Coupling Metabolic and Hydrodynamic Compartmental Models for Bioreactor Simulations

by

Ittisak Promma

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2025

© Ittisak Promma 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Jinfeng Liu, Professor Department of Chemical and Materials Engineering University of Alberta
Supervisors:	Hector Budman, Professor Department of Chemical Engineering University of Waterloo
	Nasser Mohieddin Abukhdeir, Assistant Professor Department of Chemical Engineering University of Waterloo
Internal Member:	Marc Aucoin, Professor Dept. of Chemical Engineering University of Waterloo
	Joshua Pulsipher, Assistant Professor Department of Chemical Engineering, University of Waterloo
Internal-External Member:	Brian Ingalls, Professor Department of Applied Mathematics University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Portions of this thesis are adapted from a research article co-authored with Marc G. Aucoin, Nasser Mohieddin Abukhdeir, Hector Budman, and the author of this thesis. Specifically, parts of Sections 2.2.2, 2.4.2 and 3.1 to 3.3, as well as Chapters 4 and 5, have been modified from Promma et al. [1].

The conceptualization, model development, implementation, and analysis of results were jointly contributed by Marc G. Aucoin, Nasser Mohieddin Abukhdeir, Hector Budman, and the author of this thesis. The final manuscript was reviewed and approved by all co-authors. These contributions were refined, expanded, and integrated into this thesis to align with its broader research goals. Proper citations to the original publication are included.

Abstract

Large-scale bioreactors are widely employed across bioprocessing industries for the production of chemicals, pharmaceuticals, and biofuels. The increasing demand for specialized pharmaceuticals has motivated industries to optimize bioreactor operations. However, the complexity of multiphase interactions and the emergence of concentration gradients and intracellular heterogeneity in bioreactors pose significant challenges in accurately predicting and optimizing the performance of bioreactors. Simulations of numerical models have become invaluable for understanding these systems; however, the high computational cost of detailed models—particularly those involving multiphysics—limits their practicality. The computational cost of these models precludes them from being used for real-time applications or for extensive design optimizations.

To address this challenge, this thesis proposes computationally efficient methods to solve coupled metabolic and hydrodynamic compartmental models that describe key process dynamics.

The compartmental model (CM) approach is based on steady-state multiphase computational fluid dynamic simulations and is designed to accurately represent hydrodynamic properties such as turbulent dissipation energy, oxygen mass transfer, and flow topology. Conventional compartmentalization methods were found to introduce nonphysical "short-cutting" effects, leading to inaccuracies in mixing time predictions. To mitigate this, a refined compartmentalization approach was developed to better capture hydrodynamic mixing patterns. Then, a metabolic model was integrated with the compartmental model to explore the metabolic and hydrodynamic interactions.

In terms of the metabolic behavior, two key scenarios were considered: i) the intracellular concentrations were assumed to reach instantaneous equilibrium with their extracellular environments at all times and ii) the intracellular concentrations were not in equilibrium with their extracellular environments.

For the first case, where intracellular and extracellular equilibration was assumed, a binary search tree metabolic model (BSTMM) was developed from a dynamic flux balance analysis model and coupled with a CM describing the extracellular environment. This method significantly reduces computational complexity by substituting traditional linear optimization solvers with an online point-location approach. The coupled BSTMM-CM successfully captured diauxic growth dynamics and demonstrated substantial computational efficiency, enabling long-term bioprocess simulations on standard desktop hardware. However, kinetic parameters for metabolic models calibrated in small-scale bioreactors could not be directly applied to large-scale systems without recalibration. This finding

suggested that intracellular heterogeneity may play a crucial role in metabolic regulation and must therefore be explicitly accounted for.

To address this, the second case considered a finite-rate adaptation mechanism governing equilibration between the extracellular and intracellular environments. A method of moments approach, assuming a truncated normal distribution, was implemented to reconstruct the number density function efficiently. This approach demonstrated that intracellular heterogeneity is most pronounced when the characteristic timescales of microbial adaptation and extracellular advection are comparable. The application of this approach to *E. coli* fermentation data reported in the literature resulted in improved fitting as compared to a model that ignores intracellular heterogeneity. Furthermore, the impact of population heterogeneity on metabolic regulation was evaluated, revealing distinct variations in growth rate and substrate uptake across different regions of the bioreactor.

To further improve computational efficiency of the coupled PBM-CM, an adaptive population compartmental scheme is proposed, which dynamically adjusts the compartmentalization over the course of a simulation to balance accuracy and computational cost. This approach was found to be particularly effective for large-scale bioreactor simulations, especially when advection rates exceed microbial adaptation rates, leading to substantial reductions in simulation time with minimal loss of predictive accuracy.

This research significantly advances the modeling of large-scale bioreactors by integrating hydrodynamic and metabolic models into a computationally efficient framework. The developed methods provide more in-depth insights into the influence of concentration gradients and intracellular heterogeneity on microbial behavior, ultimately improving the predictive accuracy and scalability of bioprocess simulations.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Professor Hector Budman and Professor Nasser Mohieddin Abukhdeir, whose guidance and support have been invaluable throughout my graduate studies. Their dedication to research and commitment to excellence have shaped my academic journey and inspired me to strive for the highest standards in my work. From the very beginning, their insightful advice and constructive feedback have played a pivotal role in refining my ideas and advancing my research. Their generosity with their time, patience, and encouragement has helped me navigate challenges and stay motivated. I deeply appreciate their mentorship, which extends beyond academia, fostering both my professional growth and personal resilience. It has been a privilege to learn from them, and I am sincerely thankful for their unwavering support. *If there were words that conveyed more than "thank you very much", I would use them to express my deepest gratitude to both of you, Sirs.*

I am deeply grateful to Dr. Alex Vasile, my friend and colleague, for his unwavering support throughout my PhD journey. I consider myself incredibly fortunate to have known him and worked alongside him. His guidance, encouragement, and generosity have left a lasting impact on both my research and personal growth.

To Daniel Rangel Martinez, thank you for your unwavering care and support. Your presence has been my greatest source of strength and comfort, making every challenge more manageable and every success more meaningful. I am endlessly grateful to have you by my side.

Thank you to my best friend, Andrew Stella, for sharing wonderful moments with me, for your unwavering support, and for always being there through the highs and lows. Your kindness, encouragement, and friendship have meant so much to me throughout this journey.

Additionally, I extend my gratitude to Arshia Fazeli for your constant support, thoughtful advice, and for always bringing positivity into my life. Your friendship has been a source of strength and laughter, making even the toughest days more bearable. I am grateful for the countless discussions, encouragement, and shared experiences that have enriched my journey.

I would also like to acknowledge my Thai friends at Waterloo—Pin, Peach, Ching, P, Aom, Pi, Fong, Gigi, Kitty, Nathy, and Linhnhi—for their friendship and support. Having them by my side has been a source of comfort, making Waterloo feel like home and shielding me from homesickness. Their presence has truly made this journey more enjoyable and meaningful.

I extend my sincere gratitude to the Department of Chemical Engineering at the University of Waterloo for providing me with the opportunity to pursue my PhD, along with exceptional resources and a supportive environment. Special thanks to the administrative staff, particularly Judy Caron, for their patience, kindness, and invaluable assistance throughout my journey. Their dedication and support have made a meaningful difference in my experience.

I am also deeply grateful to my doctoral committee members—Prof. Jinfeng Liu, Prof. Brian Ingalls, Prof. Marc Aucoin, Prof. Joshua Pulsipher, and Prof. David Simakov—for their invaluable feedback, which provided direction to my doctoral research. Their insights and perspectives introduced new dimensions of thought that enriched my work and contributed significantly to its depth and quality.

Without the unwavering support of my Promma family, this work would not have been possible. My deepest gratitude goes to my parents, Por Kong and Ma Jin, for their unconditional love, care, and endless encouragement. I am also profoundly grateful to my dearest sisters, Pang, Am, Bam, and Ben, for their unwavering love and constant support. Their presence in my life has been a source of strength and motivation throughout this journey.

Lastly, I would like to acknowledge our research sponsors, including Sanofi Pasteur Limited, the Natural Sciences and Engineering Research Council of Canada, and the Digital Research Alliance of Canada. Their invaluable support and contributions have been instrumental in the successful completion of this research.

Dedication

I dedicate this thesis to my family, whose unwavering support, encouragement, and sacrifices have made this journey possible. To my parents, Por Kong and Ma Jin, for instilling in me the values of perseverance, curiosity, and kindness, and for always believing in my dreams. To my dearest sisters, Pang, Am, Bam, and Ben, for always being in my heart and providing unwavering support from afar. To Daniel, for his care and invaluable support throughout this journey.

Lastly, I dedicate this work to all those who pursue knowledge with passion and resilience, striving to make meaningful contributions to a better world.

Table of Contents

Examining Committee ii

Author's Declaration iii

Statement of Contributions iv

Abstract v

Acknowledgements vii

Dedication ix

List of Figures xiv

List of Tables xx

1 Introduction 1

1.1 Research Motivation 1

1.2 Objectives 4

1.3 Structure of Thesis 5

2 Background 7

2.1 Understanding concentration gradients and intracellular heterogeneity in bioreactors 7

2.1.1 Formation of concentration gradients and intracellular heterogeneity 8
2.1.2 Responses of microorganisms exhibiting concentration gradients and intracellular heterogeneity . . . 10
2.1.3 Approaches for investigating concentration gradients and intracellular heterogeneity . . . 11
2.2 Hydrodynamic models . . . 13
2.2.1 Two-fluid models . . . 13
2.2.2 Compartmental models . . . 16
2.3 Metabolic models . . . 19
2.4 Constraint-based metabolic models . . . 20
2.4.1 Flux balance analysis . . . 22
2.4.2 Dynamic flux balance analysis . . . 23
2.5 Population balance models . . . 25
2.5.1 Number density functions . . . 26
2.5.2 Population balance equations . . . 28
2.5.3 Population balance equations for compartmental models . . . 32
2.5.4 Moments of the number density function . . . 33
2.5.5 Moment transport equations . . . 34
2.5.6 Moment transport equations for compartmental models . . . 35

3 Literature review **37**
3.1 Simulation of large-scale bioreactors . . . 37
3.2 Compartmentalization . . . 40
3.2.1 Determination of compartment network . . . 40
3.2.2 Determination of flow rates in multiphase compartmental models . . . 41
3.3 Numerical method for dynamic flux balance analysis . . . 43
3.3.1 Multiparametric linear programming representation of dynamic flux balance analysis . . . 43
3.3.2 Mathematical description of a critical region map . . . 44

3.3.3 Parameter point location . . . 45
3.3.4 Construction of binary search tree . . . 46
3.4 Numerical methods for solving the population balance equation . . . 47
3.4.1 Bin method . . . 48
3.4.2 Method of moments . . . 51
3.4.3 Truncated multivariate normal distribution and its moments . . . 54

4 Development of a CM for Large-Scale Bioreactor Simulations 56
4.1 Introduction . . . 56
4.2 Methodology . . . 57
4.2.1 Computational fluid dynamics simulation setup . . . 57
4.2.2 Compartmentalization . . . 60
4.3 Results and discussion . . . 61
4.3.1 Multiphase computational fluid dynamics simulation . . . 61
4.3.2 Flow-informed compartmentalization . . . 62
4.4 Conclusion . . . 64

5 Development of a BSTMM to Simulate a Coupled FBA-CM 65
5.1 Introduction . . . 65
5.2 Methodology . . . 66
5.3 Results and discussion . . . 68
5.3.1 Integration of the binary search tree metabolic model into the compartmental model . . . 68
5.3.2 Validation and analysis of simulations of the large-scale bioreactor . . . 71
5.3.3 Effects of flow oscillations . . . 83
5.4 Conclusion . . . 86

6 Development of an MM-TND to Simulate a Coupled PBM-CM **87**
6.1 Introduction . . . 87
6.2 Methodology . . . 89
6.3 Results and discussion . . . 92
6.3.1 Numerical challenges of solving PBEs using the QMOM . . . 92
6.3.2 Validation of the proposed MM-TND . . . 94
6.3.3 Effect of flow rate on gradients and intracellular heterogeneity . . . 100
6.3.4 Validation and analysis of simulations of the large-scale bioreactor . 107
6.4 Conclusion . . . 117

7 Development of an APCS to Simulate a Coupled PBM-CM **119**
7.1 Introduction . . . 119
7.2 Methodology . . . 120
7.2.1 Development of a computationally efficient compartmental map for the cell population . . . 120
7.2.2 Development of the APCS . . . 123
7.3 Results and discussion . . . 128
7.4 Conclusion . . . 133

8 Conclusions and Future Directions **135**
8.1 Conclusions . . . 135
8.2 Recommendations for Future Work . . . 138

References **140**

List of Figures

2.1 Microorganisms traverse different zones in the bioreactor: (a) a scenario where the cell adaptation rate is of the same order of magnitude as the advection rate, causing states that differ from the zone environment, leading to heterogeneity in the mixing zone; (b) a scenario where cells immediately adjust their states to match the zone environment, resulting in a homogeneous population in the mixing zone. 9

2.2 Classification of metabolic model 19

2.3 A diagram of the metabolic network of *E. coli*, taken from Mahadevan et al. [20], comprises 54 metabolites and 85 reactions. It includes pathways of glycolysis, the pentose phosphate pathway, the TCA cycle with the glyoxylate bypass, anapleurotic reactions, and redox metabolism. 21

2.4 Constraint-based metabolic models 22

2.5 Number density function of cells characterized by size as an internal state . 26

3.1 Illustration depicting critical region maps in $\mathbb{R}^2$ featuring labelled geometric elements of significance, including critical regions as Θ_i, hyperplanes as $\mathbf{D}_j^0$, and vertices as $\mathbf{y}_{i,k}$ 44

4.1 Illustration depicting the mesh for the reactor geometry and the computational mesh employed; presenting (a) the external surface mesh, (b) the components inside the tank, (c) mesh within the rotational domain, and (d) mesh at the bottom of the tank. 58

4.2 30m^3 bioreactor (with 22m^3 of working volume): (a) Line integral convolution representation of velocity fields colored by the liquid volume fraction; Averaged velocity components (b) u_x, (c) u_y, (d) u_z; (e) Compartments from the flow-informed compartmental model. 62

4.3 The presence of "short-cutting" paths (indicated by the arrows) introduced by compartments (highlighted in orange and yellow) between local recirculatory zones. The dashed line indicates a plane where the flow flux perpendicular to the plane is nearly zero. 63

5.1 A flowchart outlining the process of integrating the compartmental model with the binary search tree metabolic model, with the species balance serving as a central junction connecting both models 67

5.2 Potential positions of critical regions in relation to a hyperplane ($\mathbf{D}_j^0$); critical regions located (a) in $\mathbf{D}_j^+$, (b) in both $\mathbf{D}_j^+$ and $\mathbf{D}_j^-$, (c) in $\mathbf{D}_j^-$. 68

5.3 Comparison of simulation concentration results with experimental data in a lab-scale bioreactor: (a) glucose, (b) acetate, (c) biomass. The vertical line $L1$ denotes the starting point of the fed-batch with an exponential feed rate. $L2$ and $L3$ denote the starting point of fed-batch with different constant feed rates (Table 5.2). 73

5.4 Comparison of simulation concentration results with experimental data in a large-scale bioreactor: (a, c, e) refer to glucose, while (b, d, f) refer to acetate, studied at three distinct locations: (a, b) the top probe (6.48 m from the bottom of the tank), (c, d) the middle probe (4.04 m from the bottom of the tank), and (e, f) the bottom probe (1.12 m from the bottom of the tank). The vertical line $L1$ denotes the starting point of the fed-batch with an exponential feed rate. $L2$ and $L3$ denote the starting point of fed-batch with different constant feed rates (Table 5.2). 77

5.5 Oxygen concentration simulation results in a large-scale bioreactor 78

5.6 Comparison of simulation concentration results of biomass with experimental data at the top section in a large-scale bioreactor 78

5.7 Comparison of the specific glucose uptake rate in laboratory-scale and large-scale bioreactors. The single rate, denoted by the solid line, is reported for a well-mixed lab-scale bioreactor, while rates at different locations in the large-scale bioreactor are reported as a range bounded by the dashed lines. The dynamics in the upper bounds are mostly due to dynamics in glucose concentration at the top part of the tank, where most of the glucose was fed at different rates (Table 5.2) and consumed throughout the course of operation. This result indicates that microorganisms in the lab-scale bioreactor exhibit an approximately average glucose consumption rate of those rates occurring in the large-scale bioreactor. 79

5.8 Reaction rates observed at various heights within the system. **Red**: bottom at 1.12 m, **blue:** middle at 4.04 m, and **green:** top at 6.48 m. 81

5.9 Metabolic fluxes measured at the following locations: (a) the lower probe at 1.12 m, (b) the middle probe at 4.04 m, and (c) the top probe at 6.48 m. Each flux is associated with a specific metabolic reaction as outlined: $v_{1,j}$: $39.43A + 35O_2 \rightarrow X$, $v_{2,j}$: $9.46G + 12.92O_2 \rightarrow X$, $v_{3,j}$: $9.84G + 12.73O_2 \rightarrow 1.24A + X$, and $v_{4,j}$: $19.23G \rightarrow 12.12A + X$ 82

5.10 Glucose concentration profile at different heights. The range of concentrations results from measurements taken across all compartments at each fixed height, reflecting the uncertainty of the angular positions in the simulation. 83

5.11 Comparing simulation results with turbulent-induced oscillation flow rates (where $M = 0.7$ and $\omega = 3\ \mathrm{h}^{-1}$) to experimental data in a large-scale bioreactor: (a,c,e) refer to glucose, while (b, d, f) refer to acetate, studied at three distinct locations: (a, b) the top probe (6.48 m from the bottom of the tank), (c, d) the middle probe (4.04 m from the bottom of the tank), and (e, f) the bottom probe (1.12 m from the bottom of the tank). 85

6.1 Diagram of the MM-TND solver. 89

6.2 Numerical analysis results highlight the relative error between the numerical integration value, computed using Simpson's 1/3 rule, and the expected value as the standard deviation approaches zero for $\int_{-\infty}^{\infty} g\phi dz$. This relative error is illustrated for various functions: (a) the zeroth-order moment where $g = 1$, the first-order moment where $g = z$, and the second-order moment where $g = z^2$, and (b) the Monod equation where $g(z) = z/(k + z)$. The shaded area in the figure represents the range of error introduced by variations in the position of the population mean relative to the integration grid. 91

6.3 Demonstration of numerical challenges due to the use of a structured metabolic model, leading to discontinuities in the values of mass transfer or reaction rates, w. 93

6.4 Demonstration of numerical challenges due to the use of a normal distribution, ϕ, instead of a truncated normal distribution, τ, forcing cells to reside in an unphysical domain. 94

6.5 Concentration profiles of (a) glucose (c_G), (b) acetate (c_A), and (c) biomass (c_X), obtained using the bin method (BM) with varying numbers of bins, and the method of moments with an assumed truncated normal distribution (MM-TND). The model predictions agree across methods, resulting in a single line in each plot. 98

6.6 (a) Mean (μ_z) and (b) standard deviation (σ_z) of the growth rate, r_X, as determined using the bin method (BM) with different number of bins and the method of moments with an assumed truncated normal distribution (MM-TND). 99

6.7 CM with 9 compartments of a large 30 m^3 bioreactor. 100

6.8 Characteristic time profiles for various phenomena in the bioreactor simulation, computed using equations from Table 6.3, under the following flow rates: (a-c) $q = 0.01152$L/h, (d-f) $q = 1,152$L/h, and (g-i) $q = 1,152,000$L/h. The results are shown for (a, d, g) glucose, (b, e, h) acetate, and (c, f, i) cell population. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments. 104

6.9 Comparison of simulated concentration profiles across all compartments for (a) glucose, (b) acetate, and (c) biomass (average concentration) at different flow rates. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments. 105

6.10 Comparison of simulation results of growth rate heterogeneity across various compartments when (a) $q = 0.01152$ L/h (b) $q = 1,152$ L/h (c) $q = 1,152,000$ L/h. The numbered tags with different colors indicate the compartments from which the solutions are taken, while the shaded bands represent the range of growth rate values—from the mean, μ_z, to three standard deviations, $3\sigma_z$. 106

6.11 Comparison of the simulation results with the experimental data [22] for the concentration profiles of (a) glucose, (b) acetate, (c) biomass, and (d) inhibitor (only the simulation result presented) in a lab-scale reactor. . . . 110

6.12 Growth rate profile in a lab-scale reactor. At each time point, the mean growth rate, μ_z, approaches the optimal growth rate, z^*, set by the external substance concentration. 111

6.13 Comparison of simulation results with experimental data [22] for the average biomass concentration in a bioreactor with a working volume of 22 m^3. . . 112

6.14 Comparison of simulation results with experimental data [22] for concentration profiles in a bioreactor with a working volume of 22 m^3. Glucose profiles are shown in (a, c, e), while acetate profiles are shown in (b, d, f) at three probe locations: (a, b) top probe at 6.48 m, (c, d) middle probe at 4.04 m, and (e, f) bottom probe at 1.12 m from the bottom of the reactor. 113

6.15 Characteristic time profiles for various phenomena in the fed-batch bioreactor simulation with active volume of 22 m^3. The results are shown for (a) glucose, (b) acetate, and (c) cell population. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments. 115

6.16 Growth rate profile in a bioreactor with a working volume of 22 m^3. At each time point, the mean growth rate, μ_z, approaches (a) the volume-averaged optimal growth rate, μ_{z^*}, of (b) the optimal rates, z_j^*, set by the external substance concentration in different compartments. 116

6.17 Comparison of simulation results for concentration profiles of (a) glucose, (b) acetate, and (c) biomass (average concentration) at different heights in a bioreactor with a working volume of 22 m^3: without the population balance equation (solid line), with the population balance equation (dashed line), and experimental data [22]. 117

6.18 Comparison between the actual consumption/production rates (Actual) and the optimal rates determined by the external substance concentrations (Optimal) for (a) glucose and (b) acetate in a bioreactor with a working volume of 22 m^3. 117

7.1 Microorganisms from Compartment 1 rapidly transit through Compartment 2 to Compartment 3 without sufficient time to adapt their internal state. . 121

7.2 Diagram of the adaptive population compartmental scheme (APCS). . . . 124

7.3 Growth rate over time. At each time point, the mean growth rate, μ_z, approaches the volume-averaged optimal growth rate, μ_{z^*} of the optimal rates, z_j^*, set by the external substance concentration in different compartments. . 129

7.4 Standard deviation (σ_z) and relative error comparing the implementation of APCS with the manual PopComp map fixed at $n_{PC} = n_{EC}$ for (a) $n_{EC} = 9$ and (b) $n_{EC} = 36$. 131

7.5 (a) Evolution of the PopComp map and (b) number of PopComp compartments in the map when the APCS was implemented for $n_{EC} = 9$ and $n_{EC} = 36$. 132

7.6 Comparison of CPU time required to simulate the large-scale bioreactor with different methods: the APCS and the manual PopComp map fixed at $n_{PC} = n_{EC}$ for (a) $n_{EC} = 9$ and (b) $n_{EC} = 36$. The scope of each process is shown in Figure 7.2. 133

List of Tables

2.1 Variables and parameters in Equation 2.4. n_s denotes the number of substrates. n_{EC} denotes the number of compartments. 18

3.1 Overview of the current state-of-the-art in the use of metabolic models integrated with CFD or compartmental models. EE: Euler-Euler method, CM: Compartmental Model, PBE: Population Balance Equation 38

3.2 Summary of studies on population balance models used to investigate microbial population heterogeneity. 39

4.1 Physical properties of water and air used for the simulation. 59

5.1 Model parameters 72

5.2 Feed profiles of glucose in L/h and the associated parameters (Xu et al.[22]). 74

5.3 Initial concentrations and control parameters 74

5.4 Comparison of the corrected Akaike information criteria 75

5.5 Coordinates (X, Y, Z) of concentration probes and the feed location used in the simulation. The origin coordinate is at the center of the bottom hemisphere. 75

6.1 Initial concentrations 96

6.2 Model constants 97

6.3 Various characteristic times at Compartment j for further analysis 103

6.4 Programs for the glucose feed flow rate, $q_{ex\rightarrow j}$ (in L/h), and the glucose feed concentration, $c_{G,ex\rightarrow j}$, for lab-scale (15 L) and large-scale (30 m^3) bioreactors. 108

6.5 Comparison of AICc values for different models 112
6.6 Root mean squared error (RMSE) in g/L for Glucose, Acetate, and Biomass 114

Chapter 1

Introduction

1.1 Research Motivation

Bioprocesses are widely employed in the production of various substances, including food ingredients, chemicals, and pharmaceuticals. Aerated stirred bioreactors are key pieces of equipment in pharmaceutical production and represent a highly intricate multiphase system. Dynamic interactions among gas-phase bubbles, solid-phase microorganisms, and a liquid-phase nutrient solution lead to complex behavior in these bioreactors, necessitating a thorough understanding of the underlying processes for efficient operation. Since the products must be manufactured efficiently and cost-effectively at a large scale, increasingly large bioreactors—several orders of magnitude larger than those feasible for laboratory use—are required. For example, bioreactors used in industrial applications can range from 1,000 liters to 500,000 liters, with some systems even exceeding 1,000,000 liters depending on the process requirements [2]. These large-scale bioreactors often face mass transfer limitations, leading to incomplete mixing and the formation of gradients in key operational variables.

Computational simulations have emerged as an invaluable tool for understanding bioreactor operation, as well as providing critical insights into both extracellular properties (e.g., velocity, concentrations, and pH) and intracellular properties (e.g., cell size and internal concentrations). These insights are often difficult to obtain through physical experimentation due to cost constraints and the lack of direct sensors or analytical techniques for measuring variables of interest. Moreover, simulations are crucial for predicting key bioprocess performance metrics, such as yield and productivity, because they enable the analysis of complex interactions between biological, chemical, and physical phenomena that are

often difficult or costly to observe experimentally. The ability to simulate bioreactor operations under various conditions provides several benefits, including aiding in developing more efficient bioprocesses and determining optimal operating conditions.

Conducting these simulations necessitates the development of mathematical models capable of accurately representing the intricate physics of these phases and their interactions, including the cellular biology of microorganisms. Bioreactors are inherently multiphase systems, where different phases, such as liquid, gas, and microorganisms, are transported over large scales. Therefore, bioreactor models must account for the conservation of mass, momentum, and energy in each phase, as well as include constitutive models that describe interphase transport and biological metabolism.

To effectively model the inherent multiphase flows within bioreactors, several feasible multiphase models exist, including Euler-Euler [3, 4, 5, 6], Lagrangian-Euler [7, 8, 9, 10], and volume-of-fluid approaches [11, 12, 13, 14]. These models can be solved using computational fluid dynamics (CFD), which enables a detailed analysis of fluid flow and interactions at a level where the continuity of phase properties is valid. Although multiphase models enable accurate simulations, their computational cost limits their application in scenarios requiring fast computation [15]. For example, the application of model-based real-time control and optimization, or the use of a model for optimizing bioreactor design, requires fast model predictions that generally cannot be achieved with CFD approaches. To address this, it is necessary to trade accuracy for computational efficiency by using reduced-order models that are less accurate than CFD but still provide valuable insights into system behavior.

Given that stirring contributes significantly to the hydrodynamics of all phases, it is reasonable to assume that the majority of the bioreactor operates under a convection-dominated regime. This allows for a one-way coupling approximation, in which the hydrodynamics of each phase are described such that their flow topology is negligibly affected by temporal changes in intraphase properties, such as concentration, temperature, and other factors. Consequently, the fluid acts as a carrier for these properties, transporting them within the system [16]. Within this framework, compartmental models (CMs) have emerged as effective tools for simulating the system hydrodynamics.

CMs represent a system as a set of interconnected compartments, dividing the spatial domain into an ideal reactor network. This approach reduces computational complexity while capturing the essential features of flow dynamics [17]. CMs can be developed by compartmentalizing a steady-state hydrodynamic simulation obtained from a continuum model, primarily through the use of CFD; this technique is known as a CFD-derived CM. This approach has significantly improved the assessment of flow rates between compart-

ments and provides precise evaluations of various local physical properties at the locations represented by the compartments, including the volume fraction of each phase in multiphase systems, rate of turbulent energy dissipation, and oxygen mass transfer.

Expanding upon CMs, phenomena such as mass transfer, metabolic reactions, and cell division are modeled based on their constitutive principles. This research focuses on two key challenges in bioreactor modeling: heterogeneity (gradients) in (i) extracellular and (ii) intracellular concentrations. The extracellular gradients arise from spatial variations in nutrient availability, dissolved gases, and other environmental factors, while the intracellular heterogeneity pertains to differences in cellular states—such as enzyme levels or metabolic activity—within a population of microorganisms. These two forms of heterogeneity frequently occur and present significant obstacles in scaling bioprocesses from controlled laboratory environments to large-scale industrial production systems [5].

The presence of extracellular heterogeneities primarily contributes to the emergence of intracellular heterogeneities. However, many simulation-based studies have focused exclusively on extracellular concentration gradients. Intracellular heterogeneity has been largely overlooked due to model complexity and high computational demands. Hence, past simulations work have generally assumed that microorganisms immediately adjust their intracellular states in response to environmental conditions, leading to heterogeneity among these states. Consequently, modeling microorganisms based solely on external conditions or a uniform intracellular state is less accurate, given the inherently heterogeneous nature of microbial populations. In this study, an approach for modeling intracellular heterogeneity is developed, involving the tracking of cells distributed across various intracellular states using the population balance equation (PBE) coupled with moment transport equations (MTEs).

This study presents a method for solving MTEs by incorporating a truncated normal distribution (MM-TND). This approach improves the accuracy of microbial population modeling by maintaining realistic physical constraints on the internal state distribution, while capturing finer-scale variations compared to the bin method in the literature [18].

To effectively model metabolic reactions in the presence of these heterogeneities, structured (or conditional) metabolic models are recommended over basic Monod-type models that incorporate yield coefficients. These models account for variations in cellular metabolic structures by incorporating conditional switches within the kinetic framework or adjusting parameters across different extracellular and intracellular states [19]. However, this added level of detail requires more model parameters and conditionals to adequately capture these dynamics.

Dynamic flux balance analysis (DFBA) [20], a well-established conditional metabolic

model, has the potential to represent microbial adaptation and regulation under different conditions. DFBA enables dynamic switching of metabolic pathways to achieve optimal biological goals, such as growth rate and ATP production, while requiring fewer kinetic parameters. This makes DFBA particularly appealing among conditional metabolic models. However, implementing DFBA poses an additional numerical challenge for its integration with conservation models, as it requires solving a linear programming (LP) problem at each time step. As a result, computational complexity is significantly increased, requiring the implementation of "time-saving" strategies.

In this research, a binary search tree metabolic model (BSTMM) is developed from the FBA framework, utilizing multiparametric optimization [21] to address the computational complexity. This approach enables rapid calculation of metabolic rates at each stage of microbial growth within the bioreactor by replacing the linear optimization problem with a much simpler point-location problem.

By incorporating a more detailed representation of these phenomena through advanced modeling approaches, large-scale bioreactor simulation accuracy and reliability can be significantly enhanced. This, in turn, enables a more in-depth understanding of the complex interactions governing bioprocesses, which is ultimately expected to improve process optimization, efficiency, and scalability in industrial biotechnology.

1.2 Objectives

This thesis consists of four main research studies, each contributing to the advancement of bioreactor modeling through computational approaches, as follows:

(i) Development of a flow-informed CM derived from CFD simulations to analyze a 30 m^3 aerated stirred bioreactor. This study covers CFD simulation setup, compartmentalization, and the resulting CM

(ii) Development of a BSTMM to reduce computational complexity in dynamic metabolic models coupled with CM. The BSTMM is integrated with the CM to simulate diauxic growth of *E. coli*. In this part, the intracellular population is assumed to be homogeneous.

(iii) Extension of the CM by incorporating MTEs derived from a PBE to account for intracellular heterogeneity. The MTEs are solved using MM-TND. The primary objective is to assess whether intracellular heterogeneity enhances simulation accuracy.

A scaling analysis is performed by adjusting advection time scales and operating modes to investigate concentration gradients and intracellular heterogeneity.

(iv) Development of an adaptive population compartmental scheme (APCS) that dynamically adjusts compartmentalization during simulations to optimize computational efficiency while maintaining predictive accuracy.

In summary, the specific objectives of this research are:

1. Develop a flow-informed CFD-derived CM to analyze a large-scale, multiple-impeller aerated bioreactor during *E. coli* fed-batch fermentation [22].
2. Develop a BSTMM from a DFBA to reduce computational complexity.
3. Integrate the BSTMM with the flow-informed CM and validate the coupled model by comparing its simulation results with experimental data from the large-scale bioreactor.
4. Develop a computational framework to solve MTEs using the MM-TND.
5. Validate the coupled MTEs/CM model by comparing simulation results with experimental data from the large-scale bioreactor.
6. Develop an APCS to dynamically adjust compartmentalization during simulations.

Overall, this research aims to develop computationally efficient models for simulating concentration gradients and intracellular heterogeneity in aerated stirred bioreactors. The availability of such models is expected to facilitate their application in real-time control, optimization, or design processes requiring repetitive simulations.

1.3 Structure of Thesis

This thesis consists of seven additional chapters.

Chapter 2 provides the background necessary for understanding large-scale bioreactor modeling, covering the formation of concentration gradients and intracellular heterogeneity, along with key modeling approaches, including hydrodynamic, metabolic, and population balance models.

Chapter 3 presents a literature review, evaluating different existing computational modeling approaches for bioreactor simulations and identifying gaps in current methods that motivate this research.

Chapter 4 introduces a flow-informed CM derived from CFD to improve hydrodynamic accuracy while maintaining computational efficiency.

Chapter 5 integrates DFBA with the CM, introducing a BSTMM to reduce computational complexity and validating the coupled model against experimental data.

Chapter 6 extends the modeling framework by integrating a PBM with the CM and applying MM-TND to solve MTEs, enabling efficient tracking of intracellular heterogeneity.

Chapter 7 enhances computational efficiency with an APCS, which dynamically adjusts compartmentalization during simulations to balance accuracy and speed.

Finally, Chapter 8 summarizes key findings and discusses future research directions.

Chapter 2

Background

Scaling up bioprocesses from laboratory settings to industrial production poses considerable challenges due to the development of spatial and temporal variations within large-scale bioreactors. These variations, such as concentration gradients and intracellular heterogeneity, can significantly influence microbial metabolism, growth behavior, and overall process efficiency. Gaining insights into the formation of such heterogeneity, microbial responses to fluctuating environmental conditions, and the methodologies for investigating these variations, is essential for optimizing bioreactor performance.

This chapter presents a comprehensive review of these concepts, along with key aspects of large-scale bioprocess modeling. It covers fundamental modeling approaches, including hydrodynamic, metabolic, and population balance models, which serve as the basis for simulating bioprocess dynamics.

2.1 Understanding concentration gradients and intracellular heterogeneity in bioreactors

Gradients (extracellular heterogeneity) and intracellular heterogeneity commonly occur in large-scale bioreactors operating in fed-batch or continuous mode. These variations have been shown to impact process performance significantly, including yield, product quantity, and productivity [23].

This section provides a concise but thorough overview of the formation of concentration gradients and intracellular heterogeneity in large-scale bioreactors, microbial responses, and the approaches used to simulate these phenomena.

2.1.1 Formation of concentration gradients and intracellular heterogeneity

While undergoing advection within a reactor, microorganisms consume substrates for the synthesis of metabolic products, cell maintenance, and reproduction. A concentration gradient of a metabolite develops, particularly when an imbalance exists between the rate of consumption/production by microorganisms and the rate of transport throughout the system [23]. For example, in regions where microorganisms consume nutrients faster than they are replenished by mixing or diffusion, local depletion occurs. Meanwhile, in some areas, the concentration may remain consistently high, particularly near a feeding port. This uneven consumption and production of metabolites, coupled with imperfect mixing, results in the formation of concentration gradients across the bioreactor.

