. *Modeling the dynamics of life: calculus and probability for life scientists*. Nelson Education, 2012.

[2] Konstantin Afanasiev and Michael Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. *Lecture Notes in Pure and Applied Mathematics*, pages 317–332, 2001.

[3] Ravi P Agarwal and Donal O'Regan. *An introduction to ordinary differential equations*. Springer Science & Business Media, 2008.

[4] Mohammad Ahsanullah, B M Golam Kibria, and Mohammad Shakil. Normal Distribution. In *Normal and Student´ st Distributions and Their Applications*, pages 7–50. Springer, 2014.

[5] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.

[6] Naomi Altman and Martin Krzywinski. Points of significance: simple linear regression, 2015.

[7] Rajeev Alur, Costas Courcoubetis, Nicolas Halbwachs, Thomas A Henzinger, Pei-Hsin Ho, Xavier Nicollin, Alfredo Olivero, Joseph Sifakis, and Sergio Yovine. The algorithmic analysis of hybrid systems. *Theoretical computer science*, 138(1):3–34, 1995.

[8] Rajeev Alur, Costas Courcoubetis, Thomas A. Henzinger, and Pei Hsin Ho. Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems. pages 209–229. Springer, Berlin, Heidelberg, oct 1993.

[9] Emmanuel Anane, Peter Neubauer, and M Nicolas Cruz Bournazou. Modelling overflow metabolism in Escherichia coli by acetate cycling. *Biochemical engineering journal*, 125:23–30, 2017.

[10] Andreas Sommer. Numerical Methods for Parameter Estimation in Dynamical Systems with Noise. 2017.

[11] Athanasios C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics, jan 2005.

[12] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A. Kaandorp, and Joke G. Blom. Systems biology: parameter estimation for biochemical models. *FEBS Journal*, 276(4):886–902, feb 2009.

[13] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[14] Ahtchi-Ali Badreddine and Pederseri Henrik. Simulation of Switching Phenomena in Biological Systems. In *Biochemical Engineering for 2001*, pages 701–704. Springer, 1992.

[15] William G Bardsley and Jeffries Wyman. Concerning the thermodynamic definition and graphical manifestations of positive and negative co-operativity. *Journal of theoretical biology*, 72(2):373, 1978.

[16] Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of model reduction methods for parametric systems. 2013.

[17] James O Berger and Robert L Wolpert. The likelihood principle. IMS, 1988.

[18] Åke Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[19] HG Bock, T Carraro, W Jäger, S Körkel, and R Rannacher. *Model based parameter estimation: theory and applications*. 2013.

[20] Hamid Bolouri. *Computational Modeling of Gene Regulatory Networks â A Primer*. World Scientific Publishing Company, 2008.

[21] Rainer Breitling. What is systems biology? *Frontiers in physiology*, 1:159, 2010.

[22] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[23] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.

[24] Christophe Chassagnole, Naruemol Noisommit-Rizzi, Joachim W. Schmid, Klaus Mauch, and Matthias Reuss. Dynamic modeling of the central carbon metabolism ofEscherichia coli. *Biotechnology and Bioengineering*, 79(1):53–73, jul 2002.

[25] Saifon Chaturantabut and Danny C. Sorensen. Nonlinear Model Reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[26] Dominique Chu and David J Barnes. The lag-phase during diauxic growth is a trade-off between fast adaptation and high growth rate. *Scientific reports*, 6:25191, 2016.

[27] WW Cleland. 1 steady state kinetics. In *The enzymes*, volume 2, pages 1–65. Elsevier, 1970.

[28] Charles J Colbourn and Jeffrey H Dinitz. *Handbook of combinatorial designs*. CRC press, 2006.

[29] Kenneth Antonio Connors. *Chemical kinetics: the study of reaction rates in solution*. John Wiley & Sons, 1990.

[30] Rafael S Costa, Andras Hartmann, Paula Gaspar, Ana R Neves, and Susana Vinga. An extended dynamic model of Lactococcus lactis metabolism for mannitol and 2,3-butanediol production. *Molecular bioSystems*, 10(3):628–39, mar 2014.

[31] Peter V Coveney and Philip W Fowler. Modelling biological complexity: a physical scientist's perspective. *Journal of the Royal Society Interface*, 2(4):267–280, 2005.

[32] Markus W. Covert and Bernhard Palsson. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *Journal of Biological Chemistry*, 277(31):28058–28064, 2002.

[33] Markus W Covert, Christophe H Schilling, and Bernhard Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of theoretical biology*, 213(1):73–88, 2001.

[34] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[35] Daniel Horspool. An overview of the (basic) central dogma of molecular biochemistry with all enzymes labeled, 2008. 28 November 2008.

[36] Hidde de Jong, Stefano Casagranda, Nils Giordano, Eugenio Cinquemani, Delphine Ropers, Johannes Geiselmann, and Jean-Luc Gouzé. Mathematical modelling of microbes: metabolism, gene expression and growth. *Journal of The Royal Society Interface*, 14(136):20170502, nov 2017.

[37] Oleg Demin and Igor Goryanin. *Kinetic modelling in systems biology*. Chapman and Hall/CRC, 2008.

[38] Patrick P. Dennis and Hans Bremer. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1), sep 2008.

[39] Peter Deuflhard. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*, volume 35. Springer Science & Business Media, 2011.

[40] Leah Edelstein-Keshet. *Mathematical models in biology*. SIAM, 2005.

[41] Michael Ederer, Sonja Steinsiek, Stefan Stagge, Matthew Rolfe, Alexander Ter Beek, David Knies, Joost Teixeira de Mattos, Thomas Sauter, Jeffrey Green, and Robert Poole. A mathematical model of metabolism and regulation provides a systems-level view of how Escherichia coli responds to oxygen. *Frontiers in microbiology*, 5:124, 2014.

[42] J. S. Edwards and B. O. Palsson. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533, 2000.

[43] Jeremy S Edwards and Bernhard O Palsson. Systems properties of the haemophilus influenzaerd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–17416, 1999.

[44] Scott R Eliason. *Maximum likelihood estimation: Logic and practice*. Number 96. Sage, 1993.

[45] Craig K Enders. Maximum likelihood estimation. *Encyclopedia of statistics in behavioral science*, 2005.

[46] Péter Érdi and János Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.

[47] Marjan Faizi, Tomáš Zavřel, Cristina Loureiro, Jan Červený, and Ralf Steuer. A model of optimal protein allocation during phototrophic growth. *Biosystems*, 166:26–36, apr 2018.

[48] D A Fell and J R Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *The Biochemical journal*, 238(3):781–6, sep 1986.

[49] James E Ferrell. Q&A: Systems biology. *Journal of Biology*, 8(1):2, jan 2009.

[50] Edward H Flach and Santiago Schnell. Use and abuse of the quasi-steady-state approximation. *IEE Proceedings-Systems Biology*, 153(4):187–191, 2006.

[51] David Gavaghan, Sophie Kershaw, James Osborne, and Pras Pathmanathan. A hybrid approach to multi scale modelling of cancer. *Phil. Trans. R. Soc. A*, 368, 2010.

[52] Ziomara P Gerdtzen, Prodromos Daoutidis, and Wei-Shou Hu. Non-linear reduction for kinetic models of metabolic reaction networks. *Metabolic Engineering*, 6(2):140–154, 2004.

[53] Rudolf Gesztelyi, Judit Zsuga, Adam Kemeny-Beke, Balazs Varga, Bela Juhasz, and Arpad Tosaki. The Hill equation and the origin of quantitative pharmacology. *Archive for history of exact sciences*, 66(4):427–438, 2012.

[54] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.

[55] Stanley J Gill, Henry T Gaud, Jeffries Wyman, and B George Barisas. Analysis of ligand binding curves in terms of species fractions. *Biophysical chemistry*, 8(1):53–59, 1978.

[56] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta - General Subjects*, 1810(10):978–988, 2011.

[57] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer, 1971.

[58] Priscilla E Greenwood and Michael S Nikulin. *A guide to chi-squared testing*, volume 280. John Wiley & Sons, 1996.

[59] Robert L Grossman, Anil Nerode, Anders P Ravn, and Hans Rischel. *Hybrid systems*, volume 736. Springer, 1993.

[60] Cato Maxilian Guldberg and Peter Waage. Concerning chemical affinity. *Erdmann's Journal für practische Chemie*, 127(69-114), 1879.

[61] Cato Maximilian Guldberg and Peter Waage. *Etudes sur les affinités chimiques*. Brøgger & Christie, 1867.

[62] Khalil Hassan. Khalil, nonlinear systems. 2002.