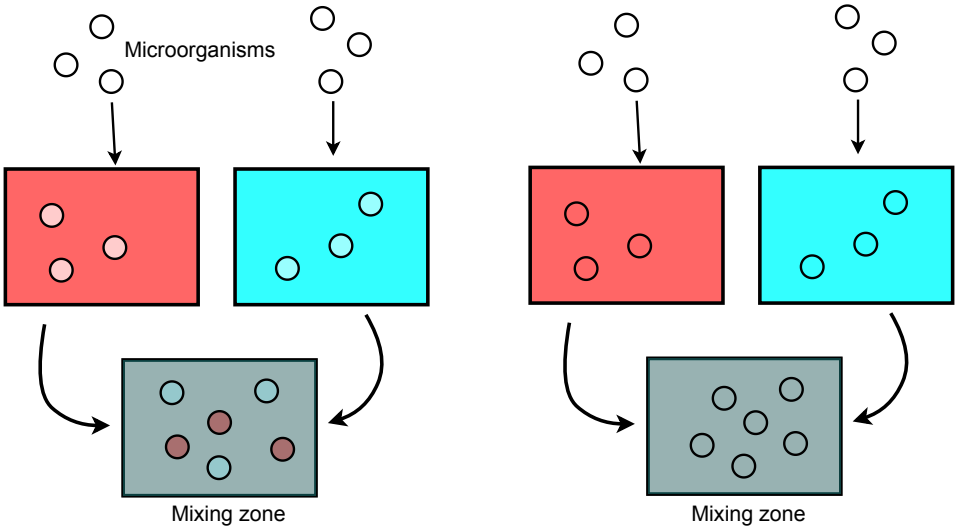
As microorganisms encounter concentration gradients—and potentially other physical property gradients—they undergo dynamic adjustments in their intracellular states in response to the surrounding extracellular environment. As cells migrate through different regions of the bioreactor, these adaptations accumulate, ultimately contributing to the emergence of intracellular heterogeneity.

Three possible scenarios can occur:

(i) In high-advection zones, such as the impeller region—where cells from different parts of the reactor are drawn in and thoroughly mixed— cells with varying intracellular states come together, as schematically shown in Figure 2.1a. If the adaptation period is on the same order as the residence time governed by advection, microorganisms do not have enough time to reach equilibrium with the extracellular conditions before leaving the location, leading to intracellular heterogeneity.

(ii) If the cells' adaptation rate to their environment is much faster than the advection rate, they will equilibrate with their surroundings almost instantly, as schematically shown in Figure 2.1b. It should be noted that intracellular heterogeneity has often been overlooked in simulations due to the complexity of the model and the considerable computational resources required. Neglecting intracellular heterogeneity is equivalent to assuming that microorganisms instantaneously adjust their intracellular states in response to environmental conditions.

(iii) If homogeneous microbes have an adaptation rate several orders of magnitude lower than the advection rate, they are effectively "swept along" by the flow, leaving insufficient time for internal differences to develop.



(a) (b)

Figure 2.1: Microorganisms traverse different zones in the bioreactor: (a) a scenario where the cell adaptation rate is of the same order of magnitude as the advection rate, causing states that differ from the zone environment, leading to heterogeneity in the mixing zone; (b) a scenario where cells immediately adjust their states to match the zone environment, resulting in a homogeneous population in the mixing zone.

In summary, intracellular heterogeneity arises when the adaptation rate is comparable to the advection rate. When the adaptation rate is significantly faster *or* slower than the advection rate, the population tends to exhibit less heterogeneity.

2.1.2 Responses of microorganisms exhibiting concentration gradients and intracellular heterogeneity

To this point, the intracellular properties of microorganisms have been treated as a single, lumped unit; however, in reality, they adapt across multiple cellular levels:

- **Metabolome:** The set of small molecules produced during metabolism, such as amino acids, lipids, and glucose, which reflect the cell's biochemical state and are influenced by genetics and the environment.
- **Proteome:** The complete set of proteins, including enzymes such as lactase and ribonuclease, that perform cellular functions and vary with gene expression and environmental conditions.
- **Transcriptome:** The collection of RNA molecules transcribed from the genome, such as messenger RNA (mRNA) for hemoglobin or regulatory RNAs such as microRNAs, providing a dynamic snapshot of gene activity in response to changing conditions.
- **Genome:** The organism's complete genetic material, which encodes instructions for biological processes, such as genes for antibiotic resistance or regulatory regions that control stress responses.

Adaptation at one level affects the others and occurs over timescales ranging from seconds to days, contributing to the complexity of cellular heterogeneity [23].

Overflow metabolism is one of the most well-known examples of proteome adaptations, which subsequently affect the metabolome and occur on a characteristic timescale of seconds. This metabolic process uses the fermentation pathway under conditions of high substrate concentrations, even in the presence of oxygen. Basan et al. [24] found that pathway switching results from efficient proteome allocation. Specifically, for the same energy flux, fermentation requires a greater carbon flux than respiration, yet it uses a smaller quantity of proteins. This metabolic process results in reduced biomass yield and the accumulation of undesirable by-products, which negatively affect cell physiology and

process performance (e.g., high acetate concentrations inhibit growth in *Escherichia coli* [25, 26]).

Alterations in gene expression at the transcriptome level due to environmental stress have been documented in the literature on a timescale of seconds. The magnitude of these changes has been found to be location-dependent within the reactor [27]. Lara et al. [28] discovered that alternating between anaerobic and aerobic conditions leads to significant alterations in gene expression. Transcriptome levels of mixed-acid fermentation genes increased by 1.5- to over 6-fold under oscillatory conditions compared to aerobic cultures. These changes were found to be related to an increase in energy demand (i.e., ATP demand for maintenance in *Escherichia coli* was reported to increase by 40–50% [29]). The practical consequence of this is that carbon sources are preferentially consumed to sustain the cell rather than to support product formation and growth.

Finally, prolonged exposure to stressful environmental conditions (ranging from hours to days) may increase the mutation rate (i.e., changes in the genome), leading to population heterogeneity. Natural selection plays a role in this process, as the subpopulations best able to endure fluctuations will dominate the system. Studies have shown that after exposure to fluctuating environmental conditions, certain species (including *Escherichia coli* [30, 31]) exhibit higher viability but lower biomass yield than cells grown in a uniform environment.

2.1.3 Approaches for investigating concentration gradients and intracellular heterogeneity

Numerous studies have investigated concentration gradients and intracellular heterogeneity in industrial-scale bioreactors. Nadal-Rey et al. [23] identified two main experimental techniques for studying these gradients: direct measurement and scale-down experiments. To quantify microbial culture heterogeneity, flow cytometry has been employed in conjunction with these methods.

Direct measurement can provide insights into the magnitude and impact of gradients on existing commercial processes. The conventional approach to measuring gradients in large-scale systems involves placing multiple online probes and sampling from various ports at different heights within the bioreactor. Although measuring gradients at large scales is highly valuable for researchers, it remains infrequent. The primary reason for this scarcity is limited access to industrial reactors for research and the infrequent release of authentic industrial data, which is often restricted due to confidentiality concerns. Even though pilot facilities may be accessible, they rarely capture meaningful gradients unless operated under suboptimal conditions. Additionally, pilot-scale data provide limited insights into gradient

formation and the potential challenges at industrial-scale. Consequently, measuring gradients in large-scale bioreactors remains a particularly compelling research area, though it is challenging to achieve.

In scale-down experiments, laboratory-scale bioreactors are used to replicate the local conditions that microbes encounter along their advection path in an industrial system [32]. These experiments allow precise control over environmental fluctuations and enable easier sample collection and analysis compared to large-scale systems.

These techniques are sufficient for investigating the relationship between environmental gradients and key process performance; however, they do not adequately address how cell cultures respond to these gradients, which is crucial for understanding the overall process. To gain a complete understanding, flow cytometry is essential.

Flow cytometry [33] is an advanced analytical technique that measures multiple physical and chemical characteristics of individual cells, such as size, granularity, and intracellular content, as they flow in suspension through a detection system. Flow cytometry operates based on the light-scattering properties of cells, which can be enhanced by dyes or monoclonal antibodies that specifically bind to extracellular surface molecules or intracellular targets. Modern flow cytometers can analyze thousands of cells per second and, when configured as cell sorters, actively separate and isolate cells with specific optical features at comparable speeds. This capability enables rapid and detailed analysis of heterogeneity within complex cell populations.

Although the techniques discussed have advanced our understanding of gradients and heterogeneity in large-scale bioreactors, the complexity of environmental factors cannot be fully replicated in laboratory-scale models, and direct measurements remain limited due to restricted access. A computational model is essential for bridging these gaps, providing a framework to simulate and predict the complex interactions between microbial cells and environmental gradients in real-time. Such a model empowers researchers to investigate the gradients and heterogeneity that arise in large bioreactors.

Moreover, Computational approaches significantly reduce the resources required for extensive design processes. By simulating a range of conditions, they facilitate bioreactor design optimization, help identify the most efficient configurations, and support the exploration of optimal operating parameters. As a result, these models are invaluable for advanced process control and improving bioprocess performance.

The simulation framework for an aerated stirred bioreactor integrates hydrodynamic and metabolic models. While the hydrodynamic model serves as a conservation-based framework governing the transport of fluids, solutes, and microbes within the system, the

metabolic model captures microbial responses to environmental conditions, detailing the complex interplay between cellular metabolism and external factors.

2.2 Hydrodynamic models

To simulate a lab-scale bioreactor, a "well-mixed" reactor model is frequently used. However, it is quickly dismissed as a viable option for the hydrodynamic model upon analyzing the gradients observed in the large-scale bioreactor.

To capture spatial gradients within a bioreactor, models must incorporate more than two independent calculation nodes (or degrees of freedom) positioned at distinct locations in the reactor. These nodes allow the model to resolve variations in key variables—such as concentration, velocity, and temperature—across the reactor. Two common modeling approaches that meet this requirement are continuum models and compartmental models, both of which are discussed in this section. Among them, the two-fluid model—a specific type of continuum model—is highlighted as the primary hydrodynamic framework employed in this study.

The models discussed here incorporate a one-way coupling assumption, simplifying the analysis. Stirring strongly influences the hydrodynamics of all phases, resulting in a convection-dominated regime throughout most of the bioreactor. This regime justifies the one-way coupling approximation, where the hydrodynamics of each phase are independent of intraphase properties such as concentration and temperature. As a result, the fluid primarily functions as a transport medium for these properties within the system.

2.2.1 Two-fluid models

Continuum models are used to describe systems where the physical quantities (like concentration, velocity, or temperature) are assumed to vary continuously across space and time. This assumption holds true when these quantities can be averaged or "smeared out" over a small volume, allowing for the neglect of microscopic details (like molecular interactions) in favor of focusing on macroscopic behavior. These models achieve high spatial and temporal resolution by solving systems of partial differential equations (PDEs). However, this detailed representation results in significant computational demands. Computational Fluid Dynamics (CFD) techniques, which are a collection of numerical methods used to solve this type of model, are often employed. Some well-known examples of continuum

models include the Navier-Stokes equations, the two-fluid model, and the Euler-Lagrange model.

In bioreactor simulations, the liquid and bubble phases are typically considered. Microorganisms are often treated using a one-way coupling approach, since their density is similar to that of the liquid and their small size yields a low Stokes number. Consequently, inertial effects are minimal, allowing microorganisms to precisely follow the fluid flow and justifying the use of a single velocity field for the fluid-cell mixture.

Two-fluid models (also known as Euler-Euler models), a type of continuum model, are frequently applied to investigate these systems. These models treat both the liquid and gas (bubble) phases as continuous, each with its set of equations describing mass, momentum, and energy transport. The two phases are assumed to interact, but they remain separate within the flow field. By solving the two-fluid model, the volume fraction profile, $\alpha_k(t, \mathbf{s})$, and the velocity profile, $\mathbf{u}_k(t, \mathbf{s})$, for each phase k at various spatial locations, $\mathbf{s}$, and times, t, are determined. These variables provide key hydrodynamic information about the system.

Equation (2.1) and Equation (2.2) show the mass and momentum balance equations, respectively, used in the two-fluid model [3].

$$\underbrace{\frac{\partial}{\partial t}\left(\alpha_k \rho_k\right)}_{\text{Accumulation}} + \underbrace{\boldsymbol{\nabla} \cdot \left(\alpha_k \rho_k \mathbf{u}_k\right)}_{\text{Convection}} = \underbrace{\mathcal{I}_k}_{\substack{\text{Interphase} \\ \text{mass transfer}}} \tag{2.1}$$

$$\begin{aligned} \underbrace{\frac{\partial(\alpha_k \rho_k \mathbf{u}_k)}{\partial t}}_{\text{Accumulation}} + \underbrace{\boldsymbol{\nabla} \cdot \left(\alpha_k \rho_k \mathbf{u}_k \mathbf{u}_k\right)}_{\text{Convection}} = \underbrace{\alpha_k \rho_k \mathbf{g}}_{\substack{\text{Gravitational} \\ \text{force}}} \underbrace{-\boldsymbol{\nabla}(\alpha_k p_k) - \boldsymbol{\nabla} \cdot \left(\alpha_k \boldsymbol{\tau}_k\right)}_{\substack{\text{Molecular forces} \\ \text{(pressure force and viscous stresses)}}} \\ \underbrace{-\boldsymbol{\nabla} \cdot \left(\alpha_k \boldsymbol{\tau}_k^t\right)}_{\text{Reynolds stresses}} + \underbrace{\mathcal{M}_k}_{\substack{\text{Interphase} \\ \text{momentum transfer}}} \end{aligned} \tag{2.2}$$

where

α_k : Volume fraction of phase $k \in \{L, G\}$ (L: liquid and G: gas).

$\mathbf{u}_k$: Column vector representing the phasic velocity of phase k.

ρ_k : Phasic density of phase k.

$\mathbf{g}$: Column vector of the gravitational force.

p_k : Phasic pressure for phase k.

$\boldsymbol{\tau}_k$: Tensor representing the viscous forces in phase k.

$\boldsymbol{\tau}_k^t$: Tensor representing the turbulent viscous stresses (or Reynolds stresses) in phase k.

To solve these equations, it is necessary to consider the energy balance equation, the turbulence model, initial and boundary conditions, and constitutive relations for $\boldsymbol{\tau}_k$, $\boldsymbol{\tau}_k^t$, and the interphase mass, momentum, and energy transfer, denoted by $\mathcal{I}_k$ for the mass transfer and $\mathcal{M}_k$ for the momentum transfer.

In many two-fluid model solvers, the continuous phase (often a liquid) is assumed to be incompressible, meaning its density is constant, while the dispersed phase (often a gas) is treated as compressible, allowing its density to vary with pressure. It is common to assume a single pressure field $p_k = p$ that all phases share. This simplifies the coupling between the phases. Moreover, because bioreactors are typically operated at a controlled, constant temperature well below the boiling point, phase changes are negligible. As a result, the energy equation is omitted, and $\mathcal{I}_k$ is assumed to be zero. However, $\mathcal{M}_k$ cannot be neglected, as it is critical for capturing the interaction between the hydrodynamics of the two phases. Interphase momentum transfer mechanisms include drag, lift, virtual forces, and surface tension. For further details on these constitutive relations, standard textbooks on multiphase flow systems are recommended [34].

Using only a two-fluid model is insufficient to simulate the dynamics of stirred bioreactors, as it does not account for the complexities introduced by the motion of the impellers, which are crucial for mixing and maintaining flow patterns in the reactor. To address this, additional numerical techniques are needed to capture the effects of impeller motion. Two common methods used to model impeller-induced flow in stirred bioreactors are the multiple reference frame (MRF) method and the sliding mesh (SM) method. Both techniques aim to simulate the movement of impellers and their interaction with the fluid, but they differ in how they approach the problem. Notably, both methods require the domain to be partitioned into multiple mesh zones of rotating zones (where the impellers are located) and a stationary zone.

In the rotating zone of MRF method [35, 36, 37], the governing equations are solved in a rotating reference frame, which introduces centrifugal and Coriolis forces into the momentum balance. The steady-state solution obtained does not resolve the transient behavior of the impeller but captures its general effects on the flow, making the method computationally efficient for many practical applications.

In contrast, the SM method [38, 39, 40] provides a more detailed representation of impeller motion by modeling the rotation of the impeller and the surrounding fluid in a time-dependent manner. As the impeller rotates, the mesh in the region of the impeller slides relative to the rest of the domain. This technique captures of transient effects, such as vortex formation and unsteady flow behavior, and can be more accurate in predicting dynamics in regions near the impeller. However, it is computationally more demanding due to the need for remeshing and interpolation at each time step.

While both the MRF and SM methods can be used to model impeller motion in stirred bioreactors, the choice between them depends on the trade-off between accuracy and computational efficiency.

A range of software packages provide a ready-to-use platform for the model discussed in this section, including the open-source OpenFOAM [41] and commercial options such as ANSYS Fluent [42] and COMSOL [43].

2.2.2 Compartmental models[1]

A traditional compartmental model (CM) consists of a network of interconnected single-phase reactors, referred to as compartments, which are fixed control volume elements. Each specified compartment exchanges mass and energy with other compartments through intermediate flows. The determination of the number of compartments, the configuration of the network, and the choice of the representative reactor type (commonly CSTRs) within the network are typically informed by an understanding of global flow behavior, which is often derived from the analysis of simulations. Once the network is established, the rates of mass and energy transfer between compartments can be calibrated by comparing them with the residence time distribution (RTD) analysis obtained from tracer experiments [44].

Under the assumptions of one-way coupling and well-mixed single phase compartments, the mass balance equation for substances in each compartment, j, is frequently expressed

[1]This section is modified from [1].

in the form of Equation (2.3).

$$\underbrace{v_j \frac{d\mathbf{c}_j}{dt}}_{\text{Accumulation}} = \underbrace{\sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} (q_{j' \to j} \mathbf{c}_{j'}) - \mathbf{c}_j \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'}}_{\text{Mass flow between compartments}} + \underbrace{q_{ex \to j} \mathbf{c}_{ex \to j}}_{\text{Inlet mass flow}} - \underbrace{q_{j \to ex} \mathbf{c}_j}_{\text{Outlet mass flow}} + \underbrace{\mathbf{r}_j v_j}_{\text{Reaction}} + \underbrace{\mathbf{g}_j v_j}_{\text{Others}} \tag{2.3}$$

where

$\mathbf{c}_j$: Column vector of substance concentration in Compartment j.

v_j : Volume of Compartment j.

$q_{a \to b}$: Volumetric flow rate from Compartment a to Compartment b. Note that the subscript j' denotes compartments apart from Compartment j, and the subscript ex indicates flow entering (i.e., $ex \to j$) due to feeding or leaving the system (i.e., $j \to ex$) through outlet streams.

$\mathbf{r}_j$: Column vector of the total reaction rate of substances in Compartment j.

$\mathbf{g}_j$: Column vector of total rates apart from reaction, including interphase mass transfer, for substances in Compartment j.

The volumetric flow rate $q_{a \to b}$ is generally extracted from the calibration of the predicted results with measured RTD data. A more detailed discussion on the calculation of $q_{a \to b}$ is provided in Section 3.2.2. To prevent misunderstanding while using the equation, it is important to emphasize that $q_{a \to b}$ must be a positive number according to the flow directionality convention used inEquations (2.3), (2.34) and (2.49).

In addition to Equation (2.3), there is a more compact representation of the substance conservation balance equation that accounts for all compartments in the system, expressed in the form of Equation (2.4).

$$\underbrace{\frac{d\mathbf{C}}{dt}\mathbf{V}}_{\text{Accumulation}} = \underbrace{\mathbf{CQ}}_{\substack{\text{Mass flow} \\ \text{inside the domain}}} + \underbrace{\mathbf{C}^{in}\mathbf{Q}^{in}}_{\text{Inlet mass flow}} - \underbrace{\mathbf{CQ}^{out}}_{\text{Outlet mass flow}} + \underbrace{\mathbf{RV}}_{\text{Reaction}} + \underbrace{\mathbf{GV}}_{\text{Others}} \tag{2.4}$$

The variables and parameters represented in this equation are organized in a matrix, and their definitions can be found in Table 2.1

Table 2.1: Variables and parameters in Equation 2.4. n_s denotes the number of substrates. n_{EC} denotes the number of compartments.

Variables/ Parameters	Size	Definition
$\mathbf{C}$	$\mathbb{R}^{n_s} \times \mathbb{R}^{n_{EC}}$	Concentration matrix with its element $c_{i,j}$ being the concentration of Substrate i in Compartment j
$\mathbf{C}^{in}$		Feed concentration matrix with its element $c^{in}_{i,j}$ being the concentration of Substrate i in Compartment j
$\mathbf{R}$		Reaction matrix with its element $r_{i,j}$ being the overall reaction rate of Substrate i in Compartment j
$\mathbf{G}$		Total rate matrix other than the reaction, including interphase mass transfer with its element $g_{i,j}$ being the overall rate of Substrate i in Compartment j
$\mathbf{Q}$	$\mathbb{R}^{n_{EC}} \times \mathbb{R}^{n_{EC}}$	Internal volumetric flow rate matrix with its element $q_{j,j' \neq j}$ being the volumetric flow rate from Compartment j to Compartment j' and element $q_{j,j}$ being the total volumetric flow rates leaving compartment i to other compartments
$\mathbf{Q}^{in}$		Inlet volumetric flow rate diagonal matrix with its element $q^{in}_{j,j}$ being the total inlet volumetric flow rate entering compartment i from outside of the system
$\mathbf{Q}^{out}$		Outlet volumetric flow rate diagonal matrix with its element $q^{out}_{j,j}$ being the total outlet volumetric flow rate leaving compartment i to outside of the system
$\mathbf{V}$		Volume diagonal matrix with its element $v_{j,j}$ being the volume of Compartment j

An advanced compartmental model can be automatically constructed when a steady-state hydrodynamic CFD simulation is provided; this method is known as the CFD-derived CM. The method has shown an improvement in the determination of flow-related parameters, e.g., the volumetric flow rate in $\mathbf{Q}$. Further details are elaborated upon in Section 3.2.

2.3 Metabolic models

Life is inherently complex, even for microbes, which represent some of the simplest cellular organisms. *Escherichia coli* (*E. coli*), the most extensively studied prokaryotic model organism, contains over 2,600 metabolites. These metabolites are interconnected through approximately 125 unique pathways and more than 2,800 reactions [45]. This complexity makes the modeling of biochemical reactions in bioreactors particularly challenging. The difficulty increases when intracellular heterogeneity is considered, as it requires accounting not only for the states of a single cell but also for all possible states within the population.

Fredrickson et al. [19] provided a comprehensive classification of mathematical models for bioprocesses. They use the term "segregated" models to describe those that explicitly account for the presence of heterogeneous individuals within a population of microorganisms, and "structured" models for formulations that represent cell responses through various cellular states. Based on their classification, models for pure cultures can be divided into four types, as illustrated in Figure 2.2.

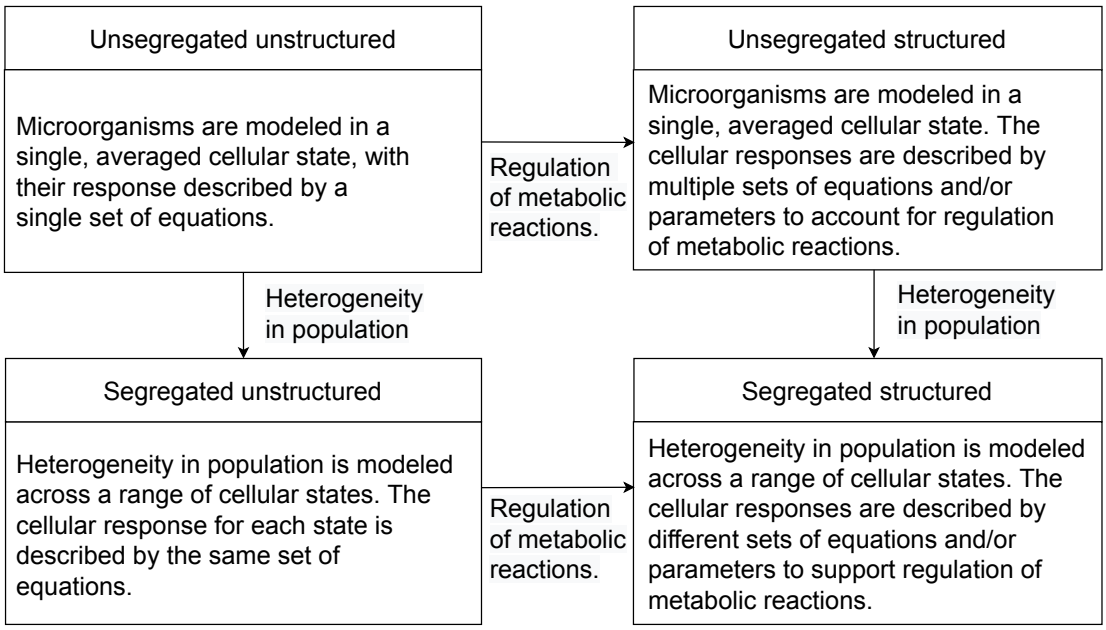


Figure 2.2: Classification of metabolic model

Among these models, the unsegregated-unstructured type (lacking cell-to-cell variability and internal state structure) is the most commonly used to describe cell cultures in

lab-scale bioreactors. Microorganisms are modeled at a single, averaged cellular state. This approach assumes homogeneity across the cell population, meaning that each cell is considered to be in the same state without accounting for individual cellular variations in response to environmental changes. As a result, a single set of equations, often based on Monod-type kinetic models, is used to describe the cellular responses and dynamics, reflecting the average behavior of the entire population. However, due to its inability to reflect microorganism regulation and adaptation to an environment, this model type is often regarded as inadequate for modeling industrial bioreactors [19].

The development of metabolic flux analysis [46, 47, 48] has significantly advanced our ability to identify and understand the metabolic pathways and regulatory feedback mechanisms that govern microbial behavior in response to changing extracellular conditions. Leveraging this knowledge, structured models can be developed that incorporate switching between different kinetic models and parameters, allowing for a more accurate representation of microbial metabolism. Furthermore, advancements in technology for monitoring single-cell characteristics (i.e., flow cytometry, microscopy, and automated image analysis) have made it feasible to observe the distribution of cell heterogeneity within microbial cultures. Leveraging this information, segregated models that can capture the behavior of a heterogeneous dynamic microbial population can also be constructed and validated.

Undoubtedly, the segregated-structured model (considering cell-to-cell variability and internal state structure) is the most advanced and computationally costly approach. Owing to its capacity to consider both the heterogeneity of cellular characteristics within a population and the metabolic processes linked to these distinct characteristics, this model appears to be the most effective at simulating a large-scale bioreactor.

In this research, dynamic flux balance analysis model (DFBA), a subset of constraint-based metabolic models, was detailed as an unsegregated structured model. In another study, the population balance model, the most general segregated model for representing heterogeneity within a microorganism population, was developed in combination with a structured model.

2.4 Constraint-based metabolic models

The metabolic processes of microorganisms can be characterized by a complex network of biochemical reactions, where various substances are transformed into numerous distinct biomolecules, aided by a range of enzymes that act as catalysts. For an illustrative example, see Figure 2.3. Numerous quantitative techniques have been proposed for developing a

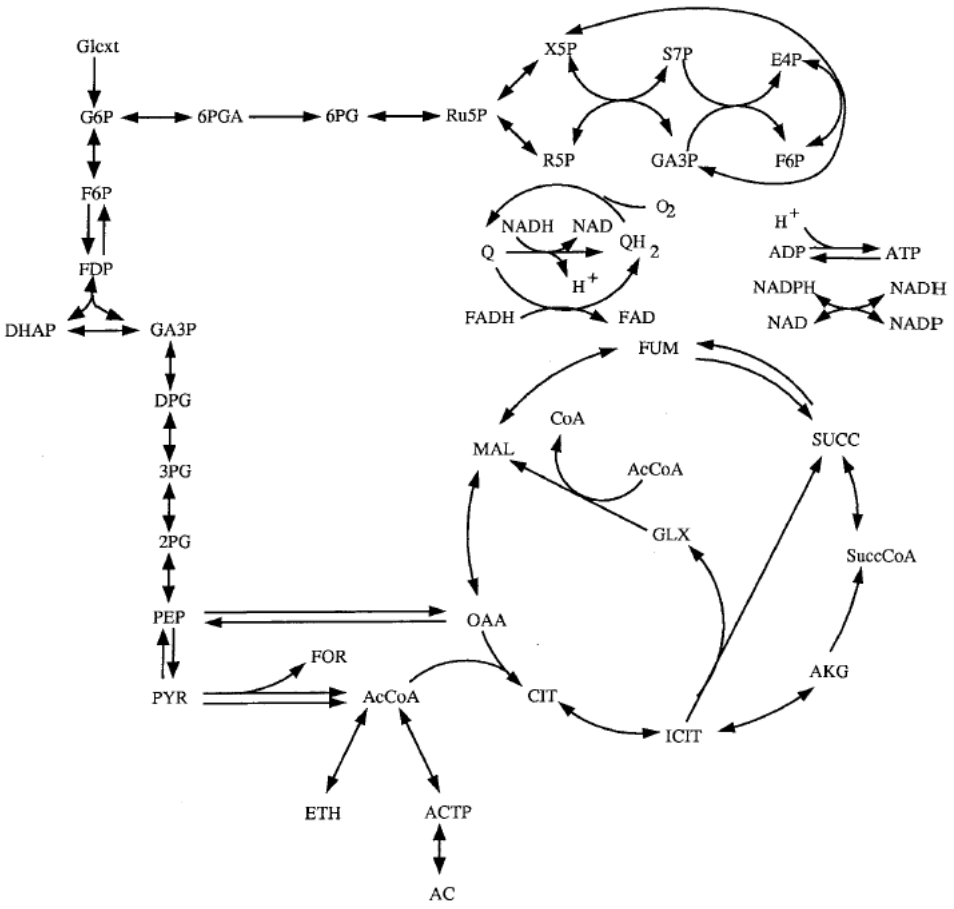


Figure 2.3: A diagram of the metabolic network of *E. coli*, taken from Mahadevan et al. [20], comprises 54 metabolites and 85 reactions. It includes pathways of glycolysis, the pentose phosphate pathway, the TCA cycle with the glyoxylate bypass, anapleurotic reactions, and redox metabolism.

dynamic model of the metabolism of microorganisms, the majority of which require kinetic equations to describe each metabolic reaction in the network. Accordingly, a complete dynamic model (at the genome scale, for example) may have up to a thousand kinetic equations and accompanying kinetic parameters, the majority of which are unknown to the researcher and impossible to obtain with the typically limited data sets available for model calibration.

To overcome this difficulty, more recent studies have simulated microorganism metabolic networks using constraint-based models. One significant advantage of such models is that they only require knowledge of the metabolic network and require less information than models which describe each reaction with a kinetic expression. The term "constraint-based" refers to techniques that employ constraints to restrict the infinite solution space of metabolic reactions within a metabolic network, as illustrated in Figure 2.4. The constrained feasible region is then obtained, which may be thought of as the potential behavior of the microorganism for a possible range of environmental conditions.

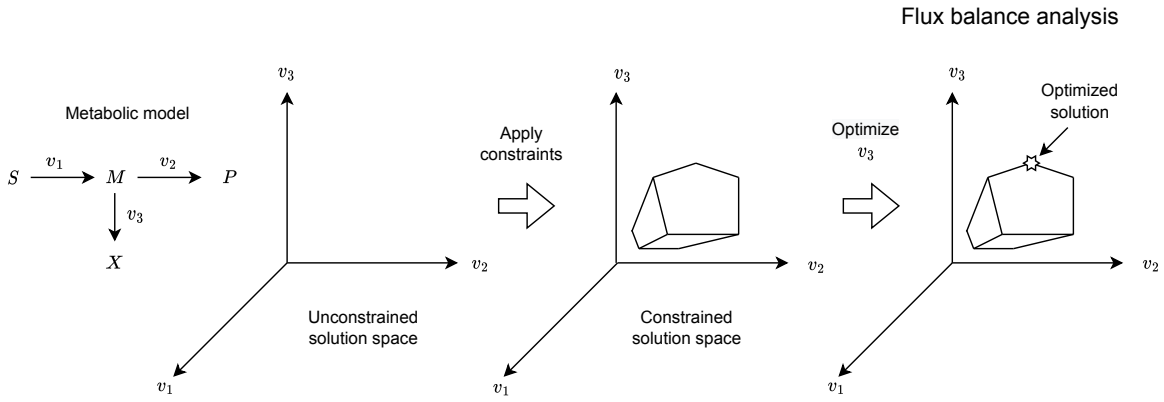


Figure 2.4: Constraint-based metabolic models

2.4.1 Flux balance analysis

Flux balance analysis (FBA) is an example of a constraint-based modeling technique. It has gained widespread popularity for simulating large-scale metabolic networks because it enables the determination of reaction fluxes for underdetermined systems, where the number of constraints is less than the number of reactions in the network.

FBA is formulated as a constrained optimization problem in which the cell is assumed to optimally allocate resources to maximize a biological objective. The objective function is chosen as a surrogate for the most plausible physiological state of the microorganism culture. Many objective functions have been adopted in the literature, including maximization of biomass growth rate, and minimization or maximization of ATP generation. A comparison of the accuracy of various objective functions is available in the literature [49].

Initially, FBA was a widely used technique for studying biochemical networks at pseudo-steady state, using genome-scale metabolic network reconstructions built over the last decade. However, the method was later extended to describe dynamic cell behavior (e.g., Verma and Palsson [50] and Mahadevan et al. [20]). The literature demonstrates that this dynamic flux balance analysis model (DFBA) has found application in process systems and control [51].

2.4.2 Dynamic flux balance analysis[2]

Dynamic flux balance analysis models (DFBAs) are formulated by integrating a FBA with mass conservation models. By combining time-varying FBA with dynamic mass conservation models, DFBA models can predict changes in metabolic pathways and concurrently track changes in metabolite concentrations over time. Two ways to formulate DFBA models have been reported [20]: static and dynamic optimization. The static optimization approach involves performing a static FBA at each location and time step, which is frequently adopted due to its computational simplicity. The dynamic optimization approach involves solving a nonlinear programming problem for the entire time interval of operation, significantly increasing computational complexity and, more importantly, making it infeasible for incorporation with traditional multiphase hydrodynamic simulations.

Focusing on a static optimization approach, the concentrations of metabolites serve as the state variables. For any discretized space j (not limited to spatial compartments in a compartmental model) at time t, the state vector $\mathbf{c}_j(t) \in \mathbb{R}^{n_x}$ represents the local concentrations of these metabolites. The metabolic flux vector $\mathbf{v}_j(t) \in \mathbb{R}^{n_v}$ is related to the state variables through an FBA, which is formulated as a linear programming problem:

$$\max_{\mathbf{v}_j} \quad \mathbf{e}^\top \mathbf{v}_j \tag{2.5a}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{v}_j \leq \mathbf{F}\theta(\mathbf{c}_j) + \mathbf{g} \tag{2.5b}$$

[2]This section is modified from [1].

The objective function of the microorganisms is represented by Equation (2.5a), where the coefficient vector of the objective function is denoted by $\mathbf{e} \in \mathbb{R}^{\mathrm{n}_v}$. Equation (2.5b) denotes the set of limiting kinetic constraints on the consumption/production of metabolites. Although it is expressed in linear form, nonlinear dependencies on the dynamic states are encapsulated within the column vector of the nonlinear function, $\theta(\mathbf{c}_j) \in \Theta \subseteq \mathbb{R}^{\mathrm{n}_\theta}$. Here, $\mathbf{A} \in \mathbb{R}^{\mathrm{n}_{cs}} \times \mathbb{R}^{\mathrm{n}_v}$ and $\mathbf{F} \in \mathbb{R}^{\mathrm{n}_{cs}} \times \mathbb{R}^{\mathrm{n}_\theta}$ are coefficient matrices defining relationships among system variables, and $\mathbf{g} \in \mathbb{R}^{\mathrm{n}_{cs}}$ accounts for constant terms in the constraints. In the expression of each parameter, n_v, n_{cs}, and n_θ represent the numbers of metabolic fluxes, constraints, and state-dependent variables, respectively.

When using a DFBA to determine the metabolic reaction rates, Equation (2.6) is employed to compute the reaction rates of all species within the discretized space. For example, in Compartment j, these rates are represented as a column vector $\mathbf{r}_j$ within the matrix $\mathbf{R}$.

$$\mathbf{r}_j = \mathbf{S}\mathbf{v}_j^* c_{X,j} \tag{2.6}$$

where $\mathbf{S} \in \mathbb{R}^{n_s} \times \mathbb{R}^{n_v}$ represents the stoichiometric matrix, $\mathbf{v}_j^*$ denotes the optimal solution of metabolic flux vector, which is obtained by solving Equation (2.5), and $c_{X,j}$ represents the biomass concentration.