[63] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N Mendoza, Anne Richelle, Almut Heinken, Hulda S Haraldsdóttir, Jacek Wachowiak, Sarah M Keating, Vanja Vlasov, Stefania Magnusdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H J Chan, Maike K Aurich, Catherine M Clancy, Jennifer Modamio, John T Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C El Assal, Luis V Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M Le, Ding Ma, Yuekai Sun, Lin Wang, James T Yurkovich, Miguel A P Oliveira, Phan T Vuong, Lemmer P El Assal, Inna Kuperstein, Andrei Zinovyev, H Scott Hinton, William A Bryant, Francisco J Aragón Artacho, Francisco J Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A Saunders, Costas D Maranas, Nathan E Lewis, Thomas Sauter, Bernhard Ø Palsson, Ines Thiele, and Ronan M T Fleming. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3):639–702, 2019.

[64] J.C. Helton and F.J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, jul 2003.

[65] Thomas A. Henzinger. The Theory of Hybrid Automata. In *Verification of Digital and Hybrid Systems*, pages 265–292. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[66] A. V. Hill. The mode of action of nicotine and curari, determined by the form of the contraction curve and the method of temperature coefficients. *The Journal of Physiology*, 39(5):361–373, dec 1909.

[67] Archibald Vivian Hill. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *The Journal of Physiology*, 40:i–vii, January 1910.

[68] P Holmes, JL Lumley, G Berkooz, and CW Rowley. *Turbulence, coherent structures, dynamical systems and symmetry*. 2012.

[69] Werner Horbelt. Maximum likelihood estimation in dynamical systems. pages 1–111, 2001.

[70] Mimmo Iannelli and Andrea Pugliese. *An Introduction to Mathematical Population Dynamics: Along the Trail of Volterra and Lotka*, volume 79. Springer, 2015.

[71] Brian Ingalls. An introduction to mathematical modelling in molecular systems biology, 2012.

[72] John L. Ingraham, Ole. Maaløe, Frederick C. (Frederick Carl) Neidhardt, and 1931. *Growth of the bacterial cell*. Sinauer Associates, 1983.

[73] Kenneth A. Johnson and Roger S. Goody. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry*, 50(39):8264–8269, oct 2011.

[74] Douglas Samuel Jones, Michael Plank, and Brian D Sleeman. *Differential equations and mathematical biology*. Chapman and Hall/CRC, 2009.

[75] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Sana, 1947.

[76] Stuart A Kauffman. The origins of order: Self-organization and selection in evolution. *Understanding Origins*, 65(December):153–181, 1992.

[77] Walter G Kelley and Allan C Peterson. *Difference equations: an introduction with applications*. Academic press, 2001.

[78] Gaetan Kerschen, Jean-claude Golinval, Alexander F Vakakis, and Lawrence A Bergman. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear dynamics*, 41(1-3):147–169, 2005.

[79] Edda Klipp, Ralf Herwig, Axel Kowald, Christoph Wierling, and Hans Lehrach. *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons, 2005.

[80] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald. *Systems biology: a textbook*. John Wiley & Sons, 2016.

[81] Eugene V Koonin. Does the central dogma still stand? *Biology Direct*, 7(1):27, 2012.

[82] Andreas Kremling. *Systems Biology*. Chapman and Hall/CRC, nov 2013.

[83] Andreas Kremling, Sophia Kremling, and Katja Bettenbrock. Catabolite repression in Escherichia coli–a comparison of modelling approaches. *The FEBS journal*, 276(2):594–602, 2009.

[84] Clemens Kreutz. An easy and efficient approach for testing identifiability. *Bioinformatics*, 34(11):1913–1921, jun 2018.

[85] Dirk P. Kroese, Tim Brereton, Thomas Taimre, and Zdravko I. Botev. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):386–392, nov 2014.

[86] Max Kuhn and Kjell Johnson. *Applied predictive modeling*, volume 26. Springer, 2013.

[87] Mustafa R S Kulenovic and Gerasimos Ladas. *Dynamics of second order rational difference equations: with open problems and conjectures*. Chapman and Hall/CRC, 2001.

[88] Peter Kunkel and Volker Mehrmann. *Differential-algebraic equations: analysis and numerical solution*, volume 2. European Mathematical Society, 2006.

[89] Juan Kuntz, Diego Oyarzún, and Guy-Bart Stan. Model reduction of genetic-metabolic networks via time scale separation. In *A systems theoretic approach to systems and synthetic biology I: models and system characterizations*, pages 181–210. Springer, 2014.

[90] Antonio C Lasaga. *Kinetic theory in the earth sciences*. Princeton university press, 2014.

[91] Fabien Lauer and Gérard Bloch. Hybrid system identification. In *Hybrid System Identification*, pages 77–101. Springer, 2019.

[92] AL Lehninger, David L Nelson, and Michael M Cox. Lehninger principles of biochemistry. 2005. *New York*.

[93] Joshua A. Lerman, Daniel R. Hyduke, Haythem Latif, Vasiliy A. Portnoy, Nathan E. Lewis, Jeffrey D. Orth, Alexandra C. Schrimpe-Rutledge, Richard D. 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(1):929, jan 2012.

[94] Ana Patrícia Lima, Vítor Baixinho, Daniel Machado, and Isabel Rocha. A Comparative Analysis of Dynamic Models of the Central Carbon Metabolism of Escherichia coli. *IFAC-PapersOnLine*, 49(26):270–276, 2016.

[95] Lin Liu and Alexander Bockmayr. Formalizing Metabolic-Regulatory Networks by Hybrid Automata. *Acta Biotheoretica*, jul 2019.

[96] Yu Liu. Overview of some theoretical approaches for derivation of the Monod equation. *Applied Microbiology and Biotechnology*, 73(6):1241–1250, jan 2007.

[97] Michel Loeve. Elementary probability theory. In *Probability Theory I*, pages 1–52. Springer, 1977.

[98] Daniel Machado, Rafael S Costa, Miguel Rocha, Eugénio C Ferreira, Bruce Tidor, and Isabel Rocha. Modeling formalisms in systems biology. *AMB express*, 1(1):45, 2011.

[99] R Mahadevan, J S Edwards, and F J Doyle 3rd. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal*, 83(3):1331–1340, 2002.

[100] Ahmad A Mannan, Yoshihiro Toya, Kazuyuki Shimizu, Johnjoe McFadden, Andrzej M Kierzek, and Andrea Rocco. Integrating kinetic model of E. coli with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. *PloS one*, 10(10):e0139507, 2015.

[101] Karla Martínez-Gómez, Noemí Flores, Héctor M Castañeda, Gabriel Martínez-Batallar, Georgina Hernández-Chávez, Octavio T Ramírez, Guillermo Gosset, Sergio Encarnación, and Francisco Bolivar. New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microbial cell factories*, 11(1):46, 2012.

[102] Ali Masoudi-Nejad and Edwin Wang. Cancer modeling and network biology: accelerating toward personalized medicine. In *Seminars in cancer biology*, volume 30, pages 1–3. Elsevier, 2015.

[103] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, may 1979.

[104] Leonor Menten and M I Michaelis. Die kinetik der invertinwirkung. *Biochem Z*, 49(333-369):5, 1913.

[105] Russell B Millar. *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, volume 111. John Wiley & Sons, 2011.

[106] Bud Mishra. Intelligently deciphering unintelligible designs: algorithmic algebraic model checking in systems biology. *Journal of The Royal Society Interface*, 6(36):575–597, 2009.

[107] Carmen G Moles, Pedro Mendes, and Julio R Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research*, 13(11):2467–2474, 2003.

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

[109] J D (James Dickson) Murray. *Mathematical biology*. Berlin ; New York : Springer-Verlag, 2nd, corr. edition, 1993.

[110] Atul Narang. Comparative analysis of some models of gene regulation in mixed-substrate microbial growth. *Journal of theoretical biology*, 242(2):489–501, 2006.

[111] Atul Narang and Sergei S Pilyugin. Bacterial gene regulation in diauxic and non-diauxic growth. *Journal of theoretical biology*, 244(2):326–348, 2007.

[112] K B Newman, S T Buckland, Byron J T Morgan, R King, D L Borchers, Diana J Cole, Panagiotis Besbeas, O Gimenez, and L Thomas. *Modelling population dynamics*. Springer, 2014.

[113] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[114] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010.

[115] Robert L Parker. Understanding inverse theory. *Annual Review of Earth and Planetary Sciences*, 5(1):35–64, 1977.

[116] Jagdish K Patel and Campbell B Read. *Handbook of the normal distribution*, volume 150. CRC Press, 1996.

[117] Luigi Preziosi. Hybrid and multiscale modelling. *Journal of mathematical biology*, 53(6):977–978, 2006.

[118] Jean-François Raskin. An introduction to hybrid automata. In *Handbook of networked and embedded control systems*, pages 491–517. Springer, 2005.