It is essential that constraint Equation (2.5b) prevents negative substance concentrations when solving the DFBA, ensuring that microorganisms cannot consume more substance than is available at a given location and time. These constraints are derived by explicitly linearizing the accumulation term in the mass balance equation using a first-order Taylor expansion in time. For instance, linearizing and rearranging the accumulation term in Equation (2.3) yields:

$$\begin{aligned} \mathbf{c}_j^{t+\Delta t} = \mathbf{c}_j^t &+ \frac{\Delta t}{v_j}\left(\sum_{j=1, j\neq j'}^{\mathrm{n_{EC}}} (q_{j'\to j}\mathbf{c}_{j'}^t) - \mathbf{c}_j^t \sum_{j=1, j\neq j'}^{\mathrm{n_{EC}}} q_{j\to j'}\right) \\ &+ \frac{\Delta t}{v_j}\left(q_{ex\to j}\mathbf{c}_{ex\to j}^t - q_{j\to ex}\mathbf{c}_j^t\right) \\ &+ \Delta t\left(\mathbf{r}_j + \mathbf{g}_j\right) \end{aligned} \tag{2.7}$$

where superscript t denotes the current time, Δt represents the time increment, and $t + \Delta t$ signifies the subsequent time step. The positive concentration constraint requires that:

$$\mathbf{c}_j^{t+\Delta t} \geq 0 \tag{2.8}$$

or

$$-\mathbf{r}_j \leq \frac{\mathbf{c}_j^t}{\Delta t} + \frac{1}{v_j}\left(\sum_{j=1,j\neq j'}^{\mathrm{n_{EC}}} (q_{j'\to j}\mathbf{c}_{j'}^t) - \mathbf{c}_j^t \sum_{j=1,j\neq j'}^{\mathrm{n_{EC}}} q_{j\to j'}\right) + \frac{1}{v_j}\left(q_{ex\to j}\mathbf{c}_{ex\to j}^t - q_{j\to ex}\mathbf{c}_j^t\right) + \mathbf{g}_j \tag{2.9}$$

Inserting Equation (2.6) for $\mathbf{r}_j$ and performing an additional rearrangement, to derive:

$$-\mathbf{S}\mathbf{v}_j \leq \frac{\mathbf{c}_j^t}{c_{X,j}^t \Delta t} + \frac{1}{c_{X,j}^t v_j}\left(\sum_{j=1,j\neq j'}^{\mathrm{n_{EC}}} (q_{j'\to j}\mathbf{c}_{j'}^t) - \mathbf{c}_j^t \sum_{j=1,j\neq j'}^{\mathrm{n_{EC}}} q_{j\to j'}\right) + \frac{1}{c_{X,j}^t v_j}\left(q_{ex\to j}\mathbf{c}_{ex\to j}^t - q_{j\to ex}\mathbf{c}_j^t\right) + \frac{\mathbf{g}_j}{c_{X,j}^t} \tag{2.10}$$

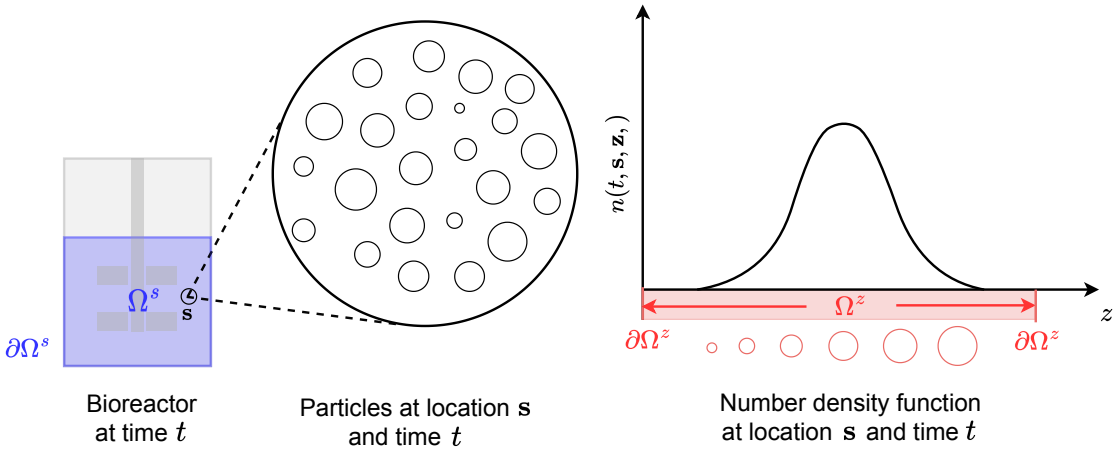
Equation (2.10) represents the positive concentration constraint that must be incorporated into the set of constraints Equation (2.5b) for DFBA coupled with a compartmental model. This linearized constraint will introduce an error into the optimal solution only when it becomes active. In such cases, the resulting local truncation error per time step is of order $O(\Delta t^2)$, corresponding to the first-order Taylor expansion in time.

2.5 Population balance models

Two well-known approaches to investigate heterogeneity within the microbe population are the Euler-Lagrange multiphase models and population balance models. The Euler-Lagrange multiphase models [7] track numerous individual cells and continuum fields, leading to challenges in implementation and significant computational complexity, especially in large-scale or highly detailed systems.

Because of these challenges, researchers often turn to population balance models [52]. These models are especially valued for their capability to handle distributions of intracellular properties across the population without requiring detailed cell tracking, making them a more computationally efficient choice for modeling microbial population heterogeneity.

This section will elaborate on the number density function as a tool for quantifying heterogeneity, and on the population balance equation, which governs the dynamics of the number density function, as a framework for tracking heterogeneity within the bioreactor.



considering

Figure 2.5: Number density function of cells characterized by size as an internal state

2.5.1 Number density functions

Microorganisms, as dispersed particles, engage in interactions with one another and with the carrier fluid as they traverse through it. This behavior can dynamically alter the properties of each cell. To quantify these properties at a given moment of time t, we define a set of cell state variables which characterize the condition of each cell within the population. Mathematically, these state variables are expressed as a vector $\mathbf{x}$, where the states are assumed to occur within the state space $\Omega \in \mathbb{R}^{n_x}$, where n_x denotes the number of dimensions. The state space is a combination of two distinct subspaces, $\Omega = \Omega^s \times \Omega^z$. $\Omega^s \in \mathbb{R}^3$ defines the geometric spatial domain. In this study, the coordinate represented by the column vector $\mathbf{s}$ defines the centroid position of an infinitesimal control volume containing the cell in spatial dimensions. $\Omega^z \in \mathbb{R}^{n_z}$ defines a space of internal variables. For this subspace, the coordinate represented by the column vector $\mathbf{z}$ represents the internal states of the cell, such as size, mass, age, rate, and other attributes, that in this study are assumed to fully characterize the status of each cell.

The intracellular heterogeneity accounts for differences in internal states $\mathbf{z}$ between dispersed cells in their population at time t and location $\mathbf{s}$. The heterogeneity can be quantified by the number density function, $n(t, \mathbf{s}, \mathbf{z})$, which represents a continuous function of the cells' number density function (number per n-dimensional hypervolume of Ω) at t and $\mathbf{s}$ over the internal states $\mathbf{z}$; refer to Figure 2.5.

Based on the definition of n, one can further specify the total cell count, $N(t)$, within

the enclosed regions of Ω^s: a defined area within the spatial dimensions, and Ω^z: a bounded range of considered properties as

$$N = \int_{\Omega^z} \int_{\Omega^s} n \, d\mathbf{s} d\mathbf{z} \tag{2.11}$$

As Ω^s and Ω^z are independent, the order of integration is not significant.

Moreover, if Ω^s is discretized into $\mathrm{n_{EC}}$ compartments, each denoted by $\Omega^s_j \subset \Omega^s$ for $j \in \{1, 2, \cdots, \mathrm{n_{EC}}\}$, then the total count of cells within Compartment j, which encompasses the space $\Omega^s_j \times \Omega^z$, $N_j(t)$, can be expressed as follows:

$$N_j = \int_{\Omega^z} \int_{\Omega^s_j} n \, d\mathbf{s} d\mathbf{z} \tag{2.12}$$

Multiplying Equation (2.12) by v_j/v_j, where v_j is the volume of Compartment j, and then rearranging terms and grouping variables, yields:

$$N_j = v_j \int_{\Omega^z} n_j d\mathbf{z} \tag{2.13}$$

where

$$n_j = \frac{\int_{\Omega^s_j} n d\mathbf{s}}{v_j} \tag{2.14}$$

which is the averaged number density function over Compartment j, meaning that $n_j(t, \mathbf{z})$ is not any longer continuous regarding the $\mathbf{s}$ coordinate.

Building upon n_j, it is also straightforward to ascertain that if Ω^z is discretized into n_b bins, each denoted by $\Omega^z_\gamma \subset \Omega^z$ for $k \in \{1, 2, \cdots, \mathrm{n}_b\}$, then the total number of cells within Compartment j and within a specific internal range, Bin γ, $N_{j,\gamma}(t)$, can be expressed. This count now encompasses the regions of interest $\Omega_{j,k} = \Omega^s_j \times \Omega^z_\gamma$ and is expressed as:

$$N_{j,\gamma} = v_j \int_{\Omega^z_\gamma} n_j \, d\mathbf{z} \tag{2.15}$$

Multiplying Equation (2.15) by v^z_k/v^z_γ, where v^z_γ is the size of Bin γ, to yield:

$$N_{j,\gamma} = v_j v^z_\gamma n_{j,\gamma} \tag{2.16}$$

where

$$n_{j,\gamma} = \frac{\int_{\Omega^z_\gamma} n_j d\mathbf{z}}{v^z_\gamma} \tag{2.17}$$

which is the averaged number density function over Compartment j and Bin γ, meaning that $n_{j,\gamma}(t)$ loses continuity in both $\mathbf{s}$ and $\mathbf{z}$ due to its averaged definition.

It is important to note that for those modeling intracellular heterogeneity via the number density function, n, and considering the discretization of the state space Ω, it is crucial to comprehend not only the definition of n but also its averaged forms, n_j and $n_{j,\gamma}$.

2.5.2 Population balance equations

The dynamic evolution of intracellular heterogeneity of cells can be described by tracking the dynamics in the number density function, n. This can be achieved by solving the dynamic population balance equation as in [52], Equation (2.18).

$$\underbrace{\frac{\partial n}{\partial t}}_{\text{Accumulation}} + \underbrace{\boldsymbol{\nabla}^z \cdot \boldsymbol{\zeta} n}_{\substack{\text{Change of} \\ \text{internal state of} \\ \text{cells}}} + \underbrace{\boldsymbol{\nabla}^s \cdot \mathbf{u} n}_{\substack{\text{Transport} \\ \text{of cells}}} = \underbrace{h}_{\substack{\text{Generation} \\ \text{of cells}}} \tag{2.18}$$

where

- $\boldsymbol{\zeta}$: Rate of change of the internal state of cells.
- $\mathbf{u}$: Velocity field of cells in physical space.
- h: Net rate of cell generation.

The rate of change of internal state of cells (population adaptation rate), described by the term $\boldsymbol{\zeta}(\mathbf{z}, \mathbf{c}) = [\zeta_1, \cdots, \zeta_{\mathrm{n}_k}]^\top$ in Equation (2.18), primarily accounts for mechanisms that redistribute the population across different internal states while preserving the total number (not the number density) of cells. Examples include internal reactions within the cell and exchange of mass between the intracellular and extracellular environments. This rate is assumed to depend on the internal state, $\mathbf{z}$, the number density function, n, and the extracellular concentration, $\mathbf{c}$.

The velocity field of the cells, $\mathbf{u}(t, \mathbf{s})$, can be approximated by the velocity field of the liquid, $\mathbf{u}_L(t, \mathbf{s})$, if the cells exhibit a low Stokes number. In this case, inertial effects are negligible, and the cells closely follow the fluid flow. This enables solving the two-fluid model first to obtain the velocity field for the liquid, which can then be used for cells in the population balance equation. If the approximation is invalid, a more sophisticated

model, such as a multiphase model coupled with the population balance equation, must be employed.

The net rate of generation of cells given by $h(\mathbf{z}, \mathbf{c}, n)$, represents mechanisms such as cell birth, death, division, and aggregation. In a closed system, this is the sole term that accounts for changes in the total number (not the number density) of cells within the system. For microorganisms, only division and death are typically relevant. The expression for the net rate of generation of microorganism, $h(\mathbf{z}, \mathbf{c}, n)$, then, can be derived as Equation (2.19).

$$h = \underbrace{-bn}_{\substack{\text{Division rate} \\ \text{sink term}}} + \underbrace{\int_{\Omega^z} a' P b' n' d\mathbf{z}}_{\substack{\text{Division rate} \\ \text{source term}}} - \underbrace{qn}_{\text{Death rate}} \tag{2.19}$$

where

$(\cdot)'$: Superscript to denote variables of the parent cell states, where the state $\mathbf{z}$ is replaced by the state $\mathbf{z}'$.

b and b': Specific division rate.

a': Average number of cells formed from the division of a single parent cell. For cell division, $a' = 2$.

P: Partition probability density function for child cells with state $\underline{z}$, formed from the division of a parent cell with state $\underline{z}'$.

q: Specific death rate.

These parameters can be modeled as functions of the internal state, $\mathbf{z}$, as well as the environmental state, $\mathbf{c}$.

If following microbial division, the offspring are assumed to inherit all the state values of their parent, the partition probability function, as given in Equation (2.20), can be defined.

$$P = \prod_{k=1}^{\mathrm{n}_z} \delta(z_k - z_k') \tag{2.20}$$

Substituting P and $a' = 2$ into the expression for h results in:

$$\begin{aligned} h &= -bn + 2\int_{\Omega^z} Pb'n'\, d\mathbf{z} - qn \\ &= -bn + 2\int_{\Omega^z} \prod_{k=1}^{\mathrm{n}_z} \delta(z_k - z_k')b'n'\, d\mathbf{z} - qn \\ &= -bn + 2bn - qn \\ &= (b-q)n \end{aligned} \tag{2.21}$$

The specific division rate, $b(\mathbf{z}, \mathbf{c})$, and the specific death rate, $q(\mathbf{z}, \mathbf{c})$, in Equation (2.21) can be functions of either intracellular states, extracellular states, or constants.

Since some mechanisms accounted for in the population balance model depend on the external state, Equation (2.18) must be solved together with an external state balance equations. For dilute species—those that dissolve in the fluid without significantly affecting the hydrodynamics of the carrier fluid—a mass balance equation in the liquid phase, under convection-dominated conditions (where diffusion is negligible), can be expressed as:

$$\underbrace{\frac{\partial(\alpha_L \mathbf{c})}{\partial t}}_{\text{Accumulation}} + \underbrace{\nabla^s \cdot (\alpha_L \mathbf{c}\mathbf{u}_L)}_{\text{Advection}} = \underbrace{\alpha_L \mathbf{r}}_{\text{Reaction}} + \underbrace{\alpha_L \mathbf{g}}_{\text{Others}}. \tag{2.22}$$

The volume fraction field α_L and the velocity field $\mathbf{u}_L$ are obtained from solving the two-fluid model. The reaction term describes how the substrates interact with the population balance model. It is generally expressed as:

$$\mathbf{r} = \int_{\Omega^z} \mathbf{w} n d\mathbf{z} \tag{2.23}$$

where $\mathbf{w}$ represents a combination of interfacial reaction and mass transfer phenomena between the intracellular and the extracellular environments. $\mathbf{w}$ is defined as a function of $\boldsymbol{\zeta}$ and h.

To solve the population balance equation for the cells together with the mass balance equation for substances in a closed system the following initial and boundary conditions are required:

1. Initial condition at time $t = 0$ for n over Ω^s and Ω^z:

$$n = n^0(\mathbf{s}, \mathbf{z}) \quad \text{at} \quad t = 0 \tag{2.24}$$

2. Boundary condition at the boundary of spatial state space $\partial\Omega^s$ for n:

$$(n\mathbf{u}_L) \cdot \hat{\mathbf{s}} = 0 \quad \text{at} \quad \mathbf{s} \in \partial\Omega^s \tag{2.25}$$

where $\hat{\mathbf{s}}$ denotes the unit normal vector at the surface of $\partial\Omega^s$.

3. Boundary condition at the boundary of internal state space $\partial\Omega^z$ for n:

$$n\boldsymbol{\zeta} = \mathbf{0} \quad \text{at} \quad \mathbf{z} \in \partial\Omega^z \tag{2.26}$$

This boundary is called the "regularity" or "containment" condition. Generally speaking, this boundary condition simply states that no cells cross over the cell state boundary into impossible or meaningless states, such as negative concentrations or maximal concentrations based on feed information.

4. Initial condition at time $t = 0$ for $\mathbf{c}$ over Ω^s:

$$\mathbf{c} = \mathbf{c}^0(\mathbf{s}) \quad \text{at} \quad t = 0 \tag{2.27}$$

5. Boundary condition at the boundary of spatial state space $\partial\Omega^s$ for $\mathbf{c}$:

$$(\alpha_L \mathbf{c}\mathbf{u}_L) \cdot \hat{\mathbf{s}} = \mathbf{0} \quad \text{at} \quad \mathbf{s} \in \partial\Omega^s \tag{2.28}$$

Boundary conditions Equation (2.24) and Equation (2.28) on $\partial\Omega^s$ are normally automatically satisfied if the hydrodynamic parameters are obtained when the boundary conditions follow:

$$(\alpha_L \mathbf{u}_L) \cdot \hat{\mathbf{s}} = \mathbf{0} \quad \text{at} \quad \mathbf{s} \in \partial\Omega^s \tag{2.29}$$

which indicates that the carrier liquid cannot enter or escape the closed system. The initial condition Equation (2.27) of extracellular concentrations will be generally known from measurements.

In this section, the population balance equation and the mass balance equation for substances are introduced as continuum models. Even when decoupled from hydrodynamics under simplifying assumptions, solving this coupled system remains computationally intensive. To mitigate this challenge, researchers commonly focus on the dynamics of the averaged number density function, n_j, over a region within Compartment j, by assuming a "well-mixed" condition for n within the compartment. The details and application of the population balance equation for compartmental models are presented in the next section.

2.5.3 Population balance equations for compartmental models

The derivation of the population balance equation (PBE) for a compartment can be accomplished through the following steps:

Step 1: Integrate the population balance equation (Equation (2.18)) over the compartment with a fixed volume of space Ω_j^s:

$$\frac{\partial}{\partial t}\int_{\Omega_j^s} n\, d\mathbf{s} + \nabla^z \cdot \int_{\Omega_j^s} n\boldsymbol{\zeta}\, d\mathbf{s} + \int_{\Omega_j^s} \nabla^s \cdot (n\mathbf{u})\, d\mathbf{s} = \int_{\Omega_j^s} h\, d\mathbf{s} \tag{2.30}$$

Step 2: Apply the Gauss-Ostrogradsky divergence theorem to the convection transport term in spatial domain:

$$\frac{\partial}{\partial t}\int_{\Omega_j^s} n\, d\mathbf{s} + \nabla^z \cdot \int_{\Omega_j^s} n\boldsymbol{\zeta}\, d\mathbf{s} + \int_{\partial\Omega_j^s} \hat{\mathbf{s}} \cdot (n\mathbf{u})\, d\mathbf{s} = \int_{\Omega_j^s} h\, d\mathbf{s} \tag{2.31}$$

Step 3: Introduce $n_j(t, \mathbf{z})$ as the averaged number density function over Compartment j, such that $\int_{\Omega_j^s} n\, d\mathbf{s} = n_j v_j$. Combining this with $\boldsymbol{\zeta} = \boldsymbol{\zeta}_j(\mathbf{z}, \mathbf{c}_j)$ and $h = h_j(\mathbf{z}, n_j)$, we obtain:

$$v_j \frac{\partial n_j}{\partial t} + v_j \nabla^z \cdot (n_j \boldsymbol{\zeta}_j) + \int_{\partial\Omega_j^s} \hat{\mathbf{s}} \cdot (n_j \mathbf{u})\, d\mathbf{s} = v_j h_j \tag{2.32}$$

Step 4: Close the convection term in the spatial domain. For the compartmental model with a finite number of compartments ($\mathrm{n_{EC}}$), the term can be closed as:

$$\int_{\partial\Omega_j^s} \hat{\mathbf{s}} \cdot (n_j \mathbf{u})\, d\mathbf{s} = -\sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j'\to j} n_j + n_j \sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'} \tag{2.33}$$

which yields:

$$v_j \frac{\partial n_j}{\partial t} + v_j \nabla^z \cdot (n_j \boldsymbol{\zeta}_j) = v_j h_j + \sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j'\to j} n_{j'} - n_j \sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'} \tag{2.34}$$

Equation (2.34) represents the population balance equation for the compartmental model, which can be directly integrated with the compartmental mass balance equations, Equation (2.3) or Equation (2.4), to simulate a bioreactor. Simulating the system requires only the specification of the following initial and boundary conditions:

1. Initial condition at time $t = 0$ for n_j in Compartment j over Ω^z:

$$n_j = n_j^0(\mathbf{z}) \quad \text{at} \quad t = 0 \tag{2.35}$$

2. Boundary condition at the boundary of internal state space $\partial\Omega^z$ for n_j:

$$n_j \boldsymbol{\zeta} = \mathbf{0} \quad \text{at} \quad \mathbf{z} \in \partial\Omega^z \tag{2.36}$$

3. Initial condition at time $t = 0$ for $\mathbf{c}_j$ in Compartment j:

$$\mathbf{c}_j = \mathbf{c}_j^0 \quad \text{at} \quad t = 0 \tag{2.37}$$

2.5.4 Moments of the number density function

Hulburt and Kartz [53] claimed that the number density function, $n(t, \mathbf{s}, \mathbf{z})$, contains too much information for many engineering applications and presented an approximation system that tracks the evolution of moments, $m_{\boldsymbol{\alpha}}(t, \mathbf{s})$, of n instead of n itself. For example, intracellular distributions is mostly measured by flow cytometry, which does not provide a high level of resolution due to noise and sensitivity limitations. To define a moment, the number density function n is multiplied by a weighting factor defined as follows:

$$g = \prod_{k=1}^{\mathrm{n}_z} z_k^{\beta_k} \tag{2.38}$$

where β_k is an exponential weight defined for the internal state z_k.

This weighting factor is then integrated over the internal state space. Different moments are characterized by various combinations of z_k-orders, specified using the vector of exponents $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k, \ldots, \beta_{\mathrm{n}_z}]$. The moments of n with respect to $\boldsymbol{\beta}$ are, then, defined as:

$$m_{\boldsymbol{\beta}} = \int_{\Omega^z} \prod_{k=1}^{\mathrm{n}_z} z_k^{\beta_k} n \, d\mathbf{z}, \tag{2.39}$$

where the order of the moment $O = m_{\boldsymbol{\beta}}$ is $\sum_{k=1}^{\mathrm{n}_z} \beta_k$. The integration process removes the dependence of $m_{\boldsymbol{\beta}}(t, \mathbf{s})$ on $\mathbf{z}$. Similarly, the moments for a control volume (Compartment j), denoted by $m_{\boldsymbol{\beta},j}(t)$, can be obtained by applying the same procedure to $n_j(t, \mathbf{z})$.

The low-order moments are typically connected to several physically significant and measurable properties of the cells [53], such as mean and variance, which are useful for

validation of the model. For example, Equation (2.11), which calculates the total number of cells in a system at time t ($N(t)$), can be expressed as a function of the zeroth moment of n, where $\boldsymbol{\beta} = \mathbf{0} = [0, \cdots, 0]$:

$$N = \int_{\Omega^s} m_{\mathbf{0}} d\mathbf{s} = \int_{\Omega^s} \int_{\Omega^z} n d\mathbf{z} d\mathbf{s} \tag{2.40}$$

The first moment of n, which is related to the single-cell mass $x \in \mathbf{z}$, can be employed to ascertain the total mass of all cells, $X(t)$, in a system at time t, where $\boldsymbol{\beta} = [\beta_x = 1, 0, \cdots, 0]$:

$$X = \int_{\Omega^s} m_{[\beta_x=1,0,\cdots,0]} d\mathbf{s} = \int_{\Omega^s} \int_{\Omega^z} x n d\mathbf{z} d\mathbf{s} \tag{2.41}$$

Provide the internal concentration, $\omega_i \in \mathbf{z}$, of species i where

$$\omega_i \; [=] \; \frac{\text{Mass of Species } i}{\text{Mass of a cell}} \tag{2.42}$$

The second moment of n, associated with the single-cell mass $x \in \mathbf{z}$ and the concentration ω_i, can be utilized to determine the total mass, $W(t)$, of species i within all the cells in a system at time t, where $\boldsymbol{\beta} = [\beta_x = 1, \beta_{\omega_i} = 1, 0, \cdots, 0]$:

$$W = \int_{\Omega^s} m_{[\beta_x=1,\beta_{\omega_i}=1,0,\cdots,0]} d\mathbf{s} = \int_{\Omega^s} \int_{\Omega^z} \omega_i x n d\mathbf{z} d\mathbf{s} \tag{2.43}$$

2.5.5 Moment transport equations

The moment transport equations (MTEs), analogous to the population balance equation, tracks the dynamic evolutions of the moments $m_{\boldsymbol{\beta}}$. MTEs could be obtained by multiplying the weight factor,Equation (2.38), with the population balance equation, Equation (2.18), and then integrating over the internal cell space Ω^z:

$$\frac{\partial m_{\boldsymbol{\beta}}}{\partial t} + \int_{\Omega^z} g \boldsymbol{\nabla}^z \cdot \boldsymbol{\zeta} n \, d\mathbf{z} + \boldsymbol{\nabla}^s \cdot \mathbf{u} m_{\boldsymbol{\beta}} = \int_{\Omega^z} g h \, d\mathbf{z} \tag{2.44}$$

Using a divergence theorem and the containment condition given by Equation (2.26), the second term on the left-hand side of Equation (2.44) can be simplified as follows:

$$\begin{aligned} \int_{\Omega^z} g \boldsymbol{\nabla}^z \cdot \boldsymbol{\zeta} n \, d\mathbf{z} &= \cancelto{0}{\int_{\partial\Omega^z} \hat{\mathbf{z}} \cdot g n \boldsymbol{\zeta} \, d\mathbf{z}} - \int_{\Omega^z} (\boldsymbol{\nabla}^z g) \cdot \boldsymbol{\zeta} n \, d\mathbf{z} \\ &= - \int_{\Omega^z} (\boldsymbol{\nabla}^z g) \cdot \boldsymbol{\zeta} n \, d\mathbf{z} \end{aligned} \tag{2.45}$$

Substituting this term back into Equation (2.44) yields

$$\frac{\partial m_{\boldsymbol{\beta}}}{\partial t} + \boldsymbol{\nabla}^s \cdot \mathbf{u} m_{\boldsymbol{\beta}} = \int_{\Omega^z} \boldsymbol{\nabla}^z g \cdot \boldsymbol{\zeta} n + g h \, d\mathbf{z} \tag{2.46}$$

when used to model a system, MTEs for different moments replace the population balance equation. These equations are coupled with other balance equations, such as the mass balance equation for substances (Equation (2.22)).

Unlike the population balance equation, the initial condition for the moments is no longer a function of the internal states $\mathbf{z}$:

$$m_{\boldsymbol{\beta}} = m^0_{\boldsymbol{\beta}}(\mathbf{s}) \quad \text{at} \quad t = 0 \tag{2.47}$$

Moreover, there is no longer a boundary condition at the boundary of the internal state space $\partial\Omega^z$, as it is automatically included in the derivation of the MTEs. However, there remains a boundary condition at the boundary of the spatial state space $\partial\Omega^s$ for n:

$$(m_{\boldsymbol{\beta}} \mathbf{u}_L) \cdot \hat{\mathbf{s}} = 0 \quad \text{at} \quad \mathbf{s} \in \partial\Omega^s \tag{2.48}$$

which is automatically satisfied for the case that the hydrodynamic parameters reflect that the carrier liquid cannot enter or escape the closed system.

Since MTEs track only the dynamics of the moments, but typical population measurements are given in terms of mean and variance, an assumed distribution is required to recover these values from the moments. Further discussion on the choice of distribution is provided in Section 3.4.2.

2.5.6 Moment transport equations for compartmental models

The MTEs for compartmental models describe the dynamics of the moments, m_j, of the averaged number density function, n_j, within a "well-mixed" compartment j. The derivation follows a procedure similar to that used for obtaining Equation (2.46), but it is applied to the population balance equation for compartmental models, Equation (2.34). This results in:

$$\frac{dm_{j,\boldsymbol{\beta}}}{dt} = \int_{\Omega^z} \boldsymbol{\nabla}^z g \cdot \boldsymbol{\zeta}_j n_j + g h_j \, d\mathbf{z} + \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} m_{j',\boldsymbol{\beta}} - m_{j,\boldsymbol{\beta}} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'} \tag{2.49}$$

For a bioreactor, these MTEs are solved together with the compartmental mass balance equations for extracellular concentrations, Equation (2.3) or Equation (2.4). With the

boundary conditions in both the spatial and internal state spaces incorporated during the derivation, Equation (2.49) is transformed into an ordinary differential equation. To solve this equation, the following initial condition is required:

$$m_{j,\boldsymbol{\beta}} = m^0_{j,\boldsymbol{\beta}} \quad \text{at} \quad t = 0 \tag{2.50}$$

$$\mathbf{c}_j = \mathbf{c}^0_j \quad \text{at} \quad t = 0 \tag{2.51}$$

It is important to note that, when modeling a system, the MTEs for multiple moments, rather than a single moment, serve as a replacement for the population balance equation.

Chapter 3

Literature review

This chapter explores existing computational approaches for bioreactor modeling, emphasizing compartmentalization strategies and the integration of metabolic and population balance models.

It reviews numerical methods for solving dynamic flux balance analysis models (DFBAs) as a structured metabolic model to capture nonlinear microbial reactions, along with population balance models (PBMs), focusing on the method of moments for moment transport equations (MTEs) to track microbial heterogeneity. By evaluating current methodologies, this chapter identifies key limitations and research gaps, providing a foundation for the modeling advancements presented in later chapters.

3.1 Simulation of large-scale bioreactors[1]

To accurately capture the diverse responses of microorganisms in varying environments and stages of operation, the coupling of the metabolism with the hydrodynamics, must be considered. Table 3.1 provides an overview of the current state-of-the-art, focusing on the simulation of large-scale stirred bioreactors using various coupled hydrodynamic and metabolic models.

Modern methods apply structured metabolic models that are based on a mapping between the cell environment and the intra-cellular kinetic rates. This enables these models to capture the *metabolic state* that the cell is expected to adopt in response to its environmental conditions.

[1] This section is modified from [1].

Table 3.1: Overview of the current state-of-the-art in the use of metabolic models integrated with CFD or compartmental models. EE: Euler-Euler method, CM: Compartmental Model, PBE: Population Balance Equation

Year	Reference	Hydrodynamic Model	Metabolic Models
1996	Larsson et al. [5]	EE	Unstructured model
2001	Vrabel et al. [54]	Regular-grided CM [55]	Structured model [26]
2014	Morchain et al. [6]	EE	See Table 3.2
2015, 2017	Pigou et al. [18, 56]	Regular-grided CM [55]	See Table 3.2
2021	Nadal-Rey et al. [57]	Volume-adaptive CM	Structured model [58]
2024	Promma et al.	CFD-derived CM [59]	BSTMM derived from DFBA[20]

Two examples of recently developed structured metabolic models include those of Nadal-Rey et al. [57] and Xu et al. [26]. Nadal-Rey et al. [57] proposed a structured model for *Saccharomyces cerevisiae* based on the assumption of five metabolic regimes: oxidation, overflow, glucose starvation, oxygen limitation, and oxygen limitation and glucose starvation. The regime is selected based on the specific uptake rates of glucose and oxygen. Xu et al. [26] introduced a kinetic model framework for *E. coli* that considers the flux of glucose consumption through both oxidative and overflow metabolisms. In their model, the main utilization of glucose flux was for oxidative metabolism. When the oxygen flux surpasses the pre-established maximum respiration rate, the excess glucose flux is redirected to the overflow metabolism

To account for the segregated aspect of intracellular heterogeneity in microorganisms, several research studies have been conducted using the Euler-Lagrange method. However, the computational complexity of this method is very high due to the large number of particles/parcels (on the order of 10^5) that need to be tracked over time [7].

Instead, population balance equations (PBEs) are used to track the distribution of cells across different intracellular states, enabling the study of microbial population heterogeneity in bioreactors. Several studies have integrated PBEs with metabolic and compartmental models to improve representation. Table 3.2 summarizes key works in this area, including models that describe biological heterogeneity in well-mixed bioreactors, the interplay of physical and biological heterogeneities in large-scale systems, and the coupling of PBEs

with flux balance analysis to study cell size distribution under balanced growth conditions.

Table 3.2: Summary of studies on population balance models used to investigate microbial population heterogeneity.

Year	Reference	Details
2013	Morchain et al. [60]	Integration of a metabolic model with a population balance model to capture biological heterogeneity in growth rates within a "well-mixed" bioreactor.
2013	Fernandes et al. [61]	Coupling of a metabolic model and a population balance model to account for cell mass heterogeneity in a "well-mixed" bioreactor for yeast cultures.
2014	Morchain et al. [6]	Investigation of the interplay between physical and biological heterogeneities in large-scale bioreactors using an Euler-Euler approach, a population balance model, and a Monod-type growth model and yield fraction for substrate consumption rates.
2015	Pigou et al. [18]	Investigation of the interplay between physical and biological heterogeneities in large-scale bioreactors using a combination of compartmental models, population balance, and a structured metabolic model.
2017	Pigou et al. [56]	Evaluation of the efficiency of various moment methods for solving population balance equations that describe the dynamic adaptation of microbial populations in large-scale bioreactors.
2017	Quedeville et al. [62]	Development of a two-dimensional population balance model incorporating cell length and anabolic rate as key variables to better understand cellular growth regulation in response to nutrient availability.
2023	Busschaert et al. [63]	Integration of flux balance analysis with a population balance model to describe cell size distribution under balanced, quasi-steady state, growth conditions.

3.2 Compartmentalization[2]

3.2.1 Determination of compartment network

The process for determining the compartment network from velocity field information is referred to as *compartmentalization*. This involves segmenting the domain into subdomains, with mass fluxes between them used to determine compartment network connectivity and exchange flows. The majority of compartmental modeling work for bioreactors uses time-averaged flow-field information from short-timescale CFD simulations to identify the compartments using various *compartmentalization* methods. A common approach for stirred-tank units, which are typically axisymmetric, is to use compartmentalize based on the decomposition of the velocity field into radial and axial components, disregarding the azimuthal velocity [59, 57]. Other work includes:

Successive subdivision: Proposed by Wells et al. [64], this method involves progressively partitioning the reactor volume within CFD mesh cell boundaries until a predetermined number of compartments is obtained. The primary advantage of this approach is its simplicity, however, there is no guarantee that the cells within a specific compartment are contiguous.

Cell agglomeration: Proposed by Bezzo et al. [65, 66], this method involves construction of compartments by aggregating CFD cells. Compartments are defined by choosing a seed CFD mesh cell and combining neighboring cells until the maximum target variable gradient within the compartment falls above the specified tolerance. Additionally, Delafosse et al. [67] employed two distinct layer-by-layer agglomeration approaches, which were found to result in compartmental networks which underestimate of mixing time when compared to full CFD simulations.

Multiple target compartmentalization: Tajsoleiman et al. [59] introduced an automated method that tackled several challenges in compartmentalization, including (i) handling an unstructured mesh, (ii) addressing non-contiguous zones, and (iii) accommodating multiple target compartmentalizations.

The compartmentalization procedure of Tajsoleiman et al. [59], noteworthy due to the degree of CFD validation reported, involves the following steps:

[2]This section is modified from [1]

Zone Clustering: Based on user chosen target variables, p_i, the CFD elements are sorted and grouped. Elements that are within a range of target variable values, Δp_i are clustered into a zone where $(p_{i,max} - p_{i,min})$ within each zone is less than the specified Δp_i. As different parts of a reactor might have a similar range of values, an identified zone can be a multi-regional zone covering several regions within the reactor.

Zone Separation: To ensure that compartments will be contiguous, multi-regional zones obtained in the previous step are further divided. By examining the connectivity of the elements, it is possible to generate spatially contiguous compartments.

Compartment Overlap: Overlapping all compartments derived from different target variables that overlap following the preceding two steps.

Compartment Merging: Merging small compartments to further reduce computational complexity based on a minimal acceptable size of a compartment, V_{min}.

It is important to note that both the compartment overlapping and merging steps have the potential to generate discontiguous zones. Consequently, the zone separation procedure is employed once again to ensure the creation of compartments that are spatially contiguous. Furthermore, an additional merging step involved joining neighboring small compartments with those that have the largest adjacent area. This prevents the formation of compartments with irregular shape, such as dendritic structures or necking structures.

Individual compartments are traditionally approximated as well-mixed compartments (see Table 2 in ref. [17]), however, recent research has introduced higher-dimensional approximations including one-dimensional plug-flow compartments [68].

3.2.2 Determination of flow rates in multiphase compartmental models

Each chemical species is assumed to be transported between compartments with a constant flow velocity of the liquid (effective phase). The required information necessary for coupling the compartmental model with the DFBA-based species balances is represented in matrix-form, including the following:

Effective internal volume matrix $\mathbf{V}$ with elements composed of *effective* volume values $V_{j,j} = \alpha_j^L v_j$ where α_j^L and v_j are the volume fraction of liquid and the volume of Compartment j, respectively.

Effective volumetric flow rate matrix $\mathbf{Q}$ with elements representing the total *unidirectional* volumetric flows of the liquid between compartments.

In general, flows between any two adjacent compartments can be bidirectional where, for a given pair of compartments, i and j, there exist interfaces where flows from i to j and, at other interfaces, flows from j to i. The *effective* volumetric flow rates from Compartment i to Compartment j is determined by,

$$q_{i\to j} = \sum_{f\in\mathbf{I}_{i/j}} \delta_f \phi_f^L \tag{3.1}$$

where $\mathbf{I}_{i/j}$ is the set of the interfaces between Compartments i and j, and ϕ_f^L is the volumetric flux of the liquid phase.

The *direction corrector* δ_f is to specifically capture the flow from Compartment i to j and not in the reverse direction, and is calculated as follows,

$$\delta_f = \begin{cases} 1; & \text{for case 1} \\ 0; & \text{for case 2} \end{cases} \tag{3.2}$$

Case 1 describes flow from Compartment i to j and involves two different scenarios:

(i) either the interface has the surface vector, $\mathbf{s}_f$, pointing from Compartment i to Compartment j with a positive value of ϕ_f^L, or

(ii) the interface has $\mathbf{s}_f$ pointing from Compartment j to Compartment i with a negative value of ϕ_f^L.

Conversely, Case 2 describes the flow from Compartment j to Compartment i and also involves two different scenarios:

(i) the interface has $\mathbf{s}_f$ pointing from Compartment i to Compartment j with a negative value of ϕ_f^L, or

(ii) the interface has $\mathbf{s}_f$ pointing from Compartment j to Compartment i with a positive value of ϕ_f^L.

Based on the definition of the direction corrector provided here, it is straightforward to understand that if ϕ_f^L is not included in $q_{i\to j}$ then it will be a portion of $q_{j\to i}$.

To facilitate the matrix operations involved in solving the species conservation equation (Equation (2.46)) of the compartmental model, the volumetric flow rate information is represented in matrix-form, $\mathbf{Q}$. Off-diagonal elements represent flow rates between compartments, $Q_{i,j} = q_{i\to j}$ and diagonal elements represent the total flow rate leaving Compartment i (to other compartments) $Q_{i,i} = -\sum_j q_{i\to j}$.

3.3 Numerical method for dynamic flux balance analysis[3]

3.3.1 Multiparametric linear programming representation of dynamic flux balance analysis

When solving a linear programming problem like Equation (2.5), optimal solutions will involve constraints (Equation (2.5b)) that are either active or inactive. Thus, Equation (2.5) has different optimal solutions corresponding to different sets of active constraints, depending on *a priori* unknown parameters which, for example, could include local bioreactor conditions. Active constraints are expressed by equalities,

$$\mathbf{A}_A \mathbf{v}_j = \mathbf{F}_A \theta(\mathbf{c}_j) + \mathbf{g}_A \tag{3.3}$$

where if the matrix $\mathbf{A}_A$ is invertible, the system of equalities is solvable with a unique *optimal* solution $\mathbf{v}_j^*$,

$$\mathbf{v}_j^* = \mathbf{G}\theta(\mathbf{c}_j) + \mathbf{h} \tag{3.4}$$

where $\mathbf{G} = \mathbf{A}_A^{-1}\mathbf{F}_A$ and $\mathbf{h} = \mathbf{A}_A^{-1}\mathbf{g}_A$. Inactive constraints are expressed as inequalities $\mathbf{A}_I \mathbf{v}_j \leq \mathbf{F}_I \theta(\mathbf{c}_j) + \mathbf{g}_I$. The result of the substitution of Equation (3.4) into the inequality constraints defines a polyhedral region, Θ_i, of θ as given in Section 3.3.1.

$$\mathbf{H}\theta(\mathbf{c}_j) \quad \leq \quad \mathbf{p} \tag{3.5}$$

where $\mathbf{H} = \mathbf{A}_I \left(\mathbf{A}_A^{-1}\mathbf{F}_A\right) - \mathbf{F}_I$ and $\mathbf{p} = \mathbf{g}_I - \mathbf{A}_I\mathbf{A}_A^{-1}\mathbf{g}_A$.

Multiparametric programming (MP) [21] is an approach for solving optimization problems in a computationally efficient way by identifying these distinct polyhedral regions, known as *critical regions*, in the parameter space $\mathbb{R}^{n_\theta}$ within which sets of active and inactive constraints are invariant. Consequently, within these *critical regions* there is a

[3]This section is modified from [1].

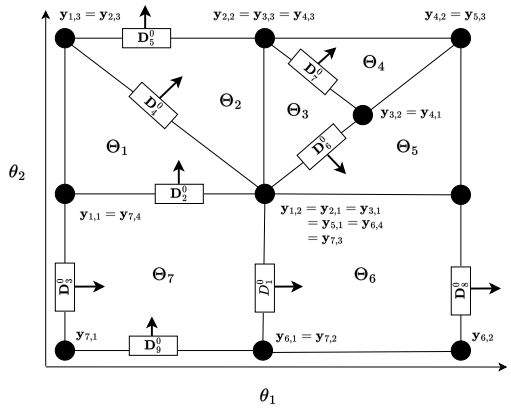


Figure 3.1: Illustration depicting critical region maps in $\mathbb{R}^2$ featuring labelled geometric elements of significance, including critical regions as Θ_i, hyperplanes as $\mathbf{D}_j^0$, and vertices as $\mathbf{y}_{i,k}$

unique optimal solution, e.g., $\mathbf{v}_j^*$ (Equation (3.4)). The term *optimal law* is subsequently utilized to denote this optimum solution, as it imparts a sense of a rule that all points within a critical region must adhere to. This reformulation of a optimization problem may be schematized as a *critical region map*, which defines distinct critical regions within the parameter space, as shown in Figure 3.1.

3.3.2 Mathematical description of a critical region map

Formally, MP optimization partitions the space $\mathbb{R}^{\mathrm{n}_\theta}$ into a set $\Theta = \bigcup_{i=1}^{\mathrm{n}_\theta} \Theta_i \subseteq \mathbb{R}^{\mathrm{n}_\theta}$ of non-overlapping polyhedra defined as:

$$\begin{aligned} &\Theta_i = \{\theta \in \mathbb{R}^{\mathrm{n}_\theta} : \mathbf{H}_i\theta \leq \mathbf{p}_i\} \\ &\text{with} \quad \Theta_i \cap \Theta_j = \emptyset \quad \text{for} \quad \forall i \neq j \end{aligned} \tag{3.6}$$

where $\mathrm{n}_{D,i}$ is the number of inequality constraints which define Θ_i, $\mathbf{H}_i \in \mathbb{R}^{\mathrm{n}_{D,i}} \times \mathbb{R}^{\mathrm{n}_\theta}$, and $\mathbf{p}_i \in \mathbb{R}^{\mathrm{n}_{D,i}}$.

Each inequality constraint defines a hyperplane and thus $\mathrm{n}_{D,i}$ is equivalent to the number of hyperplanes. Each row j of $\mathbf{H}_i$ and $\mathbf{p}_i$, i.e., $\mathbf{h}_j \in \mathbb{R}^{\mathrm{n}_\theta}$ and $p_j \in \mathbb{R}^{\mathrm{n}_{D,i}}$, defines a hyperplane $\mathbf{D}_j^0 = \{\theta \in \mathbb{R}^{\mathrm{n}_\theta} : d_j(\theta) = 0\}$ where

$$d_j(\theta) = \mathbf{h}_j^T\theta + p_j \tag{3.7}$$

It should be noted that different critical regions Θ_i may share hyperplanes with others. Therefore, all critical regions are typically represented using a set of hyperplanes drawn from the n_D unique hyperplanes that define Θ.

A polyhedron corresponding to a critical region is described by the intersection of the half-spaces defined on each side of $\mathbf{D}_j^0$, i.e., $\mathbf{D}_j^+ = \{\theta \in \mathbb{R}^{n_\theta} : d_j(\theta) > 0\}$ and $\mathbf{D}_j^- = \{\theta \in \mathbb{R}^{n_\theta} : d_j(\theta) < 0\}$. By selecting a set of n_θ linearly independent hyperplanes, expressed as $\mathbf{H}_i\theta = \mathbf{p}_i$, the points of intersection $\mathbf{W}_i$ may be identified which arise from the intersection of these n_θ hyperplanes. Among these intersection points lie the vertices of Θ_i, defined as $\mathbf{Y}_i = \{\theta \in \mathbf{W}_i \subseteq \mathbb{R}^{n_\theta} : \mathbf{H}_i\theta = \mathbf{p}_i\}$.

It is important to note that certain critical regions may have the same optimal law, which occurs when the original optimization problem has multiple optimal solutions (e.g., Shen et al. [69]). Consequently, the number of optimal laws may be less than the number of critical regions. For each Θ_i, the optimal law is represented as:

$$\mathbf{v}_k^* = \mathbf{G}_k\theta + \mathbf{h}_k \tag{3.8}$$

where n_K denotes the number of optimal laws and $n_K \leq n_\theta$.

3.3.3 Parameter point location

The multi-parametric optimization formulation of the dynamic flux balance analysis models (DFBAs) produces a set of critical regions in parameter space, each with an optimal law that is used to calculate the optimal metabolic growth. The challenge with the method is to identify *which* critical region a given state vector is in. The naive brute-force approach would involve checking each critical region to see if a given state vector, a point in state space, is inside it. This approach is inefficient, and several more efficient point location algorithms may be used to decrease computational complexity.

One class of point location methods use tree data structures which branch based on which side of a hyperplane a given point lies, i.e. whether $\theta(\mathbf{c}_j)$ lies in $\mathbf{D}_j^+$ or $\mathbf{D}_j^-$. The classical data structure for point location problems is a *binary search tree* (BST), introduced by Tøndel et al. [70]. It employs the unique hyperplanes $\mathbf{D}_j^0$ of the critical regions to efficiently locate $\theta(\mathbf{c}_j)$ using a single evaluation per hyperplane. This data structure enables effective classification and benefits from the logarithmic relationship between the tree height, h, which relates to the number of critical regions within a branch of the tree, and the total number of critical regions according to $h \approx \lceil \log_2(n_\Theta) \rceil$ for a perfectly balanced tree.

The BST data structure associated with a point location problem contains internal and leaf nodes. Each node is denoted by N_b with $b = 1 \rightarrow n_N$ where n_N is the total number of nodes in the tree [70]. The information stored at each internal and leaf node of the tree is:

Set of hyperplane indices $\mathbf{J}_b$: The set holds hyperplane indices along with $\pm$ identifiers to signify the corresponding half-space association, e.g., $\mathbf{J}_b = \{1^-, 3^+, 7^-\}$. This set defines a polyhedron, denoted as $\mathbf{P}(\mathbf{J}_b) = \mathbf{D}_1^- \cap \mathbf{D}_3^+ \cap \mathbf{D}_7^-$, where the polyhedron is the intersection of the associated half-spaces, each defined by a hyperplane in the set.

Set of critical region indices Γ_b: The set holds critical region indexes that completely or partially reside within $\mathbf{P}(\mathbf{J}_b)$, denoted as $\Gamma_b(\mathbf{J}_b) = \{i : \Theta_i \cap \mathbf{P}(\mathbf{J}_b) \text{ is full-dimensional}\}$.

Additionally, internal nodes contain a *hyperplane index* j_b that denotes the choice of $\mathbf{D}_j^0$ for partitioning Θ in the course of tree construction. On the other hand, leaf nodes store an *optimal law index* k_b associated with Θ_i, where $i \in \Gamma_b$. A method to construct a BST is shown in Section 3.3.4.

Once the BST is created, the point location algorithm is evaluated for a given query point $\theta(\mathbf{c}_j)$, starting at the root node ($b = 1$). The value $d_{j=j_b}(\theta(\mathbf{c}_j))$ is computed to determine the corresponding path, either the $+$ path ($d_{j=j_b}(\theta(\mathbf{c}_j)) \geq 0$) or the $-$ path ($d_{j=j_b}(\theta(\mathbf{c}_j)) < 0$). This step is repeated for each subsequent internal node until a leaf node is encountered, for which the value k_b is stored. Subsequently, the metabolic flux vector is computed using the equation,

$$\mathbf{v}_j^* = \mathbf{G}_{k_b}\theta(\mathbf{c}_j) + \mathbf{h}_{k_b} \tag{3.9}$$

3.3.4 Construction of binary search tree

Constructing a BST based on the MP formulation of a FBA model first involves creation of a root node N_0 with an empty hyperplane index set $\mathbf{J}_0 = \emptyset$ and critical region index set containing all regions $\Gamma_0 = \{1, ..., n_\Theta\}$. A hyperplane $\mathbf{D}_j^0$ is chosen, i.e., an index $j_0 = j$, which is associated with the node data, which is subsequently utilized to classify all Θ_i regions and determine their placement in either $\mathbf{D}_j^+$ or $\mathbf{D}_j^-$ half-space. If a region Θ_i is found to be partitioned by the hyperplane $\mathbf{D}_j^0$, this region is determined to have sub-regions in both half-spaces.

Tøndel et al. [70] determined the position of a region Θ_i relative to a hyperplane by examining the closest and farthest points $\theta \in \Theta_i$ regarding Hyperplane j. This determination was made by solving two linear programming problems,

$\min_{\theta \in \Theta_i} d_j(\theta)$

$\min_{\theta \in \Theta_i} -d_j(\theta)$

To reduce computational complexity in evaluation of whether or not a region Θ_i is divided by Hyperplane j, the following condition is used [71]:

$$\mathbf{H}_i \theta \leq \mathbf{p}_i \quad \text{and} \quad \mathbf{h}_j^T \theta = p_j \tag{3.10}$$

Evaluation of this condition is equivalent to testing the feasibility of the corresponding linear programming problem.

After partitioning, child nodes are created for the critical regions within (i) the $\mathbf{D}_j^+$ half-space and (ii) the $\mathbf{D}_j^-$ half-space. The associated information, $\mathbf{J}_b$ and Γ_b, is updated for these child nodes to enable further partitioning. A path from the parent node to each of its child nodes is established, denoted as $+$ or $-$ for the child nodes of $\mathbf{D}_j^+$ and $\mathbf{D}_j^-$, respectively. If there is more than one optimal law shared between the critical regions of the child nodes, those are classified as internal nodes, else they are considered leaf nodes. The partitioning process is repeated for internal child nodes, with hyperplane selection limited to within the child space. This iterative process continues until all leaf child nodes are determined. For a system that includes many nodes, achieving a balanced tree is desirable, with further details found in references [70, 71].

3.4 Numerical methods for solving the population balance equation

The complexity of solving the population balance equation (PBE) depends significantly on the mathematical forms of the rate of change of the internal state, $\boldsymbol{\zeta}_j$, and the net rate of particle generation, h. For instance, if h is defined as in Equation (2.19), the equation becomes a partial functional differential equation (PFDE), introducing dependencies on past values of n_j. This transformation captures memory or time-delay effects, making the model suitable for spatially distributed systems with such dynamics.

In general, PBE often takes the form of a PFDE, particularly when it incorporates non-local state interactions such as aggregation or breakage, or when history-dependent processes are involved. These interactions create coupling across the state space, adding to the complexity of the solution. However, if simpler expressions are chosen for those terms, the equation may reduce to a partial differential equation (PDE). For example, if

offspring inherit all the considered states of their parent, leading to the form of h described in Equation (2.21), the functional dependencies are eliminated.

These factors profoundly influence the choice of solution methods and the associated computational effort, as more complex formulations require advanced numerical techniques and greater computational resources.

This section outlines two primary categories of methods for solving PBE: the bin (or class or sectional) method and the method of moments. In addition to these two approaches, several other techniques are also available. Since PBEs are partial differential equations, methods commonly used in computational fluid dynamics can be applied to solve them. These include the finite difference method [72], finite volume method [73], finite element method [74, 75], and spectral methods [76]. Additionally, Monte Carlo methods, although currently too computationally expensive for practical coupling with computational fluid dynamics (CFD), can be employed. Ramkrishna [77] provides a detailed discussion on the connection between Monte Carlo simulations and population balance modeling. For further details on the methods for PBEs, the standard textbook by Ramkrishna [52] and reviews by Omar et al. [78] and Shiea et al. [79] are recommended.

To simplify the discussion in this section, the further derivations for each method will start from PBE and the moment balance equation (MBE) for the compartmental model(Equation (2.34) and Equation (2.49)), where the spatial space Ω^z has already been discretized.

3.4.1 Bin method

The bin method involves discretizing the internal state space, Ω^z, into intervals (bins, classes, or sections), transforming the PBE into a set of macroscopic balance equations within the state space. This method has been widely applied to systems governed by univariate PBEs ($\mathrm{n}_z = 1$).

As mentioned in Section 2.5.1, Ω^z is divided into n_b bins, each represented as $\Omega^z_\gamma \subset \Omega^z$ for $\gamma \in \{1, 2, \ldots, \mathrm{n}_b\}$. A pivotal point, $\boldsymbol{\xi}_\gamma = [\xi_{\gamma,1}, \ldots, \xi_{\gamma,\mathrm{n}_z}]^\top$, is assigned as the representative internal state for Bin γ. The coordinates of the pivotal point can either be the central point or one of the corner points of the bin. The population balance equation for the total number of particles, $N_{j,\gamma}$, for Bin γ in Compartment j is derived by integrating Equation (2.34)

over Ω_γ^z and applying the definition of $N_{j,\gamma}$ from Equation (2.15). This yields:

$$\frac{\partial N_{j,\gamma}}{\partial t} = -v_j \int_{\Omega_\gamma^z} \nabla^z \cdot (n_j \boldsymbol{\zeta}_j)\, d\mathbf{z} + v_j \int_{\Omega_\gamma^z} h_j\, d\mathbf{z} + \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} N_{j',\gamma} - N_{j,\gamma} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'}. \tag{3.11}$$

The integrals in Equation (3.11) remain unclosed because they depend on the distribution n_j, which cannot generally be expressed in terms of $N_{j,\gamma}$. To solve the equations, n_j within Bin γ is approximated as:

$$n_j = n_{j,\gamma} v_\gamma^z \prod_{k=1}^{\mathrm{n}_z} \delta(z_k - \xi_{\gamma,k}) \tag{3.12}$$

This expression, involving Dirac delta functions, $\delta(z_k - \xi_{\gamma,k})$, forces all particles within Ω_γ^z to concentrate at a pivotal point, $\boldsymbol{\xi}$, inside the bin.

Substituting Equation (3.12) into the first integral on the right-hand side of Equation (3.11) yields:

$$\begin{aligned} -v_j \int_{\Omega_\gamma^z} \nabla^z \cdot (n_j \boldsymbol{\zeta}_j)\, d\mathbf{z} &= -v_j \int_{\Omega_\gamma^z} \nabla^z \cdot (n_{j,\gamma} \prod_{k=1}^{\mathrm{n}_z} \delta(z_k - \xi_{\gamma,k}) \boldsymbol{\zeta}_j)\, d\mathbf{z} \\ &= -\nabla^z \cdot v_j v_\gamma^z n_{j,\gamma} \boldsymbol{\zeta}_{j,\gamma} \\ &= -\nabla^z \cdot N_{j,\gamma} \boldsymbol{\zeta}_{j,\gamma} \end{aligned} \tag{3.13}$$

The term $\boldsymbol{\zeta}_{j,\gamma}(\boldsymbol{\xi}_\gamma, \mathbf{c}_j) = [\zeta_{j,1,\gamma}, \cdots, \zeta_{j,\mathrm{n}_k,\gamma}]^\top$ is the rate of change of the internal state, evaluating at $\mathbf{z} = \boldsymbol{\xi}_\gamma$.

The same derivation can be applied to the second term on the right-hand side of Equation (3.11) to express it as a function of $N_{j,\gamma}$. For instance, if h_j is expressed as $h_j = h'_j n_j$, where $h'_j(\mathbf{z}, \mathbf{c}_j)$ is an arbitrary function of $\mathbf{z}$ and $\mathbf{c}_j$, it can be shown that:

$$\begin{aligned} v_j \int_{\Omega_\gamma^z} h_j\, d\mathbf{z} &= v_j \int_{\Omega_\gamma^z} h'_j n_{j,\gamma} v_\gamma^z \prod_{k=1}^{\mathrm{n}_z} \delta(z_k - \xi_{\gamma,k})\, d\mathbf{z} \\ &= v_j v_\gamma^z n_{j,\gamma} h'_{j,\gamma} \\ &= N_{j,\gamma} h'_{j,\gamma} \end{aligned} \tag{3.14}$$

Substituting Equation (3.13) and Equation (3.14) back into Equation (3.11) yields:

$$\frac{\partial N_{j,\gamma}}{\partial t} = -\nabla^z \cdot N_{j,\gamma} \boldsymbol{\zeta}_{j,\gamma} + N_{j,\gamma} h'_{j,\gamma} + \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} N_{j',\gamma} - N_{j,\gamma} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'}. \tag{3.15}$$

One approach to solving Equation (3.15) is to approximate the divergence term. Morchain et al. [60] discretized the divergence term in Equation (3.15) using a first-order upwind scheme, accounting for the two possible directions of transport for each coordinate z_k.

$$\nabla^z \cdot N_{j,\gamma}\boldsymbol{\zeta}_{j,\gamma} = \sum_{k=1}^{\mathrm{n}_z} \frac{\partial}{\partial z_k}(N_{j,\gamma}\zeta_{j,k,\gamma}) \tag{3.16}$$

where

$$\frac{\partial}{\partial z_k}\left(N_{j,\gamma}\zeta_{j,k,\gamma}\right) \approx \begin{cases} \dfrac{N_{j,\gamma}\zeta_{j,k,\gamma} - N_{j,\gamma_-^k}\zeta_{j,k,\gamma_-^k}}{\Delta z_{k,\gamma}}, & \text{if } \zeta_{j,k,\gamma} > 0, \\ \dfrac{N_{j,\gamma_+^k}\zeta_{j,k,\gamma_+^k} - N_{j,\gamma}\zeta_{j,k,\gamma}}{\Delta z_{k,\gamma}}, & \text{if } \zeta_{j,k,\gamma} \leq 0. \end{cases} \tag{3.17}$$

In this formulation, γ represents the bin index in the state space Ω^z but does not inherently encode directional information along z_k. To clarify the directional dependence, the subscripts γ_-^k and γ_+^k are introduced to denote the variables in the bins one step downward and one step upward, respectively, from bin γ along coordinate z_k. Here, $\Delta z_{k,\gamma}$ represents the length of Bin γ along coordinate z_k.

By applying the bin method, Equation (3.11) takes the form of an ordinary differential equation (ODE). It is important to note that although the method is straightforward and effective for stabilizing numerical solutions, it carries a notable drawback regarding accuracy. The scheme introduces artificial infinite diffusion, which smears sharp gradients in the solution. This artificial diffusion can obscure the true behavior of the system, especially when high accuracy is required to capture intricate details such as boundary layers or sharp fronts. To minimize diffusion, the grid requires refinement, which leads to an increase in computational costs.

Lastly, it should be noted that the bin method presents several challenges. First, solving hundreds of ODEs for many bins can lead to high computational costs, especially for systems with large spatial domains or when extending to multivariate population balance equations (PBEs) [80, 81]. For instance, in a system discretized into n_{EC} compartments, where n_j is defined over an internal state space of n_z dimensions and further discretized into n_b bins per dimension, the total number of coupled ODEs to be solved becomes $\mathrm{n}_{EC} \times \mathrm{n}_b^{\mathrm{n}_z}$. Additionally, implementing this method in computational fluid dynamics (CFD) is often difficult due to the significant computational burden and its incompatibility with existing solvers [82, 83]. Furthermore, this method may encounter accuracy issues related to the arbitrary selection of bin intervals, making it difficult to capture fine-scale heterogeneity in certain systems.

3.4.2 Method of moments

The method of moments refers to a group of techniques used to solve the moment transport equations (MTEs) as elaborating in Sections 2.5.4 to 2.5.6.

MTEs, e.g., Equation (2.49), are generally not in closed form, except in a few simple cases. One reason for this issue is that transport equations formulated for a set of moments may include terms that depend on higher-order moments from the integration term.

Consider a toy case of univariate MTEs ($\mathrm{n}_z = 1$), where $\boldsymbol{\zeta}_j$ is defined as:

$$\boldsymbol{\zeta}_j = a(b - z) \tag{3.18}$$

and h_j is expressed as:

$$h_j = czn_j \tag{3.19}$$

Here, a, b, and c are constants. The MTE for this toy case is expressed as:

$$\begin{aligned} \frac{dm_{j,[\beta]}}{dt} &= \int_{\Omega^z} \beta z^{\beta-1} a(b-z) n_j + cz^{\beta+1} n_j \, dz + \mathrm{A} \\ &= ab\beta \int_{\Omega^z} z^{\beta-1} n_j dz - a\beta \int_{\Omega^z} z^{\beta} n_j dz + c \int_{\Omega^z} z^{\beta+1} n_j dz + \mathrm{A} \\ &= ab\beta m_{j,[\beta-1]} - a\beta m_{j,[\beta]} + c m_{j,[\beta+1]} + \mathrm{A} \end{aligned} \tag{3.20}$$

where A is the advection term, expressed as:

$$A = \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} m_{j',[\beta]} - m_{j,[\beta]} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'} \tag{3.21}$$

Equation (3.20) can be expanded to the following set of moment equations:

$$\frac{dm_{j,[0]}}{dt} = cm_{j,[1]} + \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} m_{j',[0]} - m_{j,[0]} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'} \tag{3.22a}$$

$$\frac{dm_{j,[1]}}{dt} = abm_{j,[0]} - am_{j,[1]} + cm_{j,[2]} + \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j' \to j} m_{j',[1]} - m_{j,[1]} \sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'} \tag{3.22b}$$

$$\frac{dm_{j,[2]}}{dt} = 2abm_{j,[1]} - 2am_{j,[2]} + cm_{j,[3]} + \sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j'\to j} m_{j',[2]} - m_{j,[2]} \sum_{\substack{j'=1 \\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'} \tag{3.22c}$$
$$\dots$$

It can be seen from the toy model that the MTE of moment β is a function of the higher-order moment $\beta+1$. Adding additional transport equations for these higher-order moments does not resolve the issue, as these new equations only introduce a dependence on even higher-order moments. However, if n_j is known, the higher-order moments can be easily computed. In general, the knowledge of n_j is also needed to calculate $\boldsymbol{\zeta}$ and h in the space of internal coordinates.

This issue is referred to as the method of moment closure problem. Several approaches have been developed to close MTEs, including interpolative closures [84], reconstruction of the number density function using an assumed functional form [85, 86, 87], and quadrature-based moment methods (QBMOMs) [88, 89, 90].

In the case of multivariate population balance models (PBMs), QBMOMs are typically the preferred methods [79]. A fundamental contribution to this field was made by McGraw [88], who introduced the quadrature method of moments (QMOM) as a solution to univariate MTEs in growth problems. It allows one to capture the essential behavior of a distribution through its key moments, thus keeping the computational requirements manageable.

QMOM approximates the moments of n_j using an n_p-point Gaussian quadrature, where n is expressed as a sum of delta functions located at the quadrature abscissas (nodes):

$$n_j = \sum_{p=1}^{\mathrm{n}_p} w_{j,p}\delta(z - \xi_{j,p}) \tag{3.23}$$

where $\xi_{j,p}(t)$ and $w_{j,p}(t)$ are the abscissa and weight of the node p in Compartment j. $\delta(z - \xi_{j,p})$ denotes the Dirac delta function. The moment of order β of the approximated n_j can be expressed as follows:

$$m_{j,[\beta]} \approx \sum_{p=1}^{\mathrm{n}_p} w_{j,p}\xi_{j,p}^{\beta} \tag{3.24}$$

The given relationship suggests that by knowing the first $2\mathrm{n}_p$ moments, it is possible to identify the n_p weights and n_p abscissas required for the quadrature approximation in Equation (3.23). These nodes, $\xi_{j,p}$, and their corresponding weights, $w_{j,p}$, are extracted

from the set of moments in Equation (3.24) using moment inversion algorithms such as the product-difference (PD) algorithm [91] or the Wheeler algorithm [92]. McGraw [88] demonstrated that Gaussian quadrature could effectively approximate the growth term without imposing restrictions on its mathematical form. Additionally, QMOM was shown to accurately predict the time evolution of moments m_0 to m_5, with a relative error of less than 0.2% when compared to the analytical solution.

One challenge in solving QMOM is ensuring that the computed moments remain realizable—that is, they must lie within the moment space throughout the simulation. Moments cannot take arbitrary values; they must adhere to the constraints imposed by the non-negative number density function. When the inversion algorithm is applied to a set of realizable moments, it produces weights and abscissas that satisfy the properties of Gaussian quadrature, with the weights always being positive. In numerical schemes, a set of weights and abscissas is considered to represent realizable moments if the weights are positive and the abscissas fall within the physical range. The issue of realizability primarily arises because the numerical methods used to solve the discretized moment transport equations differ from the exact equations [79].

It should be emphasized that QMOM is particularly attractive for applications where capturing the full details of the distribution is less critical than ensuring certain properties are conserved, such as total cell number or mass, and identifying primary trends. However, when experimental data provide moment-related measurements (e.g., mean and variance), direct comparison with simulation results is expected. This necessitates the reconstruction of the number density function, n_j.

The reconstruction process is crucial not only for experimental validation, but also for coupling a structured metabolic model with a PBM to capture the nonlinear responses of microbes. In such metabolic models, discontinuities in the adaptation rate, ζ_j, growth rate, h_j, and interfacial rate, w, could be expected. Consequently, solving such a system requires reconstructing n_j at every time step to accurately determine the right-hand side of the MTEs (Equation (2.49)) and the MBEs (Equation (2.3)). Section 6.3.1 will provide a numerical example illustrating the error that arises when QMOM is applied to solve reaction rates without reconstructing n_j.

In the literature, QMOM with reconstruction of n_j is known as the extended quadrature method of moments (EQMOM) [93]. This approach couples QMOM with the kernel density element method (KDEM). In this approach, n_j is reconstructed as a weighted sum of well-known distributions (e.g., normal, log-normal, beta, and gamma distributions).

3.4.3 Truncated multivariate normal distribution and its moments

In general, the normal distribution is often a suitable approximation for natural variations due to the central limit theorem, which states that the sum of independent random effects tends to follow a normal distribution. However, many physical properties, such as size or concentration, are restricted by certain unphysical bounds (e.g., they cannot take negative values or exceed physical limits). These considerations often make the truncated normal distribution a natural candidate for modeling the underlying behavior of n_j. In the study by Pigou et al. [56], the EQMOM was used to reconstruct n_j as a weighted sum of up to three normal distributions. This, however, raises the question of how rate calculations were handled when one of the distributions approached the unphysical bounds described earlier. Hence, in this study, the truncated normal distribution is adopted as the functional form for the average number density function, n_j, to correct the distribution near the physical bounds. To test this concept, a truncated normal distribution is assumed for n_j, ensuring that n_j follows a unimodal distribution.

The functional form for the average number density function, n_j, is expressed as:

$$n_j = \frac{N_j}{v_j}\tau_j = c_{N,j}\tau_j \tag{3.25}$$

where

$\mathbf{c}_{N,j}$: Cell number concentration, representing the scalar magnitude of cell population per unit volume.

τ_j : Truncated normal distribution given by a probability density function representing the statistical structure of the population.

Since τ_j is a probability density function, it follows that: $m_{\mathbf{0},j} = \int_{\Omega^z} n_j d\mathbf{z} = \mathbf{c}_{N,j}$.

Then, the truncated normal distribution, $\tau(z) \equiv \tau(z; z_{\min}, z_{\max}, \mu, \sigma^2)$, for a univariate system ($\mathrm{n}_z = 1$), evaluated at the mean μ and variance σ^2, within the physical bounds $[z_{\min}, z_{\max}]$, is expressed as:

$$\tau(z) = \begin{cases} \dfrac{\phi(z)}{L} & z_{min} \leq z \leq z_{max} \\ 0 & \text{otherwise} \end{cases} \tag{3.26}$$

where $\phi(z) \equiv \phi(z; \mu, \sigma^2)$, is the probability density function of a univariate normal distribution, expressed as:

$$\phi = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{3.27}$$

and $L(z_{min}, z_{max}; \mu, \sigma^2)$ is expressed as:

$$L = \int_{z_{\min}}^{z_{\max}} \phi dz \tag{3.28}$$

The moments of τ (for $n_z = 1$) are given as follows:

$$y_\beta = \int_{z_{min}}^{z_{max}} z^\beta \tau dz \tag{3.29}$$

Kan and Robotti [94] developed a computationally efficient recursive algorithm to determine the moments of the truncated multivariate normal distribution y_β as follows:

$$y_\beta = \frac{\int_{z_{min}}^{z_{max}} z^\beta \phi dz}{L} = \frac{F_\beta}{L} \tag{3.30}$$

where the integral term:

$$F_\beta = \int_{z_{min}}^{z_{max}} z^\beta \phi dz \tag{3.31}$$

can be integrated by parts leading to the following:

$$F_0 = L \tag{3.32a}$$

$$F_{\beta+1} = \mu F_\beta + \beta\sigma^2 F_{\beta-1} + \sigma \left(z_{min}^\beta \phi(z_{min}) - z_{max}^\beta \phi(z_{max}) \right) \qquad (\beta \geq 0) \tag{3.32b}$$

Note that when $\beta = 0$, the term involving the undefined component F_{-1} vanishes, and F_1 can be computed solely based on F_0. Equation (3.32) provides a recurrence relation for determining higher-order moments, serving as a computationally efficient alternative for calculating the integral product in Equation (3.31). However, integration is still necessary to determine L.

Chapter 4

Development of a CM for Large-Scale Bioreactor Simulations[1]

4.1 Introduction

Computational fluid dynamics (CFD) has been widely used to simulate multiphase flow in bioreactors, providing detailed hydrodynamic information. However, direct CFD simulations are computationally expensive, particularly when applied to large-scale systems over long timescales. To address this limitation, compartmental modeling has emerged as a viable alternative, reducing computational demands while preserving essential hydrodynamic and transport phenomena.

The study in this chapter focuses on the development of a flow-informed compartmental model (CM) for a 30 m^3 bioreactor equipped with four baffles and a four-blade in-phase Rushton turbine impeller. The compartmentalization approach is based on time-averaged CFD simulations that resolve the velocity fields using the Euler-Euler model and the volume of fluid method. By leveraging time-averaged velocity data, the bioreactor is compartmentalized into well-mixed compartments, approximating local flow dynamics. The model aims to serve as a framework to capture spatial heterogeneity while offering computational efficiency suitable for large-scale bioprocess simulations.

The CM developed here was integrated with the binary search tree metabolic model (BST-MM) to investigate a large-scale bioreactor, as elaborated in Chapter 5.

[1]This chapter is modified from [1].