[119] A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser, W. Mader, T. Heinemann, J. Hasenauer, M. Schilling, T. Höfer, E. Klipp, F. Theis, U. Klingmüller, B. Schöberl, and J. Timmer. Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21):3558–3560, 2015.

[120] Andreas Raue, Marcel Schilling, Julie Bachmann, Andrew Matteson, Max Schelke, Daniel Kaschek, Sabine Hug, Clemens Kreutz, Brian D. Harms, Fabian J. Theis, Ursula Klingmüller, and Jens Timmer. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE*, 8(9), 2013.

[121] Federico Reali, Corrado Priami, and Luca Marchetti. Optimization algorithms for computational systems biology. *Frontiers in Applied Mathematics and Statistics*, 3:6, 2017.

[122] Katarzyna A Rejniak and Alexander R A Anderson. Hybrid models of tumor growth. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(1):115–125, 2011.

[123] Osbaldo Resendis-Antonio. Stoichiometric Matrix. In *Encyclopedia of Systems Biology*, pages 2014–2014. Springer New York, New York, NY, 2013.

[124] Maria Rodriguez-Fernandez, Jose A Egea, and Julio R Banga. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7(1):483, nov 2006.

[125] Charles A Rohde. *Introductory statistical inference with the likelihood function*. Springer, 2014.

[126] Shepley L Ross. *Introduction to ordinary differential equations*. John Wiley & Sons, 1980.

[127] Areejit Samal and Sanjay Jain. The regulatory network of E. coli metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC systems biology*, 2(1):21, 2008.

[128] Sandeep Sanga, Hermann B Frieboes, Xiaoming Zheng, Robert Gatenby, Elaine L Bearer, and Vittorio Cristini. Predictive oncology: a review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth. *Neuroimage*, 37:S120–S134, 2007.

[129] Wil Schilders. Introduction to model order reduction. In *Model order reduction: Theory, research aspects and applications*, pages 3–32. Springer, 2008.

[130] R Schuster, HG Holzhütter, G Jacobasch Biosystems, and undefined 1988. Interrelations between glycolysis and the hexose monophosphate shunt in erythrocytes as studied on the basis of a mathematical model. *Elsevier*.

[131] Matthew Scott and Terence Hwa. Bacterial growth laws and their applications. *Current opinion in biotechnology*, 22(4):559–565, 2011.

[132] Charles Gordon Sinclair, Bjørn Kristiansen, and John Desmond Bu'Lock. *Fermentation kinetics and modelling*. Open University Press, 1987.

[133] Lawrence Sirovich. Turbulence and the dynamics of coherent structures. i. coherent structures. *Quarterly of applied mathematics*, 45(3):561–571, 1987.

[134] Kieran Smallbone and Pedro Mendes. Large-scale metabolic models: From reconstruction to differential equations. *Industrial Biotechnology*, 9(4):179–184, 2013.

[135] Mingzhou Song, Zhengyu Ouyang, and Z Lewis Liu. Discrete dynamical system modelling for gene regulatory networks of 5-hydroxymethylfurfural tolerance for ethanologenic yeast. *IET systems biology*, 3(3):203–218, 2009.

[136] Eduardo D Sontag. Lecture Notes on Mathematical Systems Biology. page 283, 2015.

[137] Guy-Bart Stan. Modelling in biology. *Imperial College London*, 2017.

[138] Natalie J. Stanford, Timo Lubitz, Kieran Smallbone, Edda Klipp, Pedro Mendes, and Wolfram Liebermeister. Systematic Construction of Kinetic Models from Genome-Scale Metabolic Networks. *PLoS ONE*, 8(11):e79195, nov 2013.

[139] Gregory Stephanopoulos. Metabolic Fluxes and Metabolic Engineering. *Metabolic Engineering*, 1(1):1–11, 1999.

[140] Angélique Stéphanou and Vitaly Volpert. Hybrid modelling in cell biology, 2015.

[141] Angélique Stéphanou and Vitaly Volpert. Hybrid modelling in biology: a classification review. *Mathematical Modelling of Natural Phenomena*, 11(1):37–48, 2016.

[142] Gilbert Strang, Gilbert Strang, Gilbert Strang, and Gilbert Strang. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.

[143] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press, 2018.

[144] Ne-Zheng Sun and Alexander Sun. *Model calibration and parameter estimation: for environmental and water resource systems*. Springer, 2015.

[145] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. siam, 2005.

[146] René Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3):563–585, 1973.

[147] A. Varma and B. 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(10):3724–3731, 1994.