4.2 Methodology

To develop an accurate compartmental model for large-scale bioreactors, a systematic methodology was implemented, combining CFD simulations with flow-informed compartmentalization techniques. The methodology is divided into two primary steps: (1) the setup and execution of multiphase CFD simulations to obtain velocity fields, and (2) the subsequent use of these velocity fields to develop a compartmental model.

4.2.1 Computational fluid dynamics simulation setup

The steady velocity fields used for compartmentalization were obtained by solving the hydrodynamics using an integration technique that combines the Euler-Euler two-fluid and the volume of fluid model. In OpenFOAM version 10, this integration technique was accessible within the solver named `multiphaseEulerFOAM`. OpenFOAM utilizes the cell-centred finite volume method as a discretization scheme for the flow domain, enabling the straightforward acquisition of grid data for compartmentalization purposes [95].

The reactor shape and operation were adapted from Lasson et al. [5]. Figure 4.1 illustrates a diagram of the reactor geometry and the computational mesh employed in the study. The reactor was equipped with four in-phase Rushton turbines and four baffles, 90 degrees apart. The mesh was divided into two distinct domain types: the tank domain, encompassing the baffles, shaft, and sparger, and the rotating domain, comprising the impellers and a portion of the shaft. The mesh utilized in the simulations was composed of 1,360,852 hexagonal elements. Owing to the computational expense of CFD, a search for a mesh-independent solution was not conducted. Hence, it is logical to expect that various mesh-dependent solutions will yield slightly different compartmental network geometries and flow rate values. On the other hand, the obtained compartmental model is merely an approximation of the system and its suitability is assessed based on its ability to fit and predict experiments [96].

The bioreactor was operated in fed-batch mode; however, intermittent medium withdrawal was employed to maintain the total liquid volume at 22 m^3.

The following reciprocal polynomial equation of state for liquids was used to calculate the water density [41]:

$$\begin{aligned}\frac{1}{\rho} =& 0.001278 - \left[2.1055 \times 10^{-6}\right] T + \left[3.9689 \times 10^{-9}\right] T^2 \\ & - \left[4.3772 \times 10^{-13}\right] P + \left[2.0225 \times 10^{-16}\right] PT\end{aligned} \tag{4.1}$$

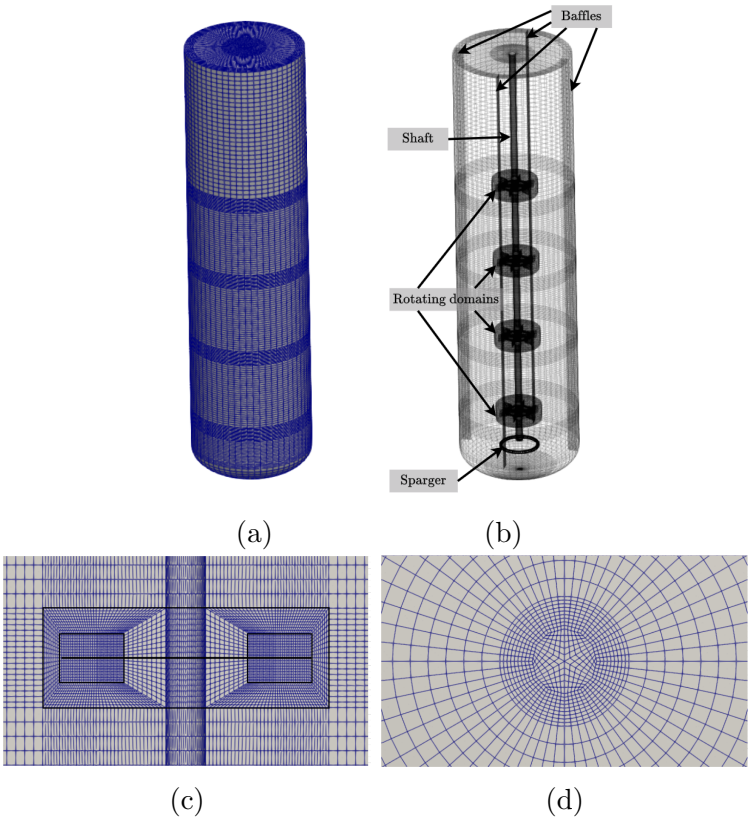


Figure 4.1: Illustration depicting the mesh for the reactor geometry and the computational mesh employed; presenting (a) the external surface mesh, (b) the components inside the tank, (c) mesh within the rotational domain, and (d) mesh at the bottom of the tank.

Table 4.1: Physical properties of water and air used for the simulation.

Properties	Water	Air
Molecular mass $(kg/kmol)$	18	28.9
Specific heat capacity $(J/kg \cdot K)$	4195*	1007*
Dynamic viscosity $(Pa \cdot s)$	3.645×10^{-4}	1.84×10^{-5}
Prandtl number	2.289	0.7
Surface tension (N/m)	0.07	

* The given value for water corresponds to the constant-volume heat capacity, whereas the value for air corresponds to the constant-pressure heat capacity.

where T is the temperature and P is the absolute pressure. The density of air was determined by using the ideal gas law. Table 4.1 lists the other physical properties of water and air used.

The air flow rate was $2.22 \times 10^{-2} kg/s$. For simplicity, instead of creating a mesh representation of the sparger, air bubbles were introduced into the system by incorporating them as a source term located at the bottom of the reactor. The isothermal dispersed-phase particle diameter model was used to model the bubble sizes, d_b, which were varied to account for changes in pressure [41]:

$$d_b = d_{b,ref} \sqrt[3]{\frac{P_{ref}}{P}} \tag{4.2}$$

where $d_{b,ref}$ is the bubble diameter at the reference pressure, P_{ref}. A value of 3 mm was used for $d_{b,ref}$ for P_{ref} of 10^5 Pa.

The interphase momentum transfer term was modelled as having contributions from drag, lift, and virtual mass. Drag was represented using the Ishii-Zuber correlation [97]. In situations where there was no clearly identifiable dispersed phase, the segregated drag model developed by Marschall [98] was utilized. Lift was represented using Tomiyama models designed for gas-liquid systems [99]. The virtual mass force was calculated using a constant coefficient virtual mass model. The turbulent heat transfer coefficient between the bubble surface and the surrounding fluid was determined using the Ranz-Marshall correlation [100]. The rotating domain was tackled by using the method of multiple reference frames [37]. The rotational speed of the impeller was set to 2.22 s^{-1}. The resulting

Reynolds number of the system, calculated with a fixed kinematic viscosity at 25 degrees Celsius, is approximately 1.2×10^6.

The simulation was run for a period of 70 seconds to confirm that the bubble reached the surface and the velocity fields attained their steady states. After this period, a transient averaging technique was initiated, and the simulations were continued for an additional 30 seconds. The averages of the results obtained during this period were subsequently utilized for the compartmentalization procedure.

The estimation of the mixing time consisted of solving a transport equation for a passive scalar (numerical tracer experiment) through the application of the OpenFOAM solver `scalarTransportFOAM`, utilizing the time-averaged velocity field. For simplicity, this estimation was conducted within a single-phase system employing a diffusion coefficient of 0.0005 m^2/s to aid in solution convergence.

4.2.2 Compartmentalization

The hydrodynamic compartmentalization method used in this work has been detailed in Section 3.2, where the largest adjacent area criterion is chosen as the primary criterion for merging small compartments. In cases where multiple merging options exist, a secondary criterion is applied based on the closest average values of the target variables.

The hydrodynamic compartmental model that was obtained was based on liquid-phase velocity profile information only. Thus, the *effective* domain, which pertains to the constant liquid portion of each compartment, was subsequently employed within the compartmental model. Vapor-phase transport was approximated through accounting for the oxygen transfer rate $g_{O_2,j}$ within Compartment j:

$$g_{O_2,j} = k_j^L a_j \left[c_{O_2^*,j} - c_{O_2,j} \right] \tag{4.3}$$

where $k_j^L a_j$ represents the volumetric mass transfer coefficient, $C_{O_2^*,j}$ is the saturation oxygen concentration, and $C_{O_2,j}$ denotes the dissolved oxygen concentration. Vapor-phase hydrodynamics were accounted for through both $k_j^L a_j$ and $c_{O_2,j}$ values, which are volume-averaged values across all CFD mesh elements defining Compartment j. To calculate the element-wise values of k^L, a, and $c_{O_2^*}$, the correlations described in the reference provided by Nadal-Ray et al. [57] were applied.

4.3 Results and discussion

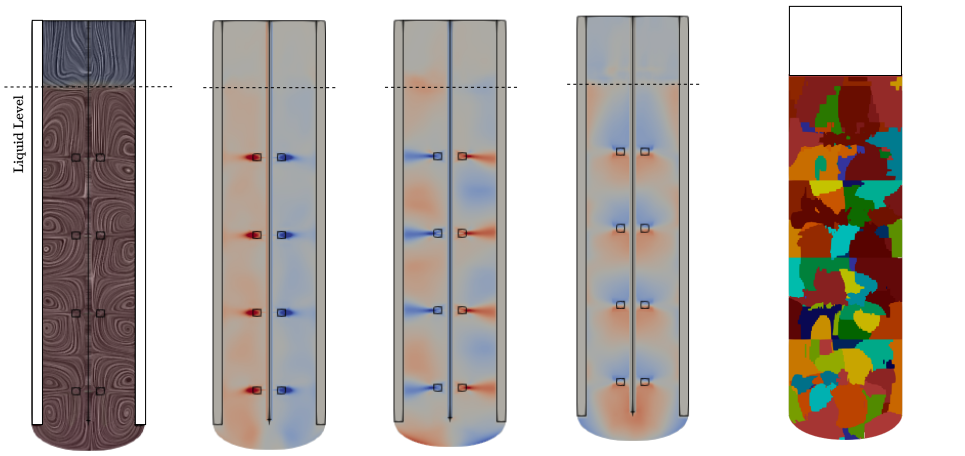
4.3.1 Multiphase computational fluid dynamics simulation

A multiphase CFD simulation of the large-scale bioreactor was performed using the dispersed multiphase two-fluid (Euler-Euler) model [95], described in Section 4.2.1. Impeller motion was captured using the MRF method.

The time-average velocity profile was obtained after developed flow conditions were identified and is shown for liquid phase is shown in Figure 4.2a. The line integral convolution (LIC) method [101, 102] for visualizing three-dimensional velocity fields is used and is colored by the local liquid volume fraction. Unsteady flow effects and turbulence are pervasively present in bioreactors using Rushton-based impellers. Thus, a time-averaged developed-flow velocity profile was used to be compatible with compartmental modeling. While past research has identified frequency-specific fluctuations of velocity [103, 104] and oscillation of partition planes between impellers [105], these effects are neglected in order to identify a single compartmentalization and are approximated through the introduction of explicit flow rate fluctuations in Section 5.3.3.

Simulations predict recirculatory flows located adjacent to each Rushton turbine, which transition from circulation loops between impellers. The presence of converging and diverging recirculatory flows elucidates the enhanced mixing effects resulting from the use of multiple impellers in a stirred tank. Figures 4.2b to 4.2d show individual velocity components of the liquid phase flow, which may be combined in order to inform a specific compartmentalization method.

In order to validate the multiphase hydrodynamics predicted by the CFD simulation, mixing time was estimated through the solution of purely convective single-phase transport of an inert tracer using the time-averaged liquid velocity field. To quantitatively assess the degree of mixing and identify a characteristic time of mixing, the coefficient of mixing (CoM) [106, 107, 108] was calculated and the mixing time was based on the point at which it surpasses 95% of the steady-state value, or in other words, when the CoM falls below 0.05. The computed mixing time was found to be 226.2s which compares to past simulations estimating mixing time of 165s [54] and 250s [18], respectively. Given the significant variation in past predictions, along with variations in impeller speeds and gas flow rates used, these results support some degree of validation of the presented multiphase CFD simulation results to inform compartmentalization.



(a) LIC (b) u_x (c) u_y (d) u_z (e) Compartments

Figure 4.2: 30m^3 bioreactor (with 22m^3 of working volume): (a) Line integral convolution representation of velocity fields colored by the liquid volume fraction; Averaged velocity components (b) u_x, (c) u_y, (d) u_z; (e) Compartments from the flow-informed compartmental model.

4.3.2 Flow-informed compartmentalization

A CM was developed based on the time-averaged velocity components of the CFD results, shown in Figures 4.2b to 4.2d. Compartmentalization parameter thresholds of $\Delta p_i = 0.5\text{m}\,\text{s}^{-1}$ and a minimum compartment size $V_{min} = 0.1\text{m}^3$ were used. Following the compartmentalization process, the resulting compartmental network includes a total of 101 compartments which are approximated as well-mixed (CSTR). The mixing time predicted by the compartmental model was found to be 60.05s (95% mixing time), significantly lower than what is predicted by CFD.

In order to determine the cause of poor mixing time prediction, a sample set of compartments were superimposed on the time-averaged velocity profile, shown in Figure 4.3. This analysis reveals that the compartmentalization method adopted from [59] produces compartments that, when assumed to be well-mixed, create direct connections between distinct recirculation zones within the bioreactor, which is referred to as a "short-cutting" effect. It can be observed in Figure 4.3 that two recirculation zones form between adjacent impellers, separated by a unidirectional flow region where the vertical flux is nearly zero. This intermediate region serves as a hydrodynamic barrier, limiting substrate advection between the recirculation zones. However, when compartments are defined within this

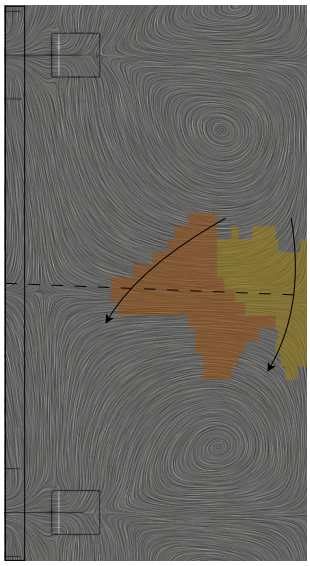


Figure 4.3: The presence of "short-cutting" paths (indicated by the arrows) introduced by compartments (highlighted in orange and yellow) between local recirculatory zones. The dashed line indicates a plane where the flow flux perpendicular to the plane is nearly zero.

region, they fail to account for the barrier and instead create a shortcut between zones, thereby underestimating the true hydrodynamic resistance to inter-zone mixing. As a result, the model overestimates convective mass transfer and inaccurately predicts mixing times.

In order to address this issue, a user-guided modification was made based on insight gained from CFD simulations. The bioreactor was manually divided into four distinct zones corresponding to sub-volumes in which recirculation was present, referring to Figure 4.2a. The division was imposed on the compartmentalization method by creating horizontal partition planes midway between each pair of impellers, such that merging of compartments across these planes is not permitted. This prevented the creation of compartments which overlap circulatory zones, reducing the short-cutting associated with the well-mixed approximation of each compartment in the resulting model. This approach generates a total of 108 compartments, shown in Figure 4.2e, and results in a compartment model which predicts 95% mixing time of 189.34s, significantly closer to CFD results of 226.2s. While the change results in a slightly high number of compartments, 108 instead of 101, the vast majority of the improvement in mixing time prediction is attributed to the prevention of the "short-cutting" and not to the presence of an extra 7 compartments.

For future studies, it is important to address the limitation of manual refinements. To this end, Vasile et al. [109] proposed higher-dimensional approximations, such as one-dimensional plug-flow compartments, to better capture flow behavior without user intervention. Morover, while the present study used only residence time data for validation, more comprehensive datasets—such as tracer concentration profiles—could further enhance validation efforts, as demonstrated in previous studies [110, 68].

4.4 Conclusion

A compartmental model for large-scale aerated bioreactors was developed as a hydrodynamic representation and validated against residence time data obtained from multiphase CFD simulations. The flow-informed compartmentalization approach, based on time-averaged velocity profiles, revealed unphysical"short-circuiting" across mixing zones that led to inaccurate mixing time predictions. To address this issue, a user-guided modification was introduced to restrict artificial shortcuts between compartments, resulting in a refined model comprising 108 compartments. This adjustment significantly improved the accuracy of predicted mixing times, bringing them closer to the CFD results while preserving computational efficiency.

These findings underscore the importance of integrating CFD-based hydrodynamic insights into compartmental modeling to better capture spatial flow structures in large-scale bioreactors. The proposed methodology provides a robust framework for enhancing model fidelity, thereby supporting more accurate process optimization and scale-up strategies. Future work should focus on further coupling hydrodynamic models with microbial kinetics to improve predictive capabilities in complex bioprocess environments.

Chapter 5

Development of a BSTMM to Simulate a Coupled FBA-CM [1]

5.1 Introduction

Large-scale bioreactors exhibit spatial variations in nutrient concentrations, metabolic byproducts, and microbial population distributions. These variations pose significant challenges in accurately modeling microbial metabolism, particularly when using simplified Monod-type models with fixed yield coefficients. To better capture metabolic heterogeneity, structured (conditional) metabolic models have been introduced, where metabolic pathways are activated or suppressed based on intracellular and extracellular conditions. However, these models require numerous parameters, increasing computational costs and the risk of overfitting.

A key drawback of structured metabolic models is the high parameter count, which can lead to overfitting experimental data, ultimately reducing predictive accuracy. An alternative approach is dynamic flux balance analysis modeling (DFBA) [20], which has gained increasing attention [111, 112, 113, 69, 114] due to its ability to model microbial metabolism with fewer parameters. DFBA models operate under the assumption that cells evolve to optimize resource utilization for specific objectives, such as maximizing growth rate or optimizing co-factor production, subject to constraints. This approach is formulated as a linear programming (LP) problem, where the constraint coefficients depend on space- and time-varying state variables (e.g., glucose concentration), leading to different active

[1] This chapter is modified from [1].

sets of constraints that define metabolic flux distributions. While DFBA models reduce parameter requirements, they significantly increase computational complexity due to the need for solving an LP problem at each simulation step.

The goal of this work is to develop a computationally efficient modeling framework for large-scale bioreactors by coupling DFBA models with hydrodynamic compartmental models. To address the computational burden of solving LP problems, multiparametric linear programming [21] is employed to reformulate the LP problem as a point-location problem. This technique, commonly used in explicit model predictive control [115], allows the pre-computation of optimal metabolic solutions, which are stored in a binary search tree (BST) for efficient retrieval during simulations.

By integrating the BST metabolic model (BSTMM) with a CM this study provides an accurate and computationally efficient tool for investigating metabolic heterogeneity in large-scale bioreactors. The proposed method is validated against experimental data, demonstrating its capability to capture spatial and temporal variations in metabolite concentrations while maintaining computational efficiency. The results offer valuable insights into the interplay between bioreactor hydrodynamics and microbial metabolism, providing a foundation for optimizing bioprocess performance.

5.2 Methodology

An overview of the formulation of the coupled DFBA/compartmental model is shown in Figure 5.1, illustrating the coupling of the compartmental model, obtained in Section 4.3.2, with the BST metabolic model. The construction of the BST representation of the DFBA model is described in Section 3.3, along with the approach to determine the relative position of the critical regions with respect to a hyperplane.

In order to solve the parameter point location problem using the BSTMM, the relative position of a critical region in relation to a hyperplane is determined through the parameter space relationship of the hyperplane and the vertices of the region. Considering $\mathbf{y}_{i,k} \in \mathbf{Y}_i$ as a vertex k within a set of vertices $\mathbf{Y}_i$ for a region Θ_i, the half-space containing $\mathbf{y}_{i,k}$ can be inferred from the value of $d_j(\mathbf{y}_{i,k})$ (Equation (3.7)). The vertices $\mathbf{Y}_i$ of a partitioned Θ_i must be present in both half-spaces created by a hyperplane that divides the region, as shown in Figure 5.2b. Correspondingly, the signs of the values of $d_j(\mathbf{y}_{i,k})$ for all vertices must be different. Following this, the classification strategy was developed, as summarized by Algorithm 1. Using that algorithm, a hyperplane j contained within the parent space and the set of critical region indexes of the parent node Γ_p, the algorithm yields $\Gamma_c(j^+)$ and $\Gamma_c(j^-)$, which represent the sets of critical region indices for the two child nodes.

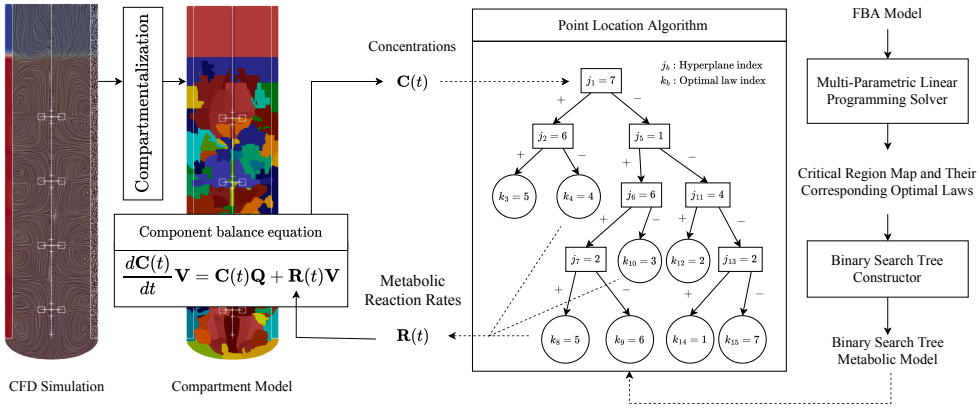


Figure 5.1: A flowchart outlining the process of integrating the compartmental model with the binary search tree metabolic model, with the species balance serving as a central junction connecting both models

Algorithm 1 Classification strategy based on region vertices

Input: Γ_p, $\mathbf{Y}_i$
Output: $\Gamma_c(j^+)$, $\Gamma_c(j^-)$
Initialize $\Gamma_c(j^+) \leftarrow \emptyset$ and $\Gamma_c(j^-) \leftarrow \emptyset$
for $i \in \Gamma_p$ **do**
  if $d_j(\mathbf{y}_{i,k}) \geq 0 \quad \forall \mathbf{y}_{i,k} \in \mathbf{Y}_i$ **then**
    $\Gamma_c(j^+) \leftarrow \Gamma_c(j^+) \cup \{i\}$ {Figure 5.2a}
  else if $d_j(\mathbf{y}_{i,k}) \leq 0 \quad \forall \mathbf{y}_{i,k} \in \mathbf{Y}_i$ **then**
    $\Gamma_c(j^-) \leftarrow \Gamma_c(j^-) \cup \{i\}$ {Figure 5.2c}
  else
    $\Gamma_c(j^-) \leftarrow \Gamma_c(j^-) \cup \{i\}$ {Figure 5.2b}
    $\Gamma_c(j^+) \leftarrow \Gamma_c(j^+) \cup \{i\}$
  end if
end for
return $\Gamma_c(j^+)$, $\Gamma_c(j^-)$

Note that subscripts p and c denote the parent and child, Γ_p and Γ_c, respectively, instead of using the placeholder b.

The simulations were performed on a personal laptop equipped with a 12$^{\text{th}}$ Gen Intel® Core™ i7-1260P processor (12 cores, 16 threads, base frequency 400 MHz, up to 4.7 GHz), 16 GB of RAM, a 512 GB SSD, an integrated Intel® Iris® Xe Graphics GPU, and running

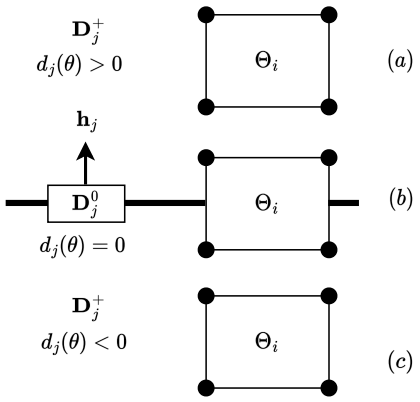


Figure 5.2: Potential positions of critical regions in relation to a hyperplane ($\mathbf{D}_j^0$); critical regions located (a) in $\mathbf{D}_j^+$, (b) in both $\mathbf{D}_j^+$ and $\mathbf{D}_j^-$, (c) in $\mathbf{D}_j^-$.

Ubuntu 64-bit.

5.3 Results and discussion

5.3.1 Integration of the binary search tree metabolic model into the compartmental model

The large-scale bioreactor [22] was operated in fed-batch mode based on expected levels of cell death and enzyme depletion. The microorganism used in the experimental study [22] was a wild-type Escherichia coli *E. coli* K-12 strain W3110 [26].

Since the DFBA model *E. coli* originally presented by Mahadevan et al. [20] was intended for batch operation, some modifications were made in order to include fed-batch operation. The original model was based on optimization of the growth rate of *E. coli* by allocating glucose and acetate among four distinct pathways. These pathways were chosen from a pool of 300 alternatives, considering factors such as biomass yield and the established physiology of *E. coli*. The model incorporates both aerobic and anaerobic pathways and includes four essential species: glucose (G), acetate (A), dissolved oxygen (O_2), and biomass (X). For simplicity and based on their similar density, these four species were considered in their dispersed forms (non-finite size particles) in the aqueous phase.

The four reaction pathways are represented as a stoichiometric coefficient matrix, $\mathbf{A}$,

given by Equation (5.1).

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_G^T \\ \mathbf{a}_{O_2}^T \\ \mathbf{a}_A^T \\ \mathbf{a}_X^T \end{bmatrix} = \begin{bmatrix} 0 & -9.46 & -9.84 & -19.23 \\ -35 & -12.92 & -12.73 & 0 \\ -39.43 & 0 & 1.24 & 12.12 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{5.1}$$

However, the original model was unable to accurately predict the long-term accumulation of glucose and acetate.

In this work, two key modifications were implemented to address these deficiencies:

1. Degradation of biomass was assumed as proposed by Hewitt et al.[30] according to a first-order degradation rate $g_{X,j} = -k_{dX} c_{X,j}$, where k_{dX} represents the specific degradation rate of biomass.

2. Uptake capacities for glucose and oxygen were based on functions of varying enzyme-to-biomass ratio, as proposed by Lin et al.[116].

These assumptions were adopted. An enzyme (E) was introduced to support the second assumption.

The components of the reaction rate matrix of the enzyme as functions of E were described by the following equation:

$$r_{E,j} = Y^{E/X} \mu_j c_{X,j} - \left[k_{dE} + k_{dX}\right] c_{E,j} \tag{5.2}$$

where $Y_{E/X}$ is the yield of enzyme from biomass and k_{dE} is a specific degradation rate of enzyme. Accordingly, specific uptake capacities for glucose were adjusted by incorporating a function of the enzyme-to-biomass ratio. This function becomes active when the ratio falls below a specific threshold r^*, as in Equation (5.3).

$$m_{i,j} = \begin{cases} \left[\dfrac{c_{E,j}}{c_{X,j} r^*}\right]^{\mathrm{n}_r}; & \dfrac{c_{E,j}}{c_{X,j}} < r^* \\ 1; & \text{otherwise} \end{cases} \tag{5.3}$$

The ratio is raised to the power of n_r to account for a nonlinear relationship with respect to the glucose consumption rate.

Following the above modifications, the *modified* DFBA model for *E. coli* used in this work is,

$$\max_{\mathbf{v}_j} \quad \mathbf{a}_X^T \mathbf{v}_j \tag{5.4a}$$

$$\text{s.t.} \quad -\mathbf{a}_{O_2}^T \mathbf{v}_j \leq K_{O_2,max} \tag{5.4b}$$

$$-\mathbf{a}_G^T \mathbf{v}_j \leq K_{G,max} \frac{c_{G,j}}{k_G + c_{G,j}} m_{G,j} = \theta_j^1(c_{G,j}, c_{E,j}, c_{X,j}) \tag{5.4c}$$

$$-\mathbf{a}_G^T \mathbf{v}_j \leq \frac{c_{G,j}}{c_{X,j}\Delta t} + \frac{F_{G,j}}{c_{X,j}V_{j,j}} = \theta_j^2(\mathbf{c}_G^T, c_{X,j}) \tag{5.4d}$$

$$-\mathbf{a}_{O_2}^T \mathbf{v}_j \leq \frac{c_{O_2,j}}{c_{X,j}\Delta t} + \frac{F_{O_2,j}}{c_{X,j}V_{j,j}} + \frac{G_{O_2,j}}{c_{X,j}} = \theta_j^3(\mathbf{c}_{O_2}^T, c_{X,j}) \tag{5.4e}$$

$$-\mathbf{a}_A^T \mathbf{v}_j \leq \frac{c_{A,j}}{c_{X,j}\Delta t} + \frac{F_{A,j}}{c_{X,j}V_{j,j}} = \theta_j^4(\mathbf{c}_A^T, c_{X,j}) \tag{5.4f}$$

with

$$F_{i,j} = \mathbf{c}_i^T \mathbf{q}_j + c_{i,j}^{in} q_{j,j}^{in} - c_{i,j} q_{j,j}^{out} \tag{5.5}$$

where $K_{O_2,max}$ is the maximum oxygen uptake rate, $K_{G,max}$ is the maximum glucose uptake rate, $\mathbf{c}_i^T$ is a row vector corresponding to species i in the concentration matrix $\mathbf{C}$, $\mathbf{q}_j$ is a column vector for compartment j in the volumetric flow rate matrix $\mathbf{Q}$, and the time interval Δt in the model is associated with numerical integration.

Equation (5.4) represents that the primary goal of the microorganisms is to maximize the biomass growth rate. Equation (5.4)b represents that the rate of oxygen consumption is constrained by a maximum uptake limit. Equation (5.4)c represents that the glucose consumption rate is restricted by an upper limit. All remaining constraints (Equation (5.4)d-f) are positivity constraints for concentrations.

The DFBA model was then reformulated into a BSTMM as described in Sections 3.3 and 5.2 where $n_\Theta = 18$ critical regions were found and an unbalanced binary search tree was used given the small number of critical regions. This resulted in a BST consisting of 67 nodes and a height of 7. Regarding the computational complexity of BST metabolic model, it required only approximately 0.040 ms to calculate the metabolic fluxes $\mathbf{v}_j$ for each time step and at each compartment. In contrast, the original DFBA, treated as a linear programming problem and solved using `scipy.optimize.linprog` (optimized SIMPLEX algorithm) [117], required 0.466 ms for the equivalent calculation.

While, in this work, the BSTMM is integrated into a hydrodynamic compartmental model, it is important to note that the reduction of computational complexity could also

benefit integration into full CFD simulation. For example, for a CFD simulation with $\approx 10^6$ degrees of freedom, similar to the one in Section 4.3.1, the computational time required to calculate the metabolic fluxes for a single time step using BST metabolic model is just 0.65 s compared to 430 s (7 minutes) required using `scipy.optimize.linprog`.

5.3.2 Validation and analysis of simulations of the large-scale bioreactor

Parameter estimation for the metabolic model was performed using experimental data for a small-scale (7 L) bioreactor for which the well-mixed approximation is justifiable, compared to large-scale bioreactors in which there are significant spatial variations of species. Metabolic model parameters were fitted to the experimental data of Xu et al. [22] by using a single-compartment CM. The system was treated as operating continuously, given that under the conditions described in ref. [22], the working volume is unchanged during operation, with the rates of media feed and intermittent supernatant removal equal.

The control methods were not reported in ref. [22] so, for simplicity, the PI controller was used to directly adjust the value of $k^L a$,

$$\frac{de}{dt} = \frac{dc_{O_2,j_{probe}}}{dt} \tag{5.6a}$$

$$\frac{d\int_0^t e\, d\tau}{dt} = c_{O_2,j_{probe}} - c_{O_2}^{set} \tag{5.6b}$$

$$u_{j_{probe}} = -K_c \left[e + \frac{1}{\tau_I} \int_0^t e\, d\tau \right] \tag{5.6c}$$

where $e = c_{O_2,j_{probe}} - c_{O_2}^{set}$ represents the error, $c_{O_2}^{set}$ is the oxygen concentration set-point, $c_{O_2,j_{probe}}$ is the measured oxygen concentration, K_c is the controller gain, τ_I is the integration time, and $u_{j_{probe}}$ is the control variable. The control variable affects $k^L a$ through the following relation,

$$k^L a = (k^L a)_0 + u_{j_{probe}} \tag{5.7}$$

where $(k^L a)_0$ is the reference value. To avoid non-physical responses from the control method, constraints are imposed where $k^L a \geq 0$. This limitation ensures that the manipulated variable remains within a positive range.

The metabolic model parameters were calibrated through an optimization process employing the Nelder-Mead method [118] to minimize the sum of squared errors between the data and the model predicted values. Simulation results for the best-fit metabolic

Table 5.1: Model parameters

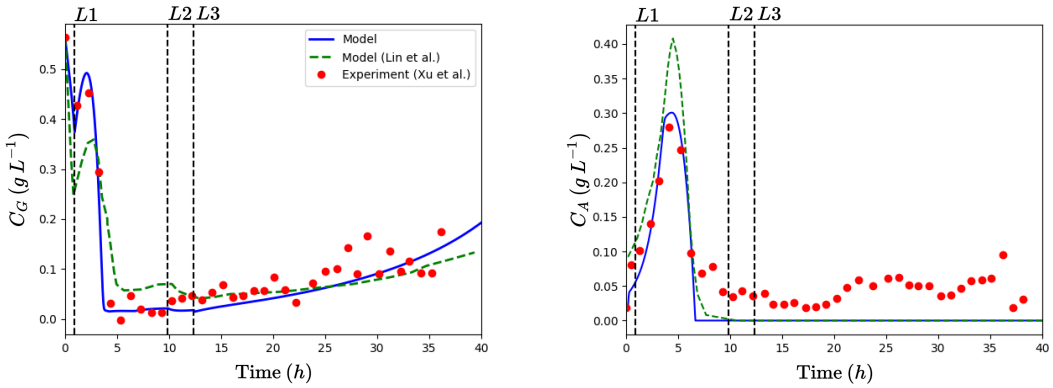
Parameters	Unit	Lab-scale bioreactor ($15L$)	Large-scale bioreactor ($30m^3$)
$K_{O_2,max}$	$mmol\ g^{-1}\ h^{-1}$	6	6
$K_{G,max}$	$mmol\ g^{-1}\ h^{-1}$	10	20
k_G	mM	0.2	0.2
k_{dX}	h^{-1}	0.04	0.04
$Y^{E/X}$	$g\ g^{-1}$	0.51	0.51
k_{dE}	h^{-1}	0.90	0.90
r^*	$g\ g^{-1}$	0.2	0.2
n_r		1.5	1.5
$(k^La)_0$	h^{-1}	12.50	See Section 4.2.2
$c_{O_2^*}$	mM	0.21	See Section 4.2.2
K_c	$h^{-1}\ mM^{-1}$	40	13
τ_I	h	100	200
$c_{O_2}^{set}$	mM	0.1	0.02

model parameters are shown in Figure 5.3 alongside past simulation results [116] and past experimental results [22]. Operating parameters are shown in Tables 5.1 and 5.3.

The single-compartment model accurately captures the gradual increase in glucose concentration during the fed-batch phase with a constant feed rate ($t > L2$). However, it is less accurate in capturing the acetate concentration, similarly to the simulations from ref. [116]. It is important to note that the metabolic model used in this work requires fewer kinetic parameters compared to that from ref. [116] and other past studies [18] of the same bioreactor system. After excluding the stoichiometric coefficients obtained from pathway analysis [119, 120, 121], the control parameters (as defined in Equation (5.6)), and the oxygen mass transfer parameters (as in Equation (4.3)) calculated from the CFD simulation, the total number of kinetic parameters needed for the presented model is 8. In contrast, the model from ref. [116] requires 16 kinetic parameters.

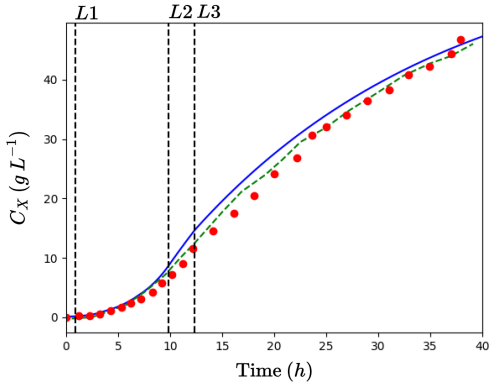
In order to compare the performance of the models quantitatively, the corrected Akaike information criterion (AIC_C) of the normalized data using the method described in Portet [122] was applied, shown in Table 5.4, supporting the improved performance of the presented due to its lower AIC_C.

Given validated estimates of the parameters for the metabolic model, simulations of a 30 m^3 bioreactor as described ref. [22] were performed. The compartmental network



(a) Glucose

(b) Acetate



(c) Biomass

Figure 5.3: Comparison of simulation concentration results with experimental data in a lab-scale bioreactor: (a) glucose, (b) acetate, (c) biomass. The vertical line $L1$ denotes the starting point of the fed-batch with an exponential feed rate. $L2$ and $L3$ denote the starting point of fed-batch with different constant feed rates (Table 5.2).

Table 5.2: Feed profiles of glucose in L/h and the associated parameters (Xu et al.[22]).

Time	Operational mode	Lab-scale bioreactor ($15L$)	Large-scale bioreactor ($30m^3$)
$0 \leq t < L1$	Batch	0	0
$L1 < t \leq L2$	Fed-batch with exponential feed rate	$0.0057\exp\{(0.27(t-L1))\}$	$15\exp\{(0.3(t-L1))\}$
$L2 < t \leq L3$	Fed-batch with constant feed rate	0.064	180
$L3 < t$	Fed-batch with constant feed rate	0.057	170
Parameters	Units		
$c_{G,feed}$	$g\ L^{-1}$	532	454
L_1	h	0.92	0.92
L_2	h	9.98	9.20
L_3	h	12.38	11.70

Table 5.3: Initial concentrations and control parameters

Initial conditions	Unit	Lab-scale bioreactor ($15L$)	Large-scale bioreactor ($30m^3$)
$c_{G,j}$	$g\ L^{-1}$	0.563	0.29
$c_{O_2,j}$	mM	0.5	$c_{O_2^*}$ (See Section 4.2.2)
$c_{A,j}$	$g\ L^{-1}$	0.018	0.050
$c_{X,j}$	$g\ L^{-1}$	0.1	0.1
$c_{E,j}$	$g\ L^{-1}$	0.065	0.065
e	mM	0.48	0.32
$\int_0^t e d\tau$	$mM\ h$	0	0

Table 5.4: Comparison of the corrected Akaike information criteria

Model	Lab-scale bioreactor ($15L$)	Large-scale bioreactor ($30m^3$)
This study	-428	-582
Other studies	-372^a	-207^b

a Lin et al. [116]
b Pigou et al. [18]

determined in Section 4.3.2 was used and resulted in a compartmental model with 108 compartments. This large-scale bioreactor was approximated as a continuous process, even though it was run as a fed-batch, since the total liquid volume was kept constant at approximately 22 m^3 through intermittent withdrawals.

The locations of the concentration probes and the feed location are listed in Table 5.5. The probe heights follow Xu et al. [22], and the positions (radius and height) of the feed are from Larsson et al. [5]. However, the angular positions of these components are not reported. Hence, the initial position was assumed along the x-axis. An investigation on the impact of the angular position on the results was performed and shown in Figure 5.10.

Table 5.5: Coordinates (X, Y, Z) of concentration probes and the feed location used in the simulation. The origin coordinate is at the center of the bottom hemisphere.

Components	X	Y	Z
Bottom probe	1.044	0	1.12
Middle probe	1.044	0	4.04
Top probe	1.044	0	6.48
Feed	0.056	0	6.5

A PI controller was used to regulate oxygen level (Equation (5.6)) with control error computed based on the oxygen level in the compartment corresponding to placement of the middle probe (described in ref. [22]). The control action at the compartment where the middle probe is located was utilized to determine the action for other compartments (u_j), with the assumption that the ratio between u_j and $k_j^L a_j$ remains constant throughout the tank. This assumption results in the following relation,

$$u_j = u_{j_{probe}} \frac{k_j^L a_j}{(k_{j_{probe}}^L a_{j_{probe}})_0} \tag{5.8}$$

and $k_j^L a_j$ was calculated as,

$$k_j^L a_j = (k_j^L a_j)_0 + u_j \tag{5.9}$$

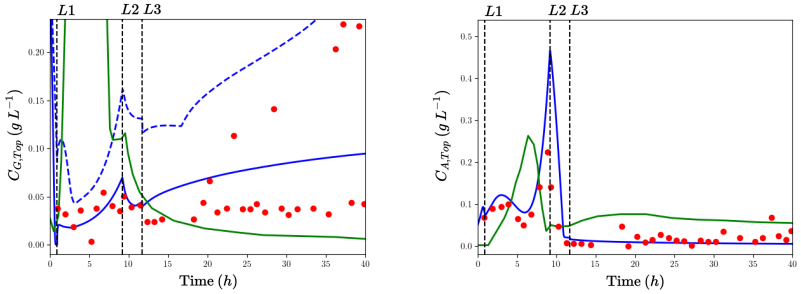
Simulation results of concentrations within the bioreactor are shown in Figures 5.4 and 5.6 alongside the experimental [22] and simulation data [18].

Using the same kinetic parameters fitted to lab-scale bioreactor data, simulations of the large-scale bioreactor show over-prediction of glucose concentration, shown by the blue-dashed line in Figures 5.4a, 5.4c and 5.4e. To accommodate for this, the maximum glucose uptake rate ($K_{G,max}$) was doubled, with no change to the other 7 kinetic parameters. The adjustment of this parameter is motivated by comparing the specific glucose uptake rates between lab-scale and large-scale bioreactors, shown in Figure 5.7. The uptake rates from the lab-scale bioreactor fall within the range of those from the large-scale bioreactor.

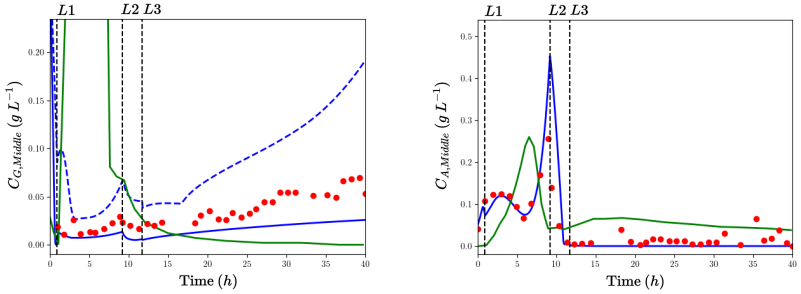
A possible explanation for the different maximum uptake rates is that in the large bioreactor, microorganisms were assumed to convect with the liquid phase throughout the entire tank on a short timescale. As a result, cells are rapidly exposed to varying extracellular concentrations along their trajectories. However, the adaptation of the cell population likely occurs on a much longer timescale compared to convection. Therefore, the maximum uptake rate exhibited by the population at any given location should reflect a distribution centered around their average maximum uptake rate, rather than being solely determined by the local extracellular condition. This hypothesis would result in a narrower distribution of maximum uptake rates in the large bioreactor, as shown in Figure 5.7, bringing it closer to that observed in the small-scale system. To accurately capture this effect, a population balance model (PBM) is required, which will be discussed in the next chapter.

The metabolic model, using a glucose uptake rate of 20 $\mathrm{mmol\,g^{-1}\,h^{-1}}$, was then used to re-simulate, with results shown by the blue-solid line in Figures 5.4a, 5.4c and 5.4e, and was found to better predict the trend of glucose concentration. As for the prediction by other reported models, it was observed that during both the batch and fed-batch phases with an exponential feed ($0 < t < L2$), the concentration decreases from the initial level. In contrast, Pigou et al. [18] and Vrábel et al. [54] reported an overestimation of glucose concentration during similar operations. This concentration overshoot observed in their models might be due to glucose being transported directly from the feed location to the measurement probes, without sufficiently accounting for cell consumption, implying that their compartmentalization was less accurate as compared to the one used in this work.

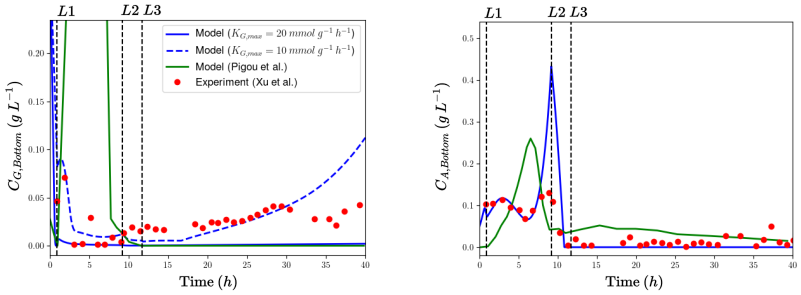
During the fed-batch phase with a constant feed ($t > L2$), the simulations predict glucose gradients. However, it tends to overestimate the concentration in the upper regions



(a) Glucose (6.48m from bottom) (b) Acetate (6.48m from bottom)



(c) Glucose (4.04m from bottom) (d) Acetate (4.04m from bottom)



(e) Glucose (1.12m from bottom) (f) Acetate (1.12 m from bottom)

Figure 5.4: Comparison of simulation concentration results with experimental data in a large-scale bioreactor: (a, c, e) refer to glucose, while (b, d, f) refer to acetate, studied at three distinct locations: (a, b) the top probe (6.48 m from the bottom of the tank), (c, d) the middle probe (4.04 m from the bottom of the tank), and (e, f) the bottom probe (1.12m from the bottom of the tank). The vertical line $L1$ denotes the starting point of the fed-batch with an exponential feed rate. $L2$ and $L3$ denote the starting point of fed-batch with different constant feed rates (Table 5.2).

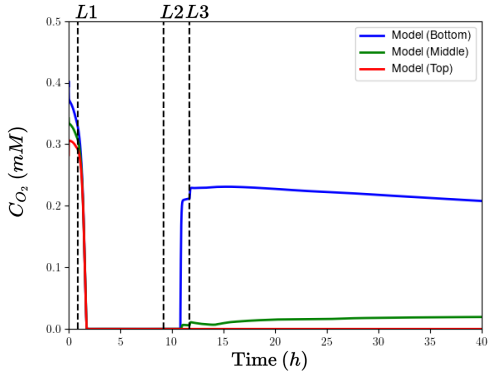


Figure 5.5: Oxygen concentration simulation results in a large-scale bioreactor

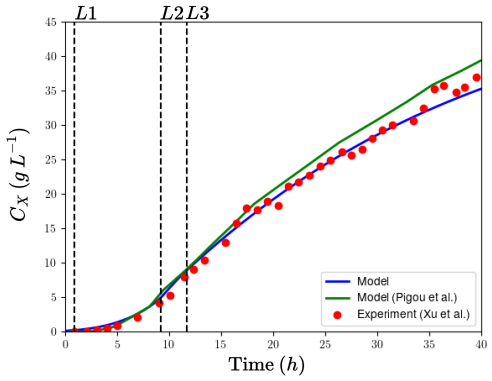


Figure 5.6: Comparison of simulation concentration results of biomass with experimental data at the top section in a large-scale bioreactor

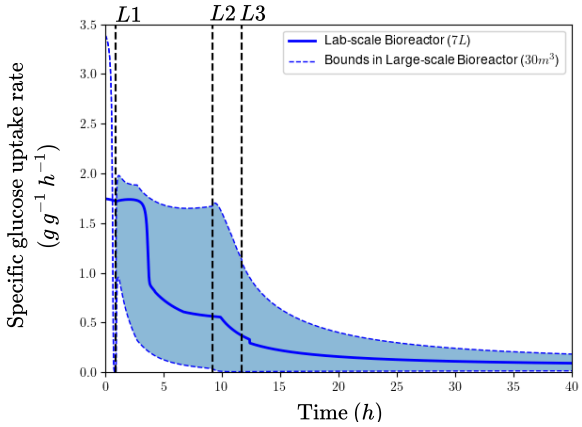


Figure 5.7: Comparison of the specific glucose uptake rate in laboratory-scale and large-scale bioreactors. The single rate, denoted by the solid line, is reported for a well-mixed lab-scale bioreactor, while rates at different locations in the large-scale bioreactor are reported as a range bounded by the dashed lines. The dynamics in the upper bounds are mostly due to dynamics in glucose concentration at the top part of the tank, where most of the glucose was fed at different rates (Table 5.2) and consumed throughout the course of operation. This result indicates that microorganisms in the lab-scale bioreactor exhibit an approximately average glucose consumption rate of those rates occurring in the large-scale bioreactor.

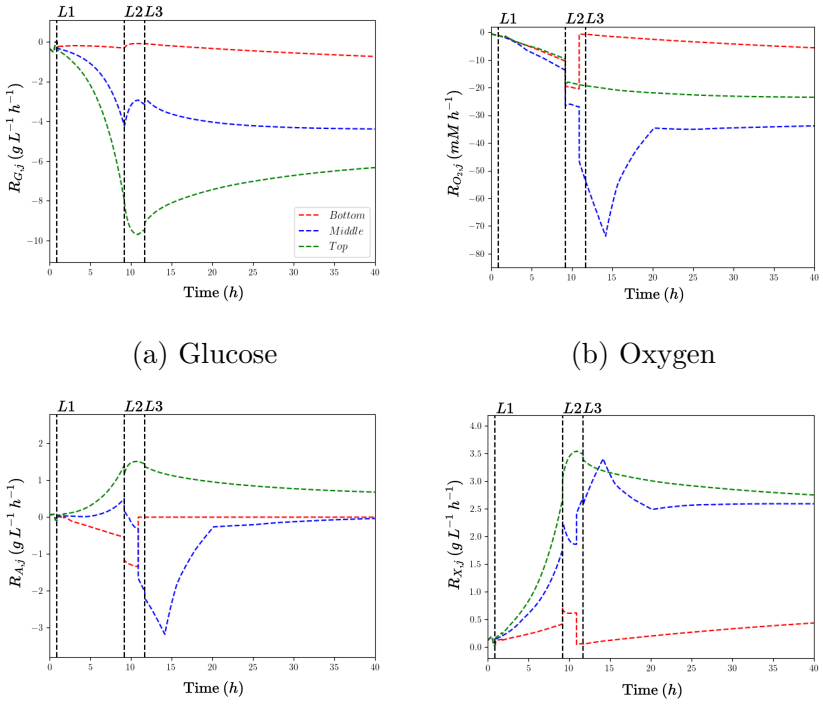
and underestimate the concentration in the middle and bottom regions of the reactor. This suggests that the extent of the concentration gradient is influenced by the hydrodynamics of the reactor. Specifically, the difference in concentration diminishes in the compartmental model when the mixing time is relatively small.

The model can accurately predict the trends in acetate (Figures 5.4b, 5.4d and 5.4f) throughout the entire operation. For example, it effectively captures the second peak in concentration resulting from oxygen deficiency due to control actions (see Figure 5.5). This peak was not predicted by the model from ref. [18] due to possibly inaccuracies in representing the hydrodynamics—particularly since the flow rates between compartments were treated as tuning parameters—and the absence of a detailed description of the oxygen control strategy. The initial peak in acetate concentration occurs when glucose consumption is constrained, causing oxygen to be utilized for acetate consumption. The second peak arises when oxygen becomes temporarily depleted (see Figure 5.5), and the integral terms in the PI control begin accumulating an error to counteract past values until the corrective action becomes sufficiently large to reintroduce additional oxygen into the system. Moreover, throughout different phases of the operation, the model demonstrates a very small extent of concentration gradient of acetate, which is consistent with the experimental data.

Despite slight variations in biomass concentration at different probe locations, the concentration gradient for biomass is negligible, which is accurately predicted by the model. To compare the performance of the presented model to that of ref.[18], the corrected Akaike information criterion was calculated for both models as shown in Table 5.4 above.

Reaction rates ($R_{i,j}$) and metabolic fluxes ($v_{i,j}$) were also determined and shown in Figure 5.8 and Figure 5.9, respectively. The concentration, rate, and metabolic flux profiles collectively indicate that glucose was primarily consumed in the upper region of the bioreactor. In this area, microorganisms utilized both aerobic ($v_{2,j}$) and anaerobic pathways ($v_{4,j}$). Within the middle section, during the fed-batch period with an exponential feed rate ($L1 < t < L2$), glucose was consumed through the anaerobic pathway. However, at approximately the 10 h mark, the controller increased the oxygen level sufficiently for the aerobic pathway to become favorable. After this point, the oxygen level was also adequate for acetate consumption during the fed-batch period with a constant feed rate ($t > L2$). At the bottom of the tank, the aerobic pathway was predominantly favored for most of the operational duration, as the oxygen levels here were higher than in the rest of the tank due to proximity to the air supply. However, the growth rate in this part of the tank remained relatively low since most of the carbon resources had already been consumed in the upper regions.

Since the angular position of the probes was not reported in the original experimental



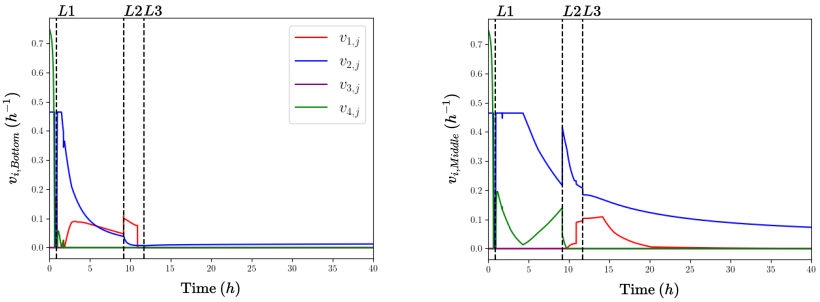
(a) Glucose

(b) Oxygen

(c) Acetate

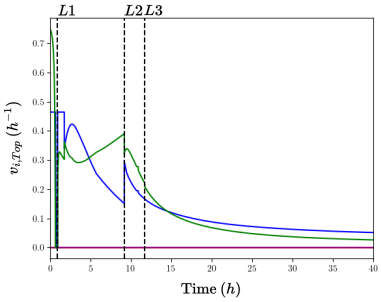
(d) Biomass

Figure 5.8: Reaction rates observed at various heights within the system. **Red**: bottom at 1.12 m, **blue:** middle at 4.04 m, and **green:** top at 6.48 m.



(a) Bottom

(b) Middle



(c) Top

Figure 5.9: Metabolic fluxes measured at the following locations: (a) the lower probe at 1.12 m, (b) the middle probe at 4.04 m, and (c) the top probe at 6.48 m. Each flux is associated with a specific metabolic reaction as outlined: $v_{1,j} : 39.43A + 35O_2 \rightarrow X$, $v_{2,j} : 9.46G + 12.92O_2 \rightarrow X$, $v_{3,j} : 9.84G + 12.73O_2 \rightarrow 1.24A + X$, and $v_{4,j} : 19.23G \rightarrow 12.12A + X$

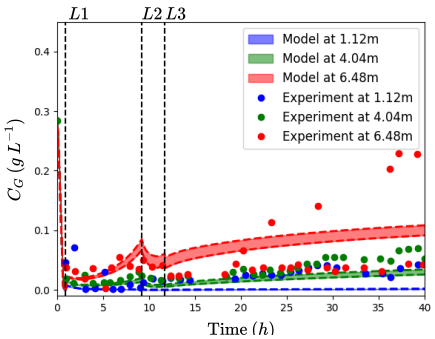


Figure 5.10: Glucose concentration profile at different heights. The range of concentrations results from measurements taken across all compartments at each fixed height, reflecting the uncertainty of the angular positions in the simulation.

work, the effect of this uncertainty was studied and the results shown in Figure 5.10. The figure displays in different colors the range of glucose concentrations occurring across all compartments over time at each of the probe heights. The narrow range of concentration values indicates a low sensitivity of glucose concentration to variations in angular position.

Lastly, the simulation time required was approximately 5 min on a laptop computer, showing the computational complexity advantages of combining the BST metabolic model and compartmental hydrodynamic models for investigating large-scale bioreactors. This represents a significant reduction in computational effort compared to a fully coupled CFD and metabolic flux model.

5.3.3 Effects of flow oscillations

One important observation from experimental data (Figure 5.4a) is the significant fluctuation of the glucose concentration measured by the top probe. These observed fluctuations were attributed to turbulence in ref. [22], which has been widely reported for bioreactors [123, 124, 5]. In order to investigate the effects of oscillations on the concentration gradients they were introduced *mathematically* through an oscillation to the volumetric flow rate matrix,

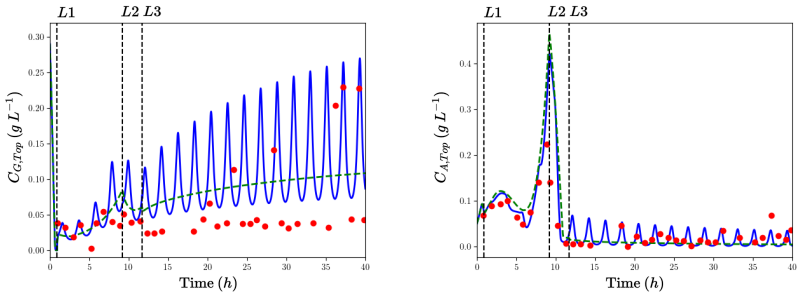
$$\mathbf{Q} = \mathbf{Q}^{bi}\left[1 + M\sin(\omega t)\right] \tag{5.10}$$

where M and ω represent an arbitrary magnitude ratio and frequency of the oscillations, respectively. The magnitudes and frequencies that are considered are based on the power spectra density of velocity measurements in a bioreactor with a Rushton impeller [104]. Moreover, based on the reported power spectra, oscillations at higher frequencies than the ones simulated here are also possible, but they do not affect the concentration profiles due to filtering by the system. To gain insight into the filtering mechanism, reference is made to the frequency response analysis [125].

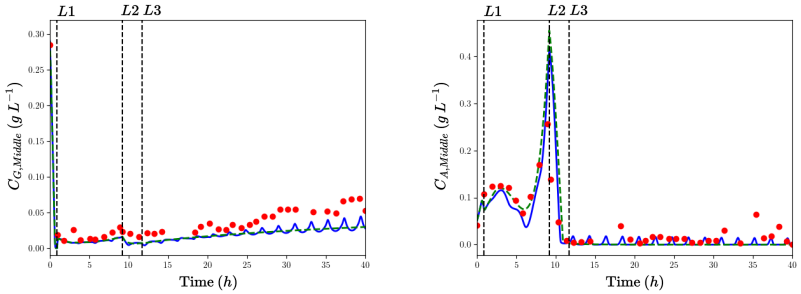
The frequency and magnitude of flow oscillations are impacted by the location relative to tank geometry and the flow regime [104, 126]. Thus, the use of Equation (5.10) may tend to overestimate these effects due to its introduction of oscillations throughout the compartment network. However, this approach provides a straightforward method to incorporate experimentally observed oscillations into compartmental models while ensuring mass conservation.

The model described in Section 5.3.2 was simulated using the oscillating volumetric flow rate method, with results shown in Figure 5.11. The agreement with experimental data was qualitatively improved, particularly for the glucose concentration at the top probe location. The concentration oscillations predicted by the simulation at the middle and bottom probes became increasingly damped, yet remained within reasonable ranges of experimental measurements as time evolved.

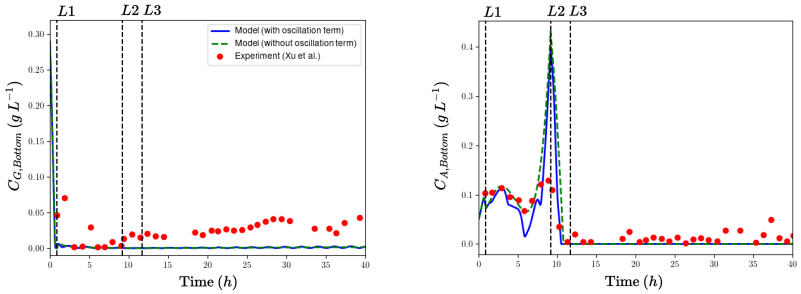
The cycle of acetate production and consumption during the fed-batch phase with a constant feed ($t > L2$) resulting from the oscillatory flow suggests the potential for acetate accumulation during the operation of the system. These findings evidently show that local hydrodynamic oscillations, whether caused by turbulence or other effects, significantly influence metabolic reactions. Consequently, these flow oscillations have a notable impact on concentration gradients within large-scale bioreactors.



(a) Glucose (6.48m from bottom) (b) Acetate (6.48m from bottom)



(c) Glucose (4.04m from bottom) (d) Acetate (4.04m from bottom)



(e) Glucose (1.12m from bottom) (f) Acetate (1.12m from bottom)

Figure 5.11: Comparing simulation results with turbulent-induced oscillation flow rates (where $M = 0.7$ and $\omega = 3\,\mathrm{h}^{-1}$) to experimental data in a large-scale bioreactor: (a,c,e) refer to glucose, while (b, d, f) refer to acetate, studied at three distinct locations: (a, b) the top probe (6.48 m from the bottom of the tank), (c, d) the middle probe (4.04 m from the bottom of the tank), and (e, f) the bottom probe (1.12 m from the bottom of the tank).

5.4 Conclusion

This chapter developed and validated a coupled metabolic flux–compartmental model for large-scale aerated bioreactors, providing an effective framework for simulating complex bioprocesses with high computational efficiency.

A key innovation of this approach is the reformulation of flux balance analysis (FBA) as a point location problem, which is efficiently solved using a binary search tree (BST) representation. By framing the FBA problem this way, optimal flux distributions can be precomputed through multiparametric linear programming. The resulting BST structure substantially reduces the computational cost of evaluating local growth rates, making it well-suited for long-term simulations. Moreover, this method integrates seamlessly with compartmental models, enabling dynamic simulations of large-scale bioreactor systems within optimization and control frameworks.

The integration of control strategies into the bioreactor simulations further enhanced model accuracy, resulting in more accurate metabolite concentration predictions and better agreement with experimental observations. The results highlight the critical role of controllers in capturing oscillations in metabolite concentrations caused by transient oxygen depletion. Additionally, simulations revealed the influence of local flow oscillations, offering insights into experimentally observed oscillatory measurements.

Ultimately, the coupled dynamic flux balance/compartmental model achieved both accuracy and computational efficiency, enabling simulations of over 40 hours of bioreactor operation to be completed within minutes on standard desktop hardware. These findings underscore the model's potential for real-time applications in model predictive control and bioreactor design optimization.

Finally, it is crucial to emphasize that the kinetic parameters used for this large-scale bioreactor require recalibration rather than direct adoption from small-scale bioreactor calibrations. This limitation suggests that intracellular heterogeneity may contribute to variations in parameter values, highlighting the need for scale-independent parameter estimation in bioprocess modeling.

Chapter 6

Development of an MM-TND to Simulate a Coupled PBM-CM

6.1 Introduction

While concentration gradients in large-scale bioreactors have been extensively studied through simulations, intracellular heterogeneity has received less attention due to the complexity and computational demands of modeling. Intracellular heterogeneity arises as microorganisms traverse varying conditions in the reactor, encountering gradients in substrate concentration, pH, and oxygen. Finite adaptation rates result in dynamic changes in intracellular states that may lag behind extracellular changes, creating a heterogeneous microbial population. For example, it was hypothesized that the discrepancies between data and model predictions reported in Chapter 5 may be due to the fact that the differences between the intracellular and extracellular environments were not accounted for. Thus, accounting for these differences as proposed in the current study may improve the fitting accuracy.

Intracellular heterogeneity is commonly modeled using PBEs. These models track the distribution of cells across different intracellular states, enabling the study of microbial population heterogeneity in bioreactors. Several studies have integrated PBMs with metabolic, and compartmental models to improve representation (see Table 3.2). Morchain et al. [60] introduced the bin method to solve a PBE, using the cell growth rate as the internal state. This method has been further used in other studies [6, 18, 56]. However, the bin method, which partitions the variable space into fixed intervals, is unable to capture finer variations within each bin. This often results in overestimation or underestimation of the real stan-

dard deviation. Furthermore, the bin method is computationally costly as it requires one differential equation for each bin. Reducing the number of intervals to reduce computation will obviously result in less accurate results (see Section 3.4.1). To address this limitation, the MOM has been applied. An important argument that favors the use of moment methods over less computational expensive is that, in practice, distributions will be measured by flow cytometry, which does not provide very fine resolution due to noise and sensitivity limitations. As a result, only moment-derived quantities (e.g., mean and variance) are typically reported. To solve intracellular distributions, Pigou et al. [56] calculated their moments by the numerical quadrature moment methods. However, three challenges were identified in that study:

1. The Gaussian kernel used in the extended quadrature method of moments (EQMOM) is observed to disregard constraints, i.e., positivity of some intracellular states (e.g., concentrations) and unrealistic upper bounds.

2. Degeneration may occur as the variance of the distribution approaches zero when the concentration in the tank becomes homogeneous, potentially leading to numerical errors.

3. It is well known that microorganisms can exhibit nonlinear behavior, and metabolic rates may be discontinuous. As a result, a structured metabolic model is required, necessitating the continuous reconstruction of the number density function to ensure accurate calculations.

To address these challenges, this study introduces an approach to solving the moment equations using a truncated normal distribution (MM-TND). This integration enhances the accuracy of microbial population modeling by capturing finer-scale variations while ensuring realistic physical bounds for the internal state distribution. To prevent inaccuracies arising from discontinuous metabolic rates, MM-TND is reconstructed at every time step. Additionally, moment inversion and a normalized integration range scheme are employed to facilitate the integration of function products of MM-TND, effectively mitigating the issue of variance approaching zero.

By focusing on the effects of intracellular heterogeneity and extracellular gradients on microbial populations, this work advances the understanding of large-scale bioprocesses and provides a robust modeling framework for future biotechnological applications.

In summary, this study aims to develop a computational framework, MM-TND, to efficiently model large-scale bioreactors and investigate the formation of intracellular heterogeneity and concentration gradients. Additionally, the proposed methods will be validated

through comparisons with existing numerical approaches and experimental data for *E. coli* fed-batch fermentation [22].

6.2 Methodology

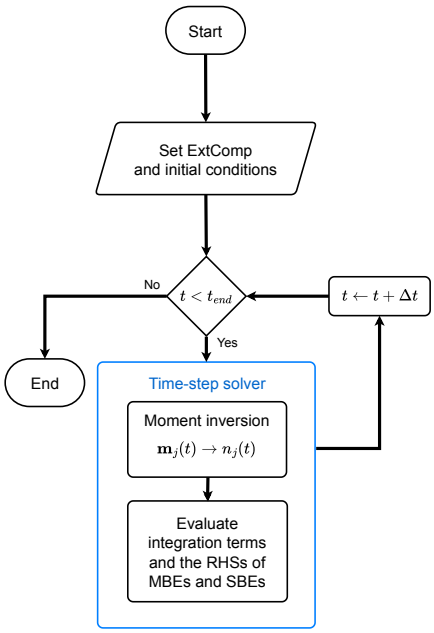


Figure 6.1: Diagram of the MM-TND solver.

In this study, the MM-TND was utilized to simulate large-scale bioreactors. The proposed method follows the same idea as EQMOM with a single distribution. A set of MTEs, given in Equation (2.49), and MBEs for external substrates, given in Equation (2.3), were solved as a system of ordinary integro-differential equations (OIDEs) using an adaptive time-step initial value problem (IVP) solver as the outer framework, see Figure 6.1. In this research, the implicit method based on backward-differentiation formulas, `scipy.integrate.BDF` [117], was used to solve the IVPs. Within each intermediate "substep" of the solver, the following steps were carried out:

1. Moment inversion: Reconstruction of the average number density function, n_j, assuming a univariate truncated normal distribution (Equation (3.25)). This requires determining the distribution parameters: mean μ and variance σ^2. Based on the relationship

between the moments and the distribution parameters of the normal distribution, only moments up to the second order are required to determine these parameters. In this study, the procedure for determining μ and σ^2 was formulated as an optimization problem aimed at finding the truncated normal distribution that best matches the known moments at each substep. At the start of the process, $\mathbf{y}_j = [y_{j,0}, y_{j,1}, y_{j,2}]^\top$, the column vector of known moments at each substep, is determined. From Equation (3.25), it follows that $y_{j,\beta} = m_{\beta,j}/c_{N,j}$. On the other hand, let $\mathbf{y}_j^*$ be the vector of moments (from zeroth to second order) of the truncated normal distribution, evaluated at any combination of μ and σ^2. Then, using the recurrence relation from Equation (3.32), the optimization problem for determining μ and σ^2 can be expressed as:

$$\min_{\mu_j, \sigma_j^2} \|\mathbf{y}_j - \mathbf{y}_j^*\|^2 \tag{6.1}$$

where $\|\cdot\|^2$ represents the squared Euclidean norm, which measures the sum of squared differences between the vector of known moments and the vector of moments computed at the sought decision variables μ and variance σ^2. The following initial guesses are suggested for the optimization scheme:

$$\mu_j^0 = \max(y_{j,1}, z_{\min}) \tag{6.2}$$

$$\sigma_j^0 = \max(y_{j,2} - y_{j,1}^2, 0) \tag{6.3}$$

These initial guesses correspond to the mean and variance of the normal distribution, respectively.

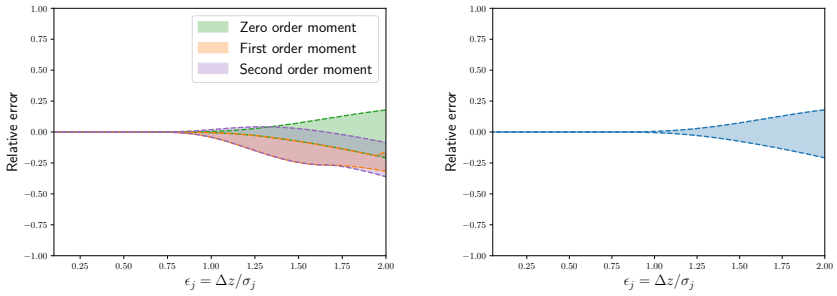
2. Integration of the integral term: Numerical evaluation of the right-hand side (RHS) terms in MTEs , Equation (2.49), and MBEs, Equation (2.3), involves numerical integrations. A naive approach is to employ a fixed high-resolution grid, but this proves impractical for two primary reasons. First, a high-resolution grid is only necessary near the region where the distribution is concentrated. Using excessive grid points far from this region unnecessarily increases computational time, while using a small number of points reduces accuracy.

Furthermore, in extreme cases where variances approach small values, the required grid resolution may be infinitely small, rendering the approach infeasible. Addressing such cases requires a mathematical treatment and the development of an adaptive integration grid that efficiently captures the distribution locally.

While adaptive integration schemes, such as `quad` from the `scipy.integrate` library, are readily available, they are not specifically optimized for scenarios involving normal distributions when computational efficiency is paramount.

Following the above, an adaptive integration scheme was developed to expedite the evaluation of functions involving truncated normal distributions. The first component of this scheme involves the integration bounds. Rather than integrating over the full range from z_{min} to z_{max}, the bounds are narrowed to $[\max(\mu_j - 5\sigma_j, z_{min}), \min(\mu_j + 5\sigma_j, z_{max})]$. This adjustment is justified by the fact that the probability density, τ_j, outside this range, is negligibly small.

The second component of the scheme addresses the selection of grid points. Given the potential transition of the distribution from a normal to a truncated form as the mean approaches one of the bounds, Simpson's 1/3 rule with a regular grid was employed. The number of grid points was determined through numerical experiments to balance accuracy and computational efficiency.



(a) (b)

Figure 6.2: Numerical analysis results highlight the relative error between the numerical integration value, computed using Simpson's 1/3 rule, and the expected value as the standard deviation approaches zero for $\int_{-\infty}^{\infty} g\phi dz$. This relative error is illustrated for various functions: (a) the zeroth-order moment where $g = 1$, the first-order moment where $g = z$, and the second-order moment where $g = z^2$, and (b) the Monod equation where $g(z) = z/(k + z)$. The shaded area in the figure represents the range of error introduced by variations in the position of the population mean relative to the integration grid.

A numerical experiment was conducted to investigate how the ratio between the interval size of the fixed integration grid, Δz, to the standard deviation, σ_j, affects the relative error. As expected, the results in Figure 6.2, indicate that the integration deviates from the analytical solution when the grid interval is too large relative to the standard deviation. In other words, the number of integration grid points must ensure that $\sigma_j > \Delta z$. Defining the ratio $\epsilon_j = \Delta z/\sigma_j$, the number of integration grid points can be calculated as $n_g =$

$\lceil (2l_j/(\epsilon_j\sigma_j)) - 1 \rceil$, where l_j represents the maximum distance from the mean, μ, to one of the bounds, and is set to $5\sigma_j$. The actual interval size is then given by $\Delta z = 2l_j/(\mathrm{n}_g - 1)$. To achieve accurate results, ϵ_j is assigned a value of 0.5, which satisfies the condition $\epsilon_j < 1$. `torchquad.integration.simpson`, which uses the GPU for computation, was then used to perform the integration over the identified fixed grid.