[148] Nikos Vlassis, Maria Pires Pacheco, and Thomas Sauter. Fast reconstruction of compact context-specific metabolic network models. *PLoS computational biology*, 10(1):e1003424, 2014.

[149] Eberhard Voit. *A first course in systems biology*. Garland Science, 2017.

[150] Stefan Volkwein. Proper orthogonal decomposition: Theory and reduced-order modelling. *Lecture Notes, University of Konstanz*, 4:4, 2013.

[151] Christian Walck. Hand-book on statistical distributions for experimentalists. Technical report, 1996.

[152] Steffen Waldherr, Diego A. Oyarzún, and Alexander Bockmayr. Dynamic optimization of metabolic networks coupled with gene expression. *Journal of Theoretical Biology*, 365:469–485, jan 2015.

[153] Walter D Wallis and John C George. *Introduction to combinatorics*. Chapman and Hall/CRC, 2016.

[154] Zhihui Wang, Joseph D Butner, Romica Kerketta, Vittorio Cristini, and Thomas S Deisboeck. Simulating cancer growth with multiscale agent-based modeling. In *Seminars in cancer biology*, volume 30, pages 70–78. Elsevier, 2015.

[155] Kathleen F Weaver, Vanessa C Morales, Sarah L Dunn, Kanya Godde, and Pablo F Weaver. *An introduction to statistical analysis in research: with applications in the biological and life sciences*. John Wiley & Sons, 2017.

[156] Neil A Weiss and Carol A Weiss. *Introductory statistics*. Pearson, 2017.

[157] IB Wilson, MA Harrison Journal of Biological Chemistry, and undefined 1961. Turnover number of acetylcholinesterase. *ASBMB*.

[158] Laurence Yang, Ali Ebrahim, Colton J. Lloyd, Michael A. Saunders, and Bernhard O. Palsson. DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression. *BMC Systems Biology*, 13(1):2, dec 2019.

[159] Michael S Zhdanov. *Inverse theory and applications in geophysics*, volume 36. Elsevier, 2015.

List of abbreviations

ODE	Ordinary Differential Equations
DAE	Differential Algebraic Equations
FBA	Flux Balance Analysis
dFBA	Dynamic Flux Balance Analysis
deFBA	Dynamic Enzyme-cost Flux Balance Analysis
RBA	Resource Balance Analysis
rFBA	Regulatory Flux Balance Analysis
MOR	Model Order reduction
POD	Proper Orthogonal decomposition
KLE	Karhunen–Loève Expansion
SVD	Singular Value Decomposition
Pr	Probability distribution
LL	Log Likelihood
MRN	Metabolic Regulatory Network
NMR	Nuclear Magnetic Resonance

Notation

$\mathbb{R}$	The real number
S	Stoichiometric matrix
r_i	The ith reactions
M	Vector of species
M_{ext}	Vector of external metabolites
M_{int}	Vector of internal metabolites
v	Vector of reaction rates
k_i	The ith constant reaction rates
n_k	Dimension of reduced model
n_m	Number of species
n_r	Number of reactions
$\mu(t)$	Growth rate at time point t
$e(t)$	Enzyme vector at time point t
$l(t), u(t)$	Lower and upper flux bounds vectors at time point t, respectively
K_M	Michaelis-Menten constant
k_{cat}	Turnover constant rate
X	Biomass
w	Average molar weight of precursor
R	Ribosome
$\mathbb{X}, \mathbb{Y}, \mathbb{Z}$	Snapshot matrices
$\hat{\sigma}$	Singular value
$\hat{\mu}$	The mean of random variable
σ^2	The variance of random variable
χ^2	Chi square distribution
$\tilde{x}(t)$	State's vector of reduced model
$\gamma, \alpha, \zeta, \beta$	Thresholds for the Hill function
h	The Hill coefficient
$\mathfrak{t}$	Discrete time
θ	Vector of parameter
$\hat{\theta}$	Vector of optimized parameters
$\mathscr{U}$	Projection subspace
U, V	Unitary matrices

$\hat{r}$	Rank of matrix
c	Vector of weights
$\overline{RP}, \overline{T_2}$	Boolean variables
n_b	Biomass reactions
v_{max}	Maximum uptake rates
$\mathscr{H}$	Hybrid automaton
T	Transpose
$\hat{\varepsilon}$	Measurement noise

Selbstständigkeitserklärung

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht.

Datum: 17.04.2020
Neveen Ali Salem Eshtewy