All simulations were carried out on a laptop running Ubuntu 64-bit, featuring a 12$^{\text{th}}$ Generation Intel® Core™ i7-1260P processor (12 cores, 16 threads, base frequency of 400 MHz and a maximum frequency of 4.7 GHz), 16 GB of RAM, a 512 GB solid-state drive (SSD), and integrated Intel® Iris® Xe Graphics.

6.3 Results and discussion

6.3.1 Numerical challenges of solving PBEs using the QMOM

Since the QMOM has been reported for describing intracellular distributions [56], it is important to emphasize the relative advantages of the proposed method, based on truncated Gaussian distributions. Pigou et al. [56] reported two versions of the quadrature method: one based on integrating the moments using the QMOM and the other based on an assumed normal distribution, known as the EQMOM. Both methods have significant limitations compared to the proposed approach.

The QMOM uses the method of collocations for performing integrations of the original population balances. These integrations are based on the values of the functions that are integrated, e.g. reaction rates, at the collocation points. However, as shown in this study, these functions are often discontinuous and thus integration based on a limited number of collocation points may lead to significant errors.

To illustrate this limitation, consider the following example. A population is distributed as a normal distribution over an arbitrary internal state, z, with an interfacial (mass or reaction) rate, w, defined as a discontinuous function of z, as shown in Figure 6.3. The reaction rate, $r = \int_0^{10} wndz$, was calculated using Equation (2.23) with both the QMOM approach and our proposed method, and the results were compared with the exact solution.

With QMOM, the reaction rate was determined by:

$$r \approx \sum_{i=1}^{\mathrm{n}_{CL}} \omega_i w(\xi_i) \tag{6.4}$$

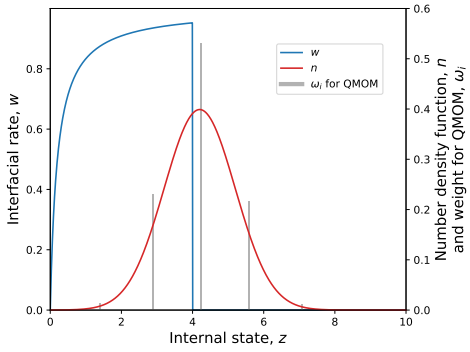


Figure 6.3: Demonstration of numerical challenges due to the use of a structured metabolic model, leading to discontinuities in the values of mass transfer or reaction rates, w.

where ξ_i and ω_i are the collocation points and their corresponding weights, respectively. The number of collocation points, n_{CL}, is given by $\mathrm{n}_{CL} = (O+1)/2$ where O is the maximum order of the set of moments. The weights and collocation points were obtained using the product-difference algorithm [91, 88]. Moments up to the ninth order ($O = 9$) were considered.

The exact value of the reaction rate, r, was 0.3955. The QMOM method, however, significantly underestimated this value, yielding 0.2270 with an error of 42.60%. In contrast, our proposed method produced a more accurate estimate of 0.4192, corresponding to an error of approximately 6%.

The higher error observed with QMOM arises from the dependence of the reaction rate on the interfacial rate, w, which is highly sensitive to the placement of collocation points in the internal state space (see the ω_i profile in Figure 6.3). This issue can be mitigated by reconstructing n using an assumed distribution, such as a truncated normal distribution. Assuming such a distribution reduces the error to only 6%. However, a normal distribution can be only assumed when its mean is far from the lower bound. Otherwise, a truncated distribution is required.

The second quadrature method, EQMOM, assumes that the population follows a normal (Gaussian) distribution. However, this assumption can lead to significant inaccuracies when part of the distribution extends beyond the physical bounds of the internal state space, assigning probability to unrealistic values such as negative concentrations. To illustrate this issue, a set of moments up to order 2 was reconstructed as both a normal distribution, $n = \phi$, and a truncated normal distribution, $n = \tau$, over an internal state space that includes an unphysical region below zero, as shown in Figure 6.4. When the nor-

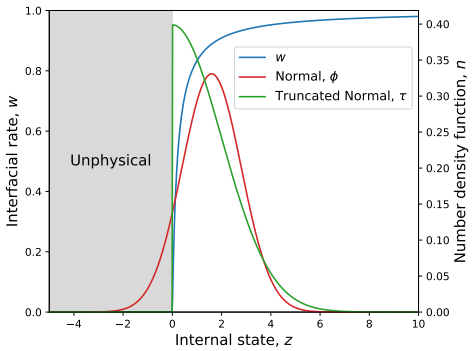


Figure 6.4: Demonstration of numerical challenges due to the use of a normal distribution, ϕ, instead of a truncated normal distribution, τ, forcing cells to reside in an unphysical domain.

mal distribution is used, a portion of it lies in the negative domain, which is not physically meaningful.

In this example, the interfacial rate w is defined to be zero within the unphysical region and follows Monod-type kinetics in the positive domain. Using the integration method proposed in Section 6.2, the reaction rate expression $r = \int_{-10}^{10} wn, dz$ was evaluated. The resulting value of r was 0.7731 when $n = \phi$, and 0.8028 when $n = \tau$. The difference in these values highlights the deviation caused by the portion of the normal distribution that incorrectly assigns probability to the unphysical region. According to the containment condition in Equation (2.26), microorganisms cannot exist in this region. Therefore, the truncated normal distribution provides a more accurate and physically consistent representation near the bounds, leading to a more reliable estimate of the reaction rate.

In summary, the truncated normal distribution method overcomes QMOM and EQ-MOM limitations by accurately capturing discontinuities and ensuring physically meaningful distributions, thereby reducing integration errors and enhancing reliability for intracellular distribution modeling.

6.3.2 Validation of the proposed MM-TND

To validate the proposed method, simulation results were compared against those obtained using the bin (or class) method described by Pigou et al. [18], as well as experimental data [22] from a batch culture of *E. coli* conducted in a 15-liter stirred tank reactor with a working volume of 7 liters. In this system, mixing and oxygen mass transfer were assumed

to be non-limiting, allowing the reactor to be treated as perfectly mixed. Consequently, the associated compartmental model comprised a single compartment ($n_{EC} = 1$). During the experiment, the oxygen concentration was fix to approximately 30% of saturation. Therefore, the conservation equation for oxygen in the extracellular environment was omitted. Following Pigou et al. [18], the cell density (ρ_c) and cell volume (v_c) are assumed to remain constant across the entire population. Furthermore, only the heterogeneity in the cell growth rate is considered, implying that the intracellular heterogeneity can be quantified by a univariate distribution, with $n_z = 1$. It follows that the variable z refers to the internal state of the cell growth rate. It is also assumed that offspring inherit all the considered states of their parent (Equation (2.21)) and that cell death is negligible (i.e., $q_{N,j} = 0$). Based on these assumptions, the resulting PBE simplifies to:

$$\frac{\partial x}{\partial t} + \nabla^z \cdot (x\boldsymbol{\zeta}) = r_X x \tag{6.5}$$

where $x = n\rho_c v_c$. The subscript j is omitted from the intracellular state, since only one component is considered.

Given that x takes the form $x = c_X \tau$, the following MTE is obtained:

$$\frac{dm_\beta}{dt} = c_X \int_0^{z_{max}} (\beta z^{\beta-1}\zeta + z^\beta r_X)\tau dz \quad \text{for} \quad \beta \in \{0, 1, 2\} \tag{6.6}$$

While $z_{\min}$ is set to zero, as the growth rate cannot be negative, $z_{\max}$ must be set to the maximum possible r_X over the entire operating time. The maximum rate can be estimated from a priori knowledge of the range of concentrations that will occur during the fermentation. Following Pigou et al. [18], the following rate of change of the internal state, ζ_j, was included:

$$\zeta_j = (\frac{1}{t_A} + z)(z_j^* - z) \tag{6.7}$$

Where, t_A is the characteristic time constant that controls the dynamics of cell adaptation, and $z_j^*(\mathbf{c}_j)$ is the optimal specific growth rate as a function of the external concentrations, $\mathbf{c}_j$. The subscript j can be omitted in the case of a system with a single compartment.

The following mass balance equations for glucose and oxygen were included:

$$\frac{dc_G}{dt} = c_X \int_0^{z_{max}} r_G \tau d\mu \tag{6.8}$$

$$\frac{dc_A}{dt} = c_X \int_0^{z_{max}} r_A \tau d\mu \tag{6.9}$$

The equations for $z^*(\mathbf{c})$, $r_A(z,\mathbf{c})$, $r_G(z,\mathbf{c})$, and $r_X(z,\mathbf{c})$ from Pigou et al. [18] were used (Note that in Pigou et al. [18], μ_j^a was used instead of $r_{X,j}$; in this research, μ refers to the population mean). The initial condition is provided in Table 6.1, while the model constants are given in Table 6.2.

Table 6.1: Initial concentrations

Symbol	Values			Unit
	Section 6.3.2	Section 6.3.4 (Lab-scale,$7L$)	Sections 6.3.3, 6.3.4 and 7.3 (Large-scale, $22m^3$)	
External concentrations				
$c^0_{G,j}$	13.800	0.5630	0.2900	$g_G L^{-1}$
$c^0_{A,j}$	0	0.0800	0.0500	$g_A L^{-1}$
$c_{O,j}$	2.6336	2.6336	1.7557	$mg_O L^{-1}$
$c^0_{I,j}$	N/A	0	0	$g_I L^{-1}$
Moments				
$m^0_{0,j} = c^0_{X,j}$	0.0770	0.0100	0.0500	$g_X L^{-1}$
$m^0_{1,j}$	0.0308	0.0040	0.0200	$g_X L^{-1} h^{-1}$
$m^0_{2,j}$	0.0123	0.0016	0.0080	$g_X L^{-1} h^{-2}$

As shown in Figure 6.5, the concentration profiles for glucose, acetate, and biomass closely matched the experimental data. Both the MM-TND and bin methods (using 100 bins) produced similar results, thus validating the proposed method.

The mean, μ_z, and standard deviation, σ_z, (which measures internal heterogeneity) for MM-TND were conducted based on Section 6.2. The calculations for μ_z and σ_z in the bin method are done as per the following equations:

$$\mu = \sum_{i=1}^{\mathrm{n}_b} z_i f_i, \tag{6.10}$$

$$\sigma^2 = \left(\sum_{i=1}^{\mathrm{n}_b} f_i z_i^2 \right) - \mu^2, \tag{6.11}$$

where n_b is the number of bins, z_i represents the bin centers, and f_i denotes the normalized frequencies (or probabilities) of the bins.

The results of μ_z, shown in Figure 6.6 were comparable for both methods. However, the standard deviation, σ_z, was significantly larger for the bin method with 100 bins compared to the MM-TND. As the bioreactor has a relatively low volume of 15 liters, it was

Table 6.2: Model constants

Symbol	Values		Unit
	Section 6.3.2	Sections 6.3.3, 6.3.4 and 7.3	
Molar masses			
M_X	113.1	113.1	$g_X mol_X^{-1}$
M_G	180.2	180.2	$g_G mol_G^{-1}$
M_O	32	32	$g_O mol_O^{-1}$
M_A	59	59	$g_A mol_A^{-1}$
Affinity and inhibition constants			
K_G	0.05	0.05	$g_G L^{-1}$
K_A	0.05	0.05	$g_A L^{-1}$
K_O	1×10^{-4}	1×10^{-4}	$g_O L^{-1}$
$K_{i,G}$	0.2	0.1	$g_G L^{-1}$
$K_{i,A}$	3	3	$g_A L^{-1}$
$K^o_{i,A}$	4	4	$g_A L^{-1}$
Yields			
Y_{EG}	12.05	12.05	$mol_E mol_G^{-1}$
Y^{max}_{XG}	1.32	1.35	$mol_X mol_G^{-1}$
m	250×10^{-6}	220×10^{-6}	$mol_G g_X^{-1} h^{-1}$
Y_{OG}	6	6	$mol_O mol_G^{-1}$
Y^o_{EG}	20	20	$mol_E mol_G^{-1}$
Y^f_{EG}	3	3	$mol_E mol_G^{-1}$
Y_{AG}	2	0.8	$mol_A mol_G^{-1}$
Y_{EA}	4	4	$mol_E mol_A^{-1}$
Y_{XA}	0.4	0.8	$mol_X mol_A^{-1}$
Y_{OA}	2	2	$mol_O mol_A^{-1}$
Y^o_{EA}	4.67	4.67	$mol_E mol_A^{-1}$
Y_{IX}	N/A	0.12	$g_I g_X^{-1}$
k_I	N/A	0.03	$g_G g_I^{-1}$
Biological limitations			
μ^G_{max}	0.663	1.4	h^{-1}
μ^A_{max}	0.032	0.13	h^{-1}
Φ^{max}_O	15.1×10^{-3}	14×10^{-3}	$mol_O g_X^{-1} h^{-1}$
Growth rate adaptation			
t_A	1.9	1.9	h

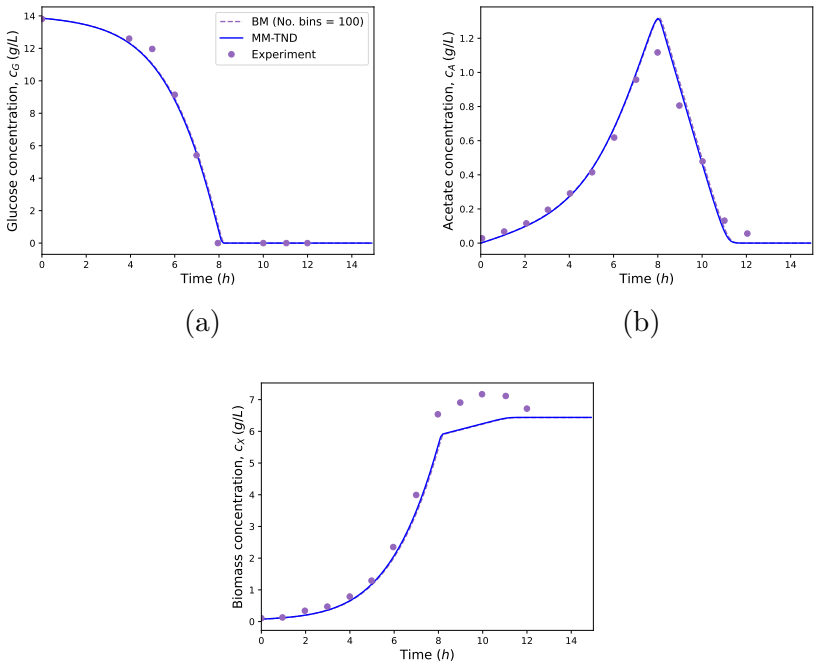
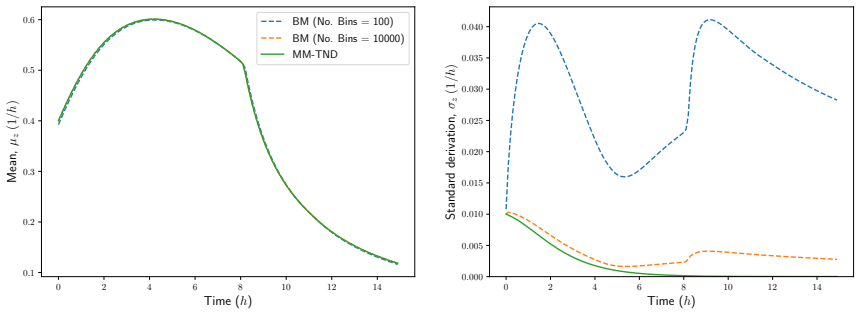


Figure 6.5: Concentration profiles of (a) glucose (c_G), (b) acetate (c_A), and (c) biomass (c_X), obtained using the bin method (BM) with varying numbers of bins, and the method of moments with an assumed truncated normal distribution (MM-TND). The model predictions agree across methods, resulting in a single line in each plot.



(a) (b)

Figure 6.6: (a) Mean (μ_z) and (b) standard deviation (σ_z) of the growth rate, r_X, as determined using the bin method (BM) with different number of bins and the method of moments with an assumed truncated normal distribution (MM-TND).

suspected to be well-mixed, and thus there should be no gradient to support the formation of heterogeneity. It was suspected that the bin method prediction of a large variance was due to the use of a relatively small number of bins.

Even though the simulations assumed an initial standard deviation (heterogeneity), the heterogeneity was expected to dissipate over time due to the small value relative to the mixing conditions. Thus, the initial use of a small number of bins was suspected to lead to an incorrect estimate of the true standard deviation. When the number of bins is increased to 10,000 σ_z approaches zero, as shown in Figure 6.6. This resolution problem does not occur for the MM-TND results, where σ_z approached zero as expected.

Hence, this comparison clearly shows that using the bin method to study heterogeneity in cell populations poses challenges not only due to the numerical errors discussed but also because of the high number of bins required for accuracy. This requirement significantly increases computation time, as shown in Section 6.3.2 where the bin method with 10,000 bins is 4 orders of magnitude slower than the proposed MM-TND approach.

Method	CPU time (s)
BM (100 bins)	2.95
BM (10,000 bins)	~92,000
MM-TND	2.85

6.3.3 Effect of flow rate on gradients and intracellular heterogeneity

As the proposed MM-TND method is specifically designed to capture intracellular heterogeneity, this section discusses an example that illustrates this capability. The MM-TND method was used to simulate a large-scale bioreactor with a batch culture of *E. coli*, focusing on the formation of gradients and internal heterogeneity influenced by the rotational speed of the impellers. Since the rotational speed directly influences the vertical volumetric flow rate, the analysis in the current example focuses on the effect of the flow rate on both gradients and intracellular heterogeneity.

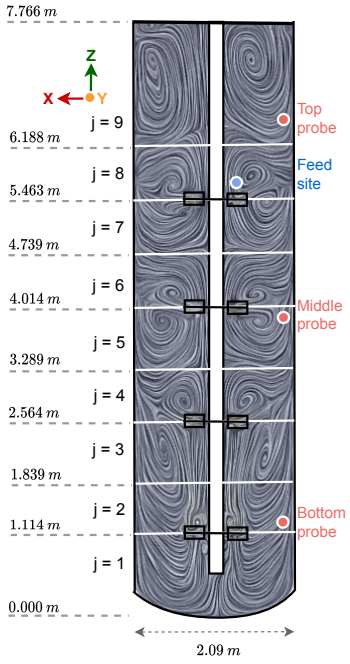


Figure 6.7: CM with 9 compartments of a large 30 m^3 bioreactor.

All assumptions made in Section 6.3.2 were also applied for the current study, except for the single-compartment model assumption. Here, a compartmental model with 9 compartments ($n_{EC} = 9$), as shown in Figure 6.7, was used to study a 30 m^3 bioreactor with a working volume of 22 m^3 [22]. The compartments were obtained by manually partitioning the mixing zone cross-sectionally at the height of the impellers and midway between them.

This compartmentalization particularly captures the main circulation patterns occurring within the compartments, as shown in Figure 6.7. In this section, 9 compartments are considered for both the extracellular and intracellular space and no merging of compartments is undertaken.

For simplicity, the flow rate between each compartment and its adjacent compartments was fixed at a constant value, q. Three cases were studied: (a) the characteristic timescale of advection, $\tilde{t}_{q,i,j}$, is much greater than that of the reaction, $\tilde{t}_{r,i,j}$, with $q = 0.01152\,\mathrm{L/h}$; (b) $\tilde{t}_{q,i,j}$ is comparable to $\tilde{t}_{r,i,j}$, with $q = 1,152\,\mathrm{L/h}$; and (c) $\tilde{t}_{q,i,j}$ is much smaller than $\tilde{t}_{r,i,j}$, with $q = 1,152,000\,\mathrm{L/h}$.

The following MTEs and MBEs were applied:

$$\begin{aligned}\frac{dm_{\beta,j}}{dt} &= c_{X,j}\int_0^{z_{max}}(\beta z^{\beta-1}\zeta_j + z^\beta r_{X,j})\tau_j dz \\ &\quad + \frac{1}{v_j}\left(\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}} q_{j'\to j}m_{\beta,j'} - m_{\beta,j}\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'}\right) \\ &\quad \text{for} \quad \beta \in \{0,1,2\}\end{aligned} \tag{6.12}$$

$$\begin{aligned}\frac{dc_{G,j}}{dt} &= \frac{1}{v_j}\left(\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}}(q_{j'\to j}c_{G,j'}) - c_{G,j}\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'}\right) \\ &\quad + c_{X,j}\int_0^{z_{max}} r_{G,j}\tau_j d\mu\end{aligned} \tag{6.13}$$

$$\begin{aligned}\frac{dc_{A,j}}{dt} &= \frac{1}{v_j}\left(\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}}(q_{j'\to j}c_{A,j'}) - c_{A,j}\sum_{\substack{j'=1\\j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'}\right) \\ &\quad + c_{X,j}\int_0^{z_{max}} r_{A,j}\tau_j d\mu\end{aligned} \tag{6.14}$$

The rate of change of the internal state, ζ_j, was determined using Equation (6.7). The metabolic model used to define $z^*(\mathbf{c})$, $r_A(z,\mathbf{c})$, $r_G(z,\mathbf{c})$, and $r_X(z,\mathbf{c})$ was modified from Pigou et al. [18], as discussed in Section 6.3.4. The initial condition is provided in Table 6.1. The microorganisms were assumed to be settled at the bottom of the reactor

at the beginning. Accordingly, the initial condition for the moments in Compartment 1 were as reported in Table 6.1, while a minimal values (10^{-8}) were assigned to the other compartments for numerical reasons. The model constants are given in Table 6.2.

Characteristic time profiles in Figure 6.8 are shown for the 3 cases mentioned above, i.e.:

- $\tilde{t}_{q,i,j}$ is significantly greater than $\tilde{t}_{q,i,j}$. (Figures 6.8a to 6.8c)
- $\tilde{t}_{q,i,j}$ is comparable to $\tilde{t}_{r,i,j}$. (Figures 6.8d to 6.8f)
- $\tilde{t}_{q,i,j}$, is significantly smaller than $\tilde{t}_{r,i,j}$. (Figures 6.8g to 6.8i)

These characteristic times are calculated as per equation in Table 6.3 for $\tilde{t}_{r,i,j}$ of:

- Glucose (Figures 6.8a, 6.8d and 6.8g)
- Acetate (Figures 6.8b, 6.8e and 6.8h)
- Cell population (Figures 6.8c, 6.8f and 6.8i)

These plots confirm that all possible ratios of advection to reaction, adaptation and growth rates are being tested. Bands of characteristic time values were obtained by collecting the characteristic time values for each of the compartments at each time interval of the simulation. For these three scenarios that considered different ratios of advection to reaction, adaptation and growth, the extracellular concentrations of glucose, acetate, and average biomass in the bioreactor are calculated for each of the 9 compartments and are plotted in Figure 6.9. Then, using the proposed method, the intracellular mean and standard deviations for growth rates at different flow rates for each of the 9 compartments are calculated and presented in Figure 6.10.

Table 6.3: Various characteristic times at Compartment j for further analysis

Process	**Expression**	
Advection of Substance i (resident time)	$\tilde{t}_{q,i,j} = \frac{v_j}{\sum_{\substack{j'=1 \\ j' \neq j}}^{\mathrm{n_{EC}}} q_{j \to j'}}$	(6.15)
Reaction of Substance i	$\tilde{t}_{r,i,j} = \left\lvert \frac{c_{i,j}}{r_{i,j}} \right\rvert$	(6.16)
Adaptation of cells (mean value)	$\tilde{t}_{\zeta,z_k,j} = \left\lvert \int_{\Omega^z} \frac{z_{k,j}}{\zeta_{k,j}} n_j d\mathbf{z} \right\rvert$	(6.17)
Generation of cells (mean value)	$\tilde{t}_{h,j} = \left\lvert \int_{\Omega^z} \frac{n_j}{h_j} d\mathbf{z} \right\rvert$	(6.18)

It is evident from Figure 6.9 that concentration gradients in glucose, acetate, and biomass, form when the characteristic timescale of advection was comparable to or longer than that of the reaction Figures 6.8a to 6.8f. These results are expected, as the impeller speed was not fast enough to homogenize the tank, unlike the case where the characteristic timescale of advection was shorter, as depicted in Figures 6.8g to 6.8i.

Internal heterogeneity of the growth rate emerged in the presence of extracellular concentration gradients and when $\tilde{t}_{q,X,j}$ was comparable to $\tilde{t}_{\zeta,z,j}$, as shown by the color bands in Figure 6.10b.

It is important to note that microbial division occurs in such a way that offspring inherit the full growth rate, z, of their parent. As a result, cell generation does not significantly contribute to changes in internal heterogeneity. Consequently, changes in the characteristic timescale of cell generation, $\tilde{t}_{h,j}$, do not significantly affect heterogeneity.

These results are consistent with expectations: within a compartment, cells with varying intracellular states come together. Furthermore, as each cell requires additional time to transition from its initial state to a new equilibrium state, intracellular heterogeneity arises.

For the case when $\tilde{t}_{q,X,j}$ ($q = 0.01152$ L/h) is longer than $\tilde{t}_{\zeta,z,j}$, heterogeneity emerges in the second compartment from the bottom of the bioreactor ($j = 2$), as shown in Figure 6.10a. This compartment functions as a local mixing zone where two homogeneous populations—one originating from Compartment 1 and the other from Compartment 3—interact. Each of these populations has distinct homogeneous internal states. When the adaptation rate is not infinitely fast ($\tilde{t}_{q,X,j} \neq 0$), heterogeneity formation in Compartment 2 is expected due to the incomplete equilibration of internal states.

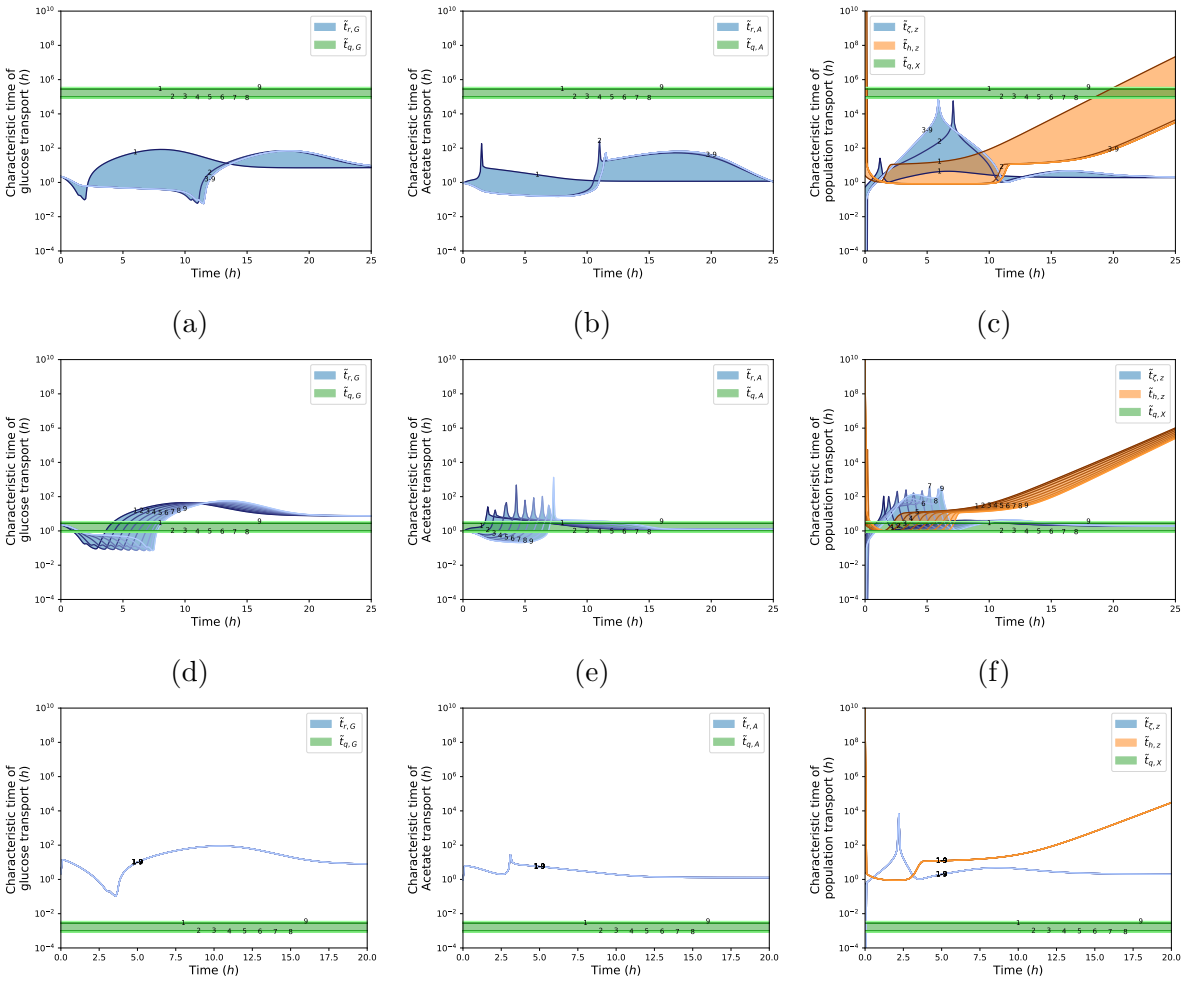


Figure 6.8: Characteristic time profiles for various phenomena in the bioreactor simulation, computed using equations from Table 6.3, under the following flow rates: (a-c) $q = 0.01152\,\mathrm{L/h}$, (d-f) $q = 1,152\,\mathrm{L/h}$, and (g-i) $q = 1,152,000\,\mathrm{L/h}$. The results are shown for (a, d, g) glucose, (b, e, h) acetate, and (c, f, i) cell population. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments.

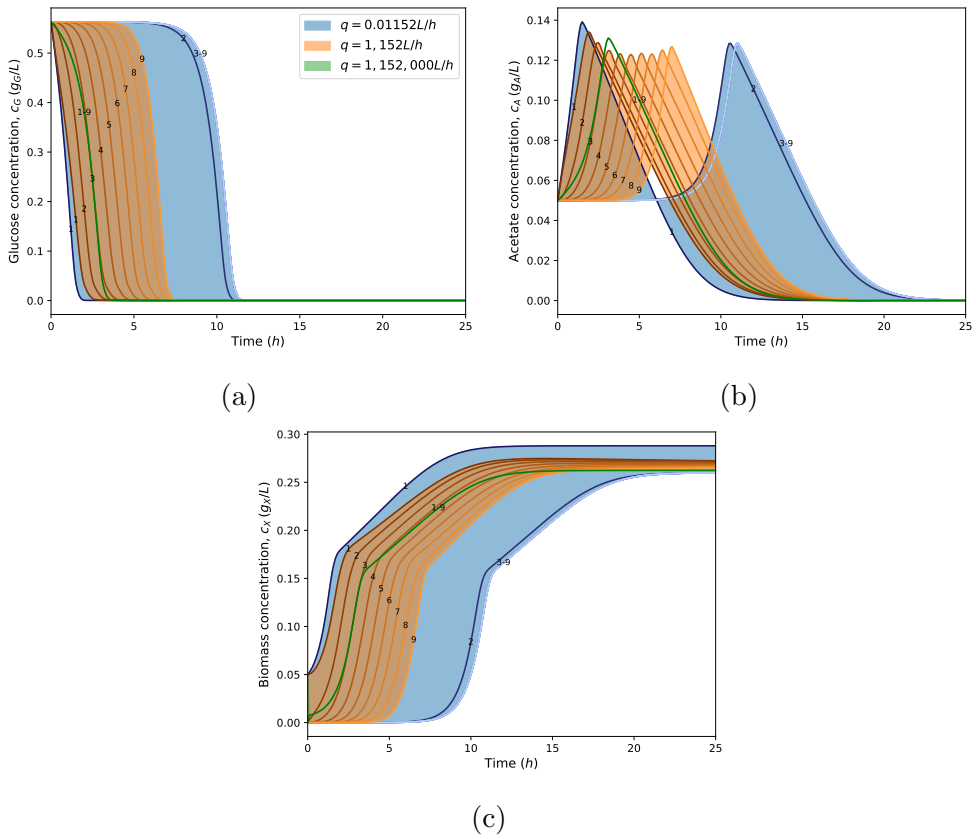
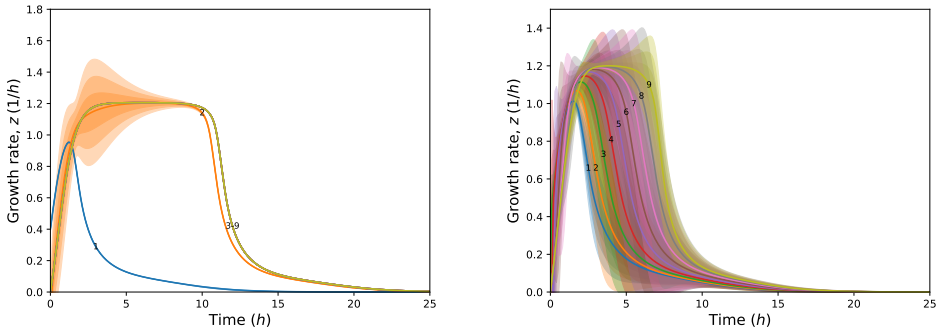
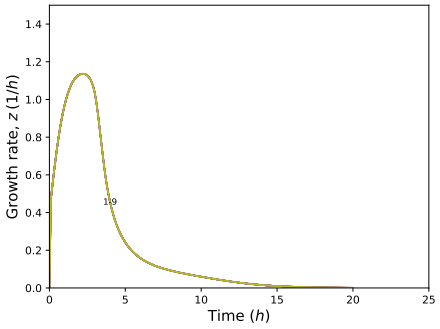


Figure 6.9: Comparison of simulated concentration profiles across all compartments for (a) glucose, (b) acetate, and (c) biomass (average concentration) at different flow rates. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments.



(a)

(b)



(c)

Figure 6.10: Comparison of simulation results of growth rate heterogeneity across various compartments when (a) $q = 0.01152$ L/h (b) $q = 1,152$ L/h (c) $q = 1,152,000$ L/h. The numbered tags with different colors indicate the compartments from which the solutions are taken, while the shaded bands represent the range of growth rate values—from the mean, μ_z, to three standard deviations, $3\sigma_z$.

In contrast, no heterogeneity emerges in Compartments 3-9 because there are no concentration gradients among these compartments (see Figures 6.9a and 6.9b). Additionally, no heterogeneity is observed in Compartment 1 because, unlike Compartment 2, it does not serve as a mixing zone where populations from multiple compartments converge. Instead, Compartment 1 only acts as an equilibration zone for the population flowing in from Compartment 2.

Lastly, when $\tilde{t}_{q,X,j}$ is shorter than $\tilde{t}_{\zeta,z,j}$ and gradients in extracellular concentrations are absent, heterogeneity effectively disappears, as shown in Figure 6.10c.

6.3.4 Validation and analysis of simulations of the large-scale bioreactor

To validate the combined model integrating the MM-TND and a CM, simulation results were compared with simulation conducted by combining the bin method and the CM described by Pigou et al. [18], and to experimental data [22] from a fed-batch *E. coli* culture conducted in a 30 m^3 stirred tank reactor with a working volume of 22 m^3.

During the experiment, dissolved oxygen levels were monitored at 1.12 meter and 4.05 meter above the bottom of the vessel and maintained above 20% air saturation throughout the fermentation. To simplify the model, the oxygen conservation equation in the extracellular space was omitted, and a constant value was assumed in the simulations.

The compartmental model was set up as described in Section 6.3.3. However, instead of fixing the flow rates, q, between adjacent compartments to constant values as done in the previous section, these were determined from CFD simulation results, as explained in [1]. This approach yielded a residence time for the overall system of 234.61 seconds, which is close to the 250 seconds reported by Pigou et al. [18], thus validating the use of the CM set-up in this study. MTEs (Equations (6.12) and (6.14)) and MBEs (Equations (6.13) and (6.14)), as described in Section 6.3.3, were incorporated into the CM.

Although the actual experimental system was operated in fed-batch mode with intermittent sampling, it was assumed for simplicity that the volume of the sample drawn for analysis counterbalanced the volume of glucose solution fed into the system, resulting in a constant working volume throughout the operation. The following glucose feed term was then added into the RHS of Equation (6.13) for Compartment 8 (the compartment where the feed port is located):

$$+\frac{1}{v_8} q_{ex\to 8} c_{G,ex\to 8} \tag{6.19}$$

The feed programs are presented in Table 6.4.

Table 6.4: Programs for the glucose feed flow rate, $q_{ex\to j}$ (in L/h), and the glucose feed concentration, $c_{G,ex\to j}$, for lab-scale (15 L) and large-scale (30 m^3) bioreactors.

Time	Operational mode	Lab-scale ($15L$)	Large-scale ($30m^3$)
$0 \leq t < L1$	Batch	0	0
$L1 < t \leq L2$	Fed-batch with exponential feed rate	$0.0057\exp\{0.27(t-L1)\}$	$15\exp\{0.3(t-L1)\}$
$L2 < t \leq L3$	Fed-batch with constant feed rate	0.064	180
$L3 < t$	Fed-batch with constant feed rate	0.057	170
Parameters	Units		
j	—	1	8
$c_{G,ex\to j}$	$g\ L^{-1}$	532	454
L_1	h	0.92	0.92
L_2	h	9.98	9.20
L_3	h	12.38	11.70

The rate of change of the internal state, ζ_j, was determined using Equation (6.7). The metabolic model used to define $z^*(\mathbf{c})$, $r_A(z,\mathbf{c})$, $r_G(z,\mathbf{c})$, and $r_X(z,\mathbf{c})$ was adapted from Pigou et al. [18], since the original model reported by Pigou was unable to fit the experimental data. In general, it is known that both glucose and oxygen levels promote growth, while acetate inhibits growth [127, 26]. Hence, it was initially hypothesized that growth rate may be expressed by the product of Monod type rates with respect to glucose and oxygen and an inhibition term with respect to acetate as per the following equation:

$$\mu_j^{G*} = \mu_{max}^{G} \frac{c_{G,j}}{K_G + c_{G,j}} \frac{c_{O,j}}{K_O + c_{O,j}} \frac{K_{i,A}}{K_{i,A} + c_{A,j}} \tag{6.20}$$

However, the data could not be explained using this equation. In particular, it was observed that there is a slow gradual accumulation of both glucose and acetate in the last 20-hour period of bioreactor operation, hinting at the accumulation of an unknown species that inhibits glucose and acetate consumption. For example, reactive oxidative species in

fermentations have been reported to accumulate towards the end of fed-batch fermentations and to inhibit growth [128, 129]. Accordingly, this research assumes that an unknown growth inhibitor accumulates during operation, thus explaining the slow accumulations of glucose and acetate at later stages of the fermentation. The mass balance of the inhibitor is expressed as:

$$\frac{dc_{I,j}}{dt} = \frac{1}{v_j}\left(\sum_{\substack{j'=1\\ j'\neq j}}^{\mathrm{n_{EC}}}(q_{j'\to j}c_{I,j'}) - c_{I,j}\sum_{\substack{j'=1\\ j'\neq j}}^{\mathrm{n_{EC}}} q_{j\to j'}\right) + c_{X,j}\int_0^{z_{max}} Y_{IX}r_{X,j}\tau_j d\mu \tag{6.21}$$

where Y_{IX} is the inhibition production rate and defined as the product of the growth rate and the yield coefficient.

It was further hypothesized that the effect of this inhibitor on glucose consumption is such that at very low inhibitor concentrations, the consumption rate adheres to a regular Monod kinetic. On the other hand, at high concentrations of the inhibitor, the consumption rate of glucose will be reduced according to a cooperative Hill kinetic rate involving both glucose and the inhibitor. Following these considerations, the optimal growth rate μ^{G*} was expressed as follows:

$$\mu_j^{G*} = \mu_{max}^{G}\frac{c_{G,j}}{K_G + c_{G,j}}\frac{c_{O,j}}{K_O + c_{O,j}}\frac{K_{i,A}}{K_{i,A} + c_{A,j}}\underbrace{\frac{c_{G,j}^{n_H}}{(k_I c_{I,j})^{n_H} + c_{G,j}^{n_H}}}_{\text{Hill term}} \tag{6.22}$$

In this expression, the association constant is modeled as being linearly proportional to the inhibition concentration, $c_{I,j}$. The Hill term will be approximately one for near-zero inhibitor concentration, i.e., $c_{I,j} \approx 0$, thus bringing the metabolic model back to the original form given by Equation (6.20) above. It should be emphasized that the new expression for the growth rate in Equation (6.22) is the only modification done to the original model reported by Pigou et al.

The original form of the model reported by Pigou et al. [18] was used to fit experimental data from both a lab-scale 15 L batch culture and an industrial-scale 30 m^3 fed-batch culture bioreactor. However, as reported by Pigou et al. [18] and Promma et al. [1], different sets of kinetic models were required to simulate the two scales. Using different kinetic parameters for different scales is not physically justified, since the parameters are supposed to represent inherent behavior of the cells, not of the reactor. Thus, a goal of the current study was to investigate whether a single metabolic model with the modified growth

rate in Equation (6.22) coupled to a CFD-based CM could fit the data at both scales. The parameters used in this model were calibrated using experimental concentration data from the lab-scale bioreactor, with the results shown in Table 6.2. The initial conditions are provided in Table 6.1. Simulation results for the lab-scale bioreactor (15 litres) with a single compartment ($n_{EC} = 1$) are shown in Figure 6.11, alongside the reported experimental results [22].

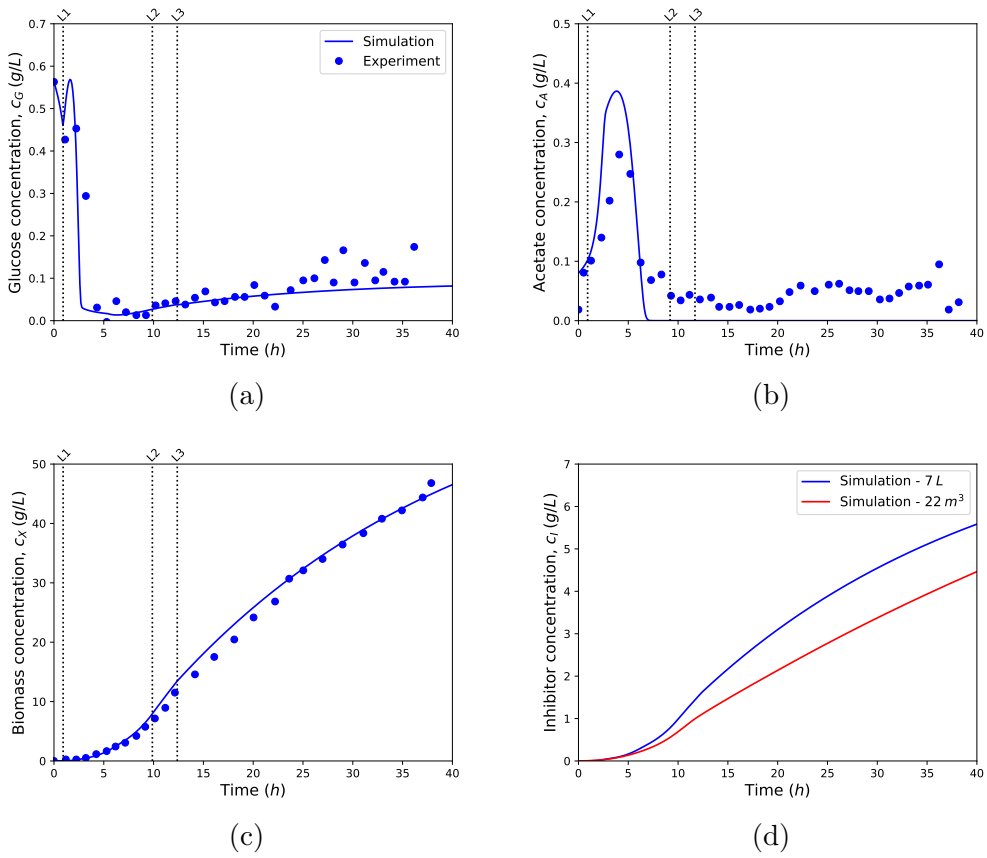


Figure 6.11: Comparison of the simulation results with the experimental data [22] for the concentration profiles of (a) glucose, (b) acetate, (c) biomass, and (d) inhibitor (only the simulation result presented) in a lab-scale reactor.

The model successfully captured the gradual increase in glucose concentration during the fed-batch phase with a constant feed rate ($t > L2$).

The growth rate profile in the bioreactor is shown in Figure 6.12. It was observed that

no heterogeneity formed within the population, and the mean growth rate approached the optimal growth rate, z^*. The optimal growth rate acted as an attractor for the mean growth rate in this system.

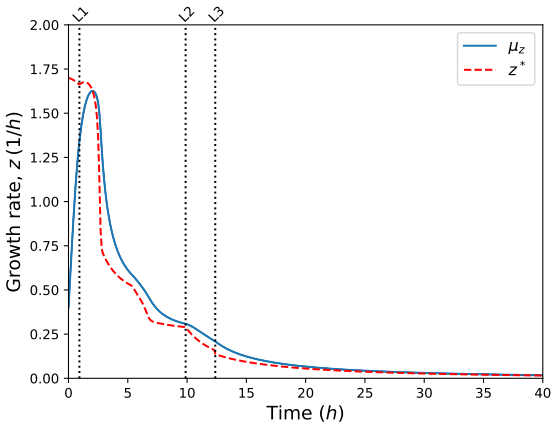


Figure 6.12: Growth rate profile in a lab-scale reactor. At each time point, the mean growth rate, μ_z, approaches the optimal growth rate, z^*, set by the external substance concentration.

Simulation results for the large-scale bioreactor (30 m^3), with $n_{EC} = 9$, are shown in Figures 6.13 and 6.14, alongside the reported experimental results [22] and simulations of the same system from other studies: Promma et al. [1] and Pigou et al. [18] at the top, middle, and bottom measuring probes (see Figure 6.7).

The relative squared errors comparing the simulations with the reported experimental results are shown in Table 6.6. The results predict the trends of glucose, acetate, and biomass concentrations well, especially for glucose at the middle and bottom probes (Figures 6.14c and 6.14e). It is evident that the glucose simulation results from this study provide a better prediction compared to the other simulations. To quantitatively compare the performance of the models, the corrected Akaike information criterion (AIC_C) was applied to the normalized data using the method described by Portet [122]. The results, presented in Table 6.5, indicate the superior performance of the proposed model, as evidenced by its lower AIC_C value. Promma et al. [1] used fewer parameters than the other models by excluding stoichiometric coefficients derived from pathway analysis and oxygen mass transfer parameters obtained from CFD simulations, arguing that these values were based on reported experimental measurements. However, since their study yielded

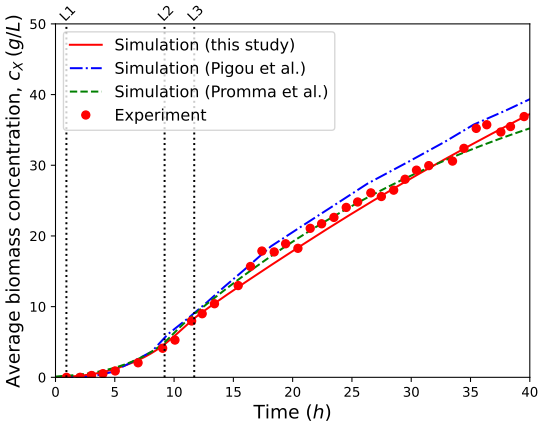


Figure 6.13: Comparison of simulation results with experimental data [22] for the average biomass concentration in a bioreactor with a working volume of 22 m^3.

significantly higher squared errors compared to this study, the reduction in the number of parameters has a less pronounced effect on AIC_C.

Table 6.5: Comparison of AICc values for different models

Model	Number of Parameters	AICc
This study	23	-790.38
Pigou et al.[18]	22	-207.00
Promma et al.[1]	8	-582.00

In the study by Promma et al. [1], it was concluded that the formation of acetate from 5 hours to 10 hours was due to oxygen shortage and the control action. However, in this study, where the oxygen concentration is maintained at excess levels, acetate production is still observed. This suggests that acetate formation is not solely governed by oxygen limitation, but may also be influenced by other factors. This raises questions about the accuracy of the models, which need to be addressed.

The oscillations of the glucose concentration observed at the top probe of the tank is believed to be caused by turbulence near the feed port [22].

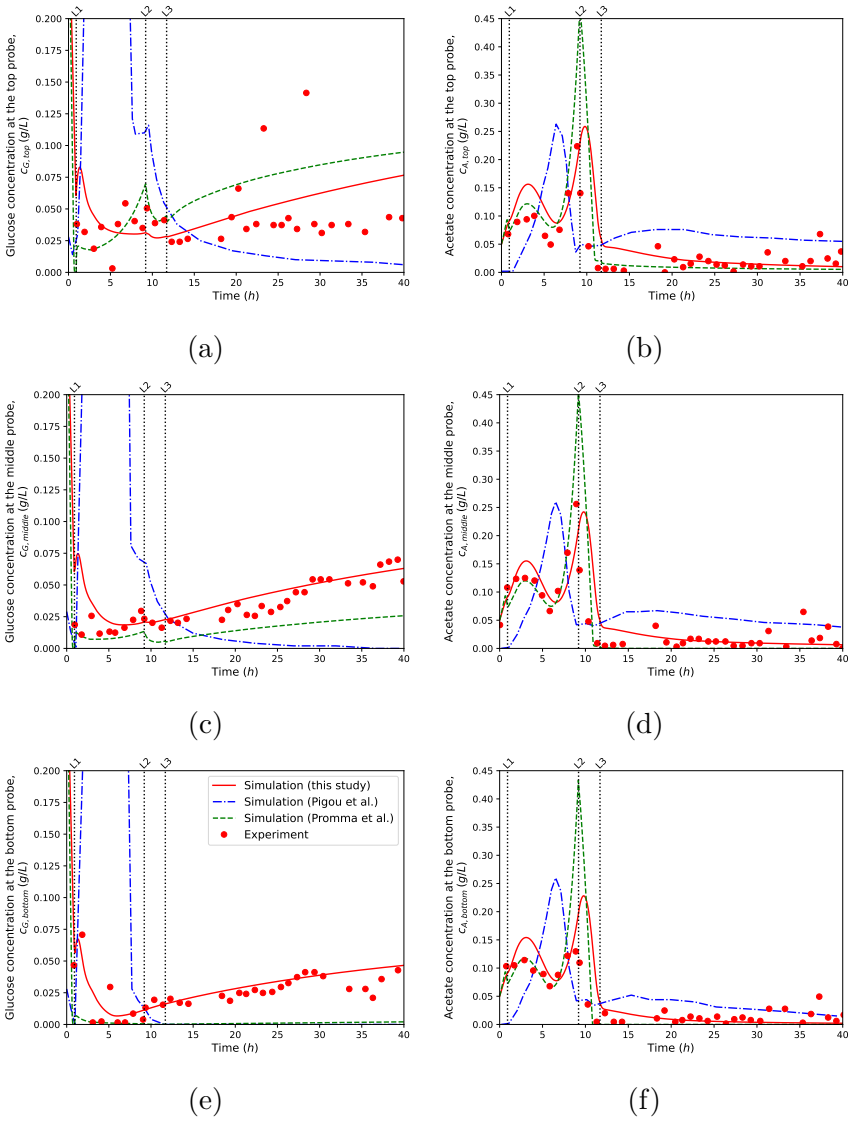


Figure 6.14: Comparison of simulation results with experimental data [22] for concentration profiles in a bioreactor with a working volume of 22 m^3. Glucose profiles are shown in (a, c, e), while acetate profiles are shown in (b, d, f) at three probe locations: (a, b) top probe at 6.48 m, (c, d) middle probe at 4.04 m, and (e, f) bottom probe at 1.12 m from the bottom of the reactor.

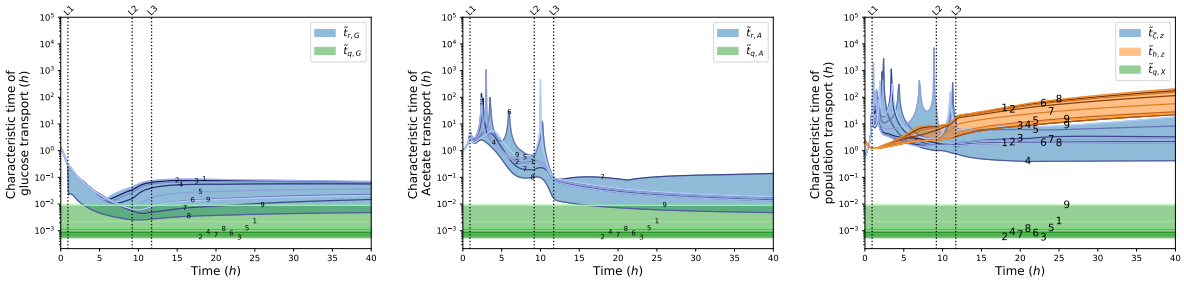
Table 6.6: Root mean squared error (RMSE) in g/L for Glucose, Acetate, and Biomass

	This Study	Pigou et al.[18]	Promma et al.[1]
Glucose (Bottom Probe)	0.0107	0.0950	0.0276
Glucose (Middle Probe)	0.0136	0.1028	0.0217
Glucose (Top Probe)	0.0500	0.1115	0.0497
Acetate (Bottom Probe)	0.0395	0.0537	0.0670
Acetate (Middle Probe)	0.0397	0.0689	0.0594
Acetate (Top Probe)	0.0468	0.0705	0.0700
Biomass	1.0143	1.6926	0.9805

Concentration gradients in glucose and acetate observed in this simulation (see dashed lines in Figure 6.17, where the simulation results are plotted together). The flow rate used produces a $\tilde{t}_{q,i,j}$ of the same magnitude as in the case of $q = 1,152,000$ L/h discussed in the previous section, where no gradients were observed. However, due to the presence of external feed, gradients still develop between locations close and far from the feeding port. In this simulation, glucose feed was accounted for via $\tilde{t}_{r,G,8}$. The volume introduced at the feed site (Compartment 8) was not considered; instead, the glucose mass itself was treated as the source term introduced into the system. The expression of $\tilde{t}_{r,G,8}$ is then modified to:

$$\tilde{t}_{r,G,8} = \left| \frac{c_{G,8}}{r_{G,8} + \frac{q_{ex\to 8} c_{G,ex\to 8}}{v_8}} \right| \tag{6.23}$$

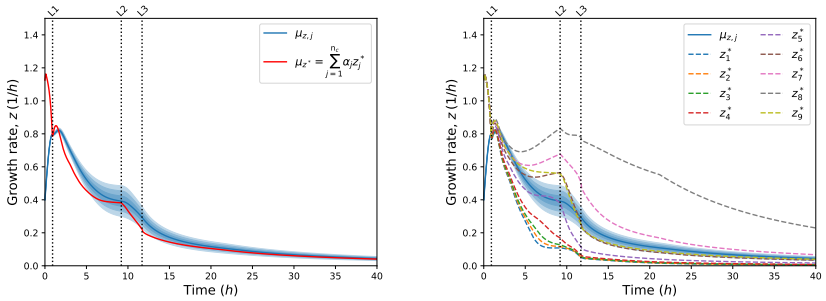
The results shown in Figures 6.15a and 6.15b confirm that glucose gradients are expected to form around 2 hours, while acetate gradients are anticipated to develop around 12 hours. This corresponds to the range where the reaction time, $\tilde{t}_{r,i,j}$ (including the glucose feed time), is on the same order of magnitude as the advection time, $\tilde{t}_{q,i,j}$.



(a) Glucose (b) Acetate (c) Cell biomass

Figure 6.15: Characteristic time profiles for various phenomena in the fed-batch bioreactor simulation with active volume of $22\,\mathrm{m}^3$. The results are shown for (a) glucose, (b) acetate, and (c) cell population. The number tag indicates the compartment where the solution is taken, while the band provides the range of values across all compartments.

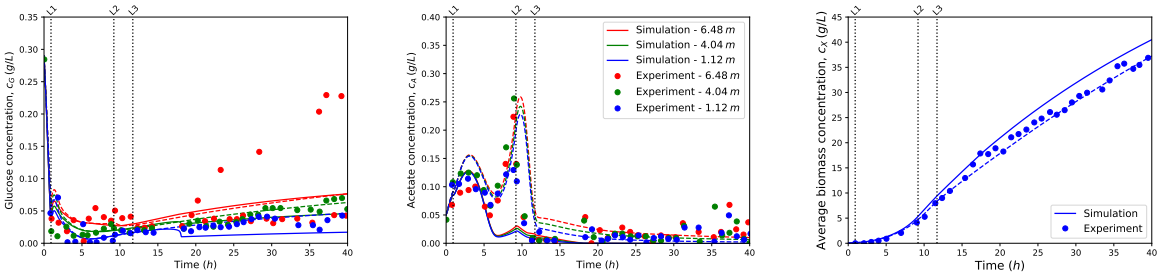
In addition, the characteristic time analysis for the cell population, shown in Figure 6.15c, aligns with the observation that the growth rate profiles across all compartments, as illustrated in Figure 6.16, are similar to each other. This is because $\tilde{t}_{q,X,j}$ is significantly shorter than both $\tilde{t}_{\zeta,z,j}$ and $\tilde{t}_{h,j}$. Specifically, the cells transit through the compartments too quickly for substantial changes to occur in their internal state or for processes such as growth or death to take place. Additionally, the mean growth rates, $\mu_{z,j}$, primarily follow the volume-averaged optimal growth rate, μ_z^* (see Figure 6.16a), while the different optimal growth rates, z_j^*, determined by the external substrate concentrations, act as external perturbations to the cells (see Figure 6.16b), thereby creating internal heterogeneity.



(a) (b)

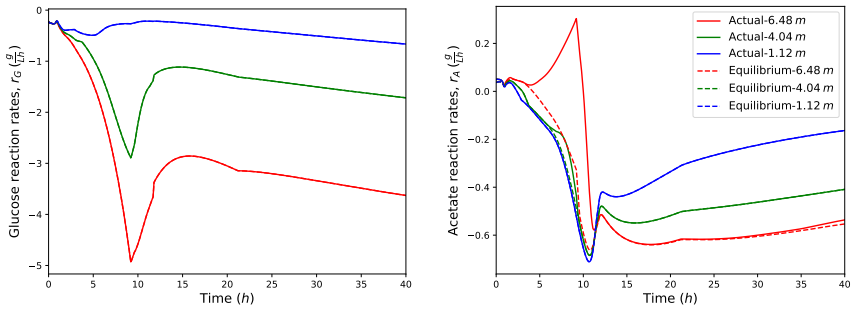
Figure 6.16: Growth rate profile in a bioreactor with a working volume of 22 m^3. At each time point, the mean growth rate, μ_z, approaches (a) the volume-averaged optimal growth rate, μ_{z^*}, of (b) the optimal rates, z_j^*, set by the external substance concentration in different compartments.

Lastly, to explore the impact of internal heterogeneity on the system, the large-scale bioreactor was simulated under the assumption of a homogeneous population, where the cellular states were fixed at their optimal rates corresponding to the external substrate concentrations (with PBE or MTEs deactivated). The resulting concentration profiles of glucose, acetate, and biomass are compared with those from the previous case, which accounted for internal heterogeneity, as shown in Figure 6.17. While the concentration profiles of glucose, acetate, and biomass for the lab-scale bioreactor remain identical to Figure 6.11 in both cases, the profiles for the large-scale bioreactor differ, particularly for acetate. Acetate was consumed more in the case of a homogeneous population, resulting in a higher biomass concentration for this case. In fact, by comparing the actual consumption/production rates of acetate for the case of the heterogeneous population with the optimal rates determined by the external substrate concentration, as shown in Figure 6.18, it was observed that the actual production rate of acetate occurs at the top part of the tank, rather than consumption, which would be expected if all cells were homogeneous and corresponding to the external concentration at those heights. Hence, modeling intracellular heterogeneity turned out to be a way to capture metabolic response that cannot be captured by only modeling the extracellular environment.



(a) (b) (c)

Figure 6.17: Comparison of simulation results for concentration profiles of (a) glucose, (b) acetate, and (c) biomass (average concentration) at different heights in a bioreactor with a working volume of 22 m^3: without the population balance equation (solid line), with the population balance equation (dashed line), and experimental data [22].



(a) (b)

Figure 6.18: Comparison between the actual consumption/production rates (Actual) and the optimal rates determined by the external substance concentrations (Optimal) for (a) glucose and (b) acetate in a bioreactor with a working volume of 22 m^3.

6.4 Conclusion

This study presents an innovative approach to modeling intracellular heterogeneity and concentration gradients in large-scale bioreactors using the MM-TND method. It effectively captures the dynamics of microbial populations under varying conditions, overcoming the limitations of traditional bin methods, such as their inability to accurately represent het-

erogeneity at finer scales and the high computational demands associated with increasing resolution.

While the MM-TND has shown great promise, future studies could explore the use of quadrature-based methods of moments (QBMMs), with modifications to account for constraints and for discontinuous reaction rates along the state space.

Through validation against experimental data and comparisons with existing methods, the proposed framework demonstrates its robustness and versatility. The findings highlight the significance of capturing both extracellular gradients and intracellular heterogeneity in large-scale bioprocesses, providing critical insights into microbial behavior and system performance.

Chapter 7

Development of an APCS to Simulate a Coupled PBM-CM

7.1 Introduction

Simulating large-scale bioreactors by coupling population balance models with compartmental models requires computing an intracellular concentration distribution for each extracellular compartment [18]. As the number of compartments increases, so does the computational load, posing significant challenges in terms of efficiency.

To address this, a simplification can be introduced, particularly in systems where stirring dominates the hydrodynamics of all phases, resulting in convection being the primary mode of microorganism transport. Under these conditions, the characteristic timescale of microbial advection is shorter than that of microbial adaptation. Observations under such conditions, as shown in Figures 6.10c and 6.16, suggest that a single distribution might, in principle, be sufficient to represent microbial heterogeneity.

However, simulation results presented in this chapter indicate that this simplification leads to a complete loss of internal heterogeneity, with the standard deviation σ approaching zero. In contrast, when compartments are treated separately, internal heterogeneity remains clearly observable.

This outcome highlights a key limitation: within a merged compartment, microorganisms are exposed only to the mean external concentration, rather than the underlying concentration gradients that drive their physiological responses.

To address this issue, an adaptive population compartmental scheme (APCS) is proposed. The APCS dynamically adjusts compartmentalization during simulations, ensuring that the predictive accuracy of microbial heterogeneity is maintained while optimizing computational efficiency. Unlike conventional compartmental models that rely on fixed spatial compartmentalization, the APCS introduces a flexible framework that adapts compartment networks and resolutions in response to real-time system conditions, such as concentration gradients and variations in microbial states. This approach mitigates the trade-off between computational cost and model fidelity, enabling more accurate yet efficient simulations of large-scale bioreactors.

In summary, this study aims to develop the APCS that dynamically refines compartmentalization during simulations. By balancing computational efficiency with predictive accuracy, the APCS provides a robust framework for capturing both spatial and physiological heterogeneity in bioprocess modeling.

7.2 Methodology

Developing the APCS involves two key steps: (1) designing a computationally efficient compartmental map that accurately captures microbial population dynamics while minimizing the number of tracked populations and computational complexity, and (2) incorporating an adaptive mechanism that dynamically adjusts compartmentalization based on the map from step (1) during simulations to enhance efficiency and accuracy.

7.2.1 Development of a computationally efficient compartmental map for the cell population

Solving one intracellular distribution for each compartment may be computationally expensive. Hence, there is a strong motivation to reduce the number of distributions needed down to the minimum needed to provide a representative status of the intracellular states. Various characteristic times can be derived and used for this task, as summarized in Table 6.3. It follows that if the mean characteristic time associated with the rate of change of the internal state of cells, $\tilde{t}_{\zeta,z_k,j}$ (Equation (6.17)), across all cell states z_k, and the net rate of cell generation, $\tilde{t}_{h,z_k,j}$ (Equation (6.18)), are significantly greater than the characteristic advection (residence) time of Compartment j, $\tilde{t}_{q,X,j}$ (Equation (6.15)), it can be inferred that cells transit through the compartment too quickly for significant changes in their internal state or for processes such as growth or death to occur (see).

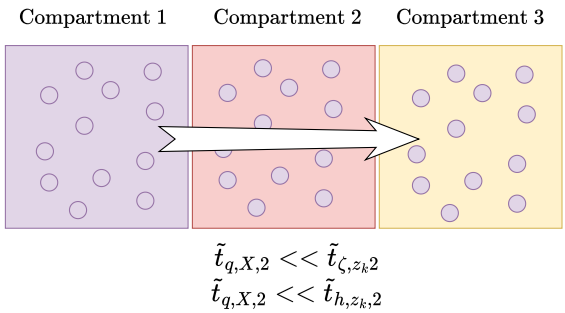


Figure 7.1: Microorganisms from Compartment 1 rapidly transit through Compartment 2 to Compartment 3 without sufficient time to adapt their internal state.

This suggests the possibility of merging Compartment 2 with Compartment 1, as a single number density function, n_j, may be sufficient to describe the microorganisms in both compartments. This idea can be extended to develop an additional compartmental map for modeling the transport of only the microorganism population, i.e. number density function, n_j. By combining multiple compartments, the number of n_j, can be reduced, leading to a decrease in computational complexity.

To implement this merging strategy of intracellular distributions while maintaining the compartmentalization of the extracellular environment that was determined based on CFD, two compartmental maps will now be employed:

- **ExtComp:** map for the transportation of external species, based on the CFD-related compartmentalization criteria.
- **PopComp:** map that describes the transportation of the number density function, n_ι, of microorganism populations

Let assume that PopComp is manually defined by merging the compartments of ExtComp that are assumed to satisfy the aforementioned characteristic time criteria, i.e. $\tilde{t}_{\zeta,z_k,j} \gg \tilde{t}_{q,X,j}$ and $\tilde{t}_{h,z_k,j} \gg \tilde{t}_{q,X,j}$, resulting in a total of $\mathrm{n_{PC}}$ compartments. Each Compartment j in ExtComp is associated with a specific Compartment ι in PopComp, and the set of Compartments j in ExtComp belonging to Compartment ι in PopComp is denoted as $\mathbf{l}_\iota$. The total volume of Compartment ι is given by $v_\iota = \sum_{j \in \mathbf{l}_\iota} v_j$, while the volume fraction of each sub-compartment j within ι is defined as $\alpha_j = v_j / v_\iota$.

The PBE for Compartment, ι, can be derived by summing Equation (2.34) across Compartment $j \in \mathbf{l}_\iota$. Assuming that the number density function remains uniform or is

averaged across $\mathbf{l}_\iota$, denoted as $n_j = n_\iota$ for all $j \in \mathbf{l}_\iota$, and considering that the advection terms between the sub-compartments within $\mathbf{l}_\iota$ cancel out, the resulting balance equation will only include the convection terms into or out of Compartment ι from/to other compartments, denoted by ι', in PopComp. By rearranging and grouping the relevant variables, the final expression is obtained as:

$$\begin{aligned} v_\iota \frac{\partial n_\iota}{\partial t} = &-v_\iota \nabla^z \cdot n_\iota \sum_{j \in \mathbf{l}_\iota} \alpha_j \boldsymbol{\zeta}_j + v_\iota \sum_{j \in \mathbf{l}_\iota} (\alpha_j h_j) \\ &+ \sum_{\substack{\iota'=1 \\ \iota' \neq \iota}}^{\mathrm{n_{PC}}} q_{\iota' \to \iota} n_{\iota'} - n_\iota \sum_{\substack{\iota'=1 \\ \iota' \neq \iota}}^{\mathrm{n_{PC}}} q_{\iota \to \iota'} \end{aligned} \tag{7.1}$$

The corresponding MTEs for this PBE are expressed as:

$$\begin{aligned} v_\iota \frac{dm_{\boldsymbol{\beta},\iota}}{dt} = &\, v_\iota \int_{\Omega^z} \nabla^z g \cdot n_\iota \sum_{j \in \mathbf{l}_\iota} \alpha_j \boldsymbol{\zeta}_j + g \sum_{j \in \mathbf{l}_\iota} (\alpha_j h_j) d\mathbf{z} \\ &+ \sum_{\substack{\iota'=1 \\ \iota' \neq \iota}}^{\mathrm{n_{PC}}} q_{\iota' \to \iota} m_{\iota', \boldsymbol{\beta}} - m_{\boldsymbol{\beta},\iota} \sum_{\substack{\iota'=1 \\ \iota' \neq \iota}}^{\mathrm{n_{PC}}} q_{\iota \to \iota'} \end{aligned} \tag{7.2}$$

The volumetric flow rate between compartments in PopComp can be derived from the known volumetric flow rates between compartments in ExtComp as follows. Let $\mathbf{Q}_{ext}$ denote the volumetric flow rate matrix (see Sections 2.2.2 and 3.2.2), where each component $q_{a \to b}$ represents the volumetric flow rate from Compartment a to Compartment b in ExtComp. The mapping between ExtComp onto PopComp compartments is described by popComp_list $= [\mathbf{l}_1, \cdots, \mathbf{l}_{\mathrm{n_{PC}}}]$, a list of lists. Each inner list $\mathbf{l}_\iota$ contains the "sorted" indices of ExtComp compartments that collectively form Compartment ι in PopComp. The algorithm to find $\mathbf{Q}_{pop}$, the volumetric flow rate matrix of PopComp, is given in Algorithm 2

Algorithm 2 Construction of $\mathbf{Q}_{pop}$ from $\mathbf{Q}_{ext}$ and popComp_list.

1: **Input:** popComp_list, $\mathbf{Q}_{ext}$
2: **Output:** $\mathbf{Q}_{pop}$
3: **for** each compartment group $\mathbf{l}_\iota$ containing multiple compartments **do**
4: Sum the flow rate values of all rows corresponding to compartments in $\mathbf{l}_\iota$ and assign the result to the row corresponding to the representative compartment (the first compartment in the group):
$\mathbf{Q}_{ext}[\mathbf{l}_\iota[1], :] \leftarrow \sum_{j \in \mathbf{l}_\iota} \mathbf{Q}_{ext}[\mathbf{l}_\iota[j], :]$
5: Sum the flow rate values of all columns corresponding to compartments in $\mathbf{l}_\iota$ and assign the result to the column corresponding to the representative compartment (the first compartment in the group): $\mathbf{Q}_{ext}[:, \mathbf{l}_\iota[1]] \leftarrow \sum_{j \in \mathbf{l}_\iota} \mathbf{Q}_{ext}[:, \mathbf{l}_\iota[j]]$
6: **end for**
7: Remove all rows and columns from $\mathbf{Q}_{ext}$ except those corresponding to the representative compartments, then assign the resulting submatrix to $\mathbf{Q}_{pop}$:
$\mathbf{Q}_{pop} \leftarrow \mathbf{Q}_{ext}[\mathbf{l}_\iota, \mathbf{l}_\iota]$
8: **return** $\mathbf{Q}_{pop}$

Note that each component $q_{a \to b}$ of $\mathbf{Q}_{pop}$ represents the volumetric flow rate from Compartment a to Compartment b in PopComp

7.2.2 Development of the APCS

It is shown in Section 7.3 that when the merged compartmental map of PopComp is applied, particularly when $n_{PC} = 1$, the simulation results indicate a complete loss of internal heterogeneity, as measured by the magnitude of σ, i.e. σ tends to zero. In contrast, internal heterogeneity is clearly observable when merging is not done, i.e. $n_{PC} = n_{EC}$. The reasons for this loss are discussed in detail in Section 7.3. Briefly, the merged compartments of PopComp eliminate the influence of concentration gradients on microorganisms. Within a PopComp compartment, microorganisms perceive only the mean concentration, rather than the detailed gradients represented by its constituent ExtComp compartments.

To mitigate the loss of internal heterogeneity when applying PopComp, the Adaptive PopComp Scheme (APCS) was developed to balance accuracy and computational efficiency. APCS automates the assignment of PopComp, allowing it to adapt dynamically throughout the simulation. In contrast, without APCS, PopComp must be manually defined at the beginning of the simulation and remains fixed thereafter. In this study, PopComp was automatically constructed using an attractor variable derived from the concentration gradients. An attractor, in this context, refers to an equilibrium internal state toward which the system naturally evolves over time. Because it is determined by environmental condi-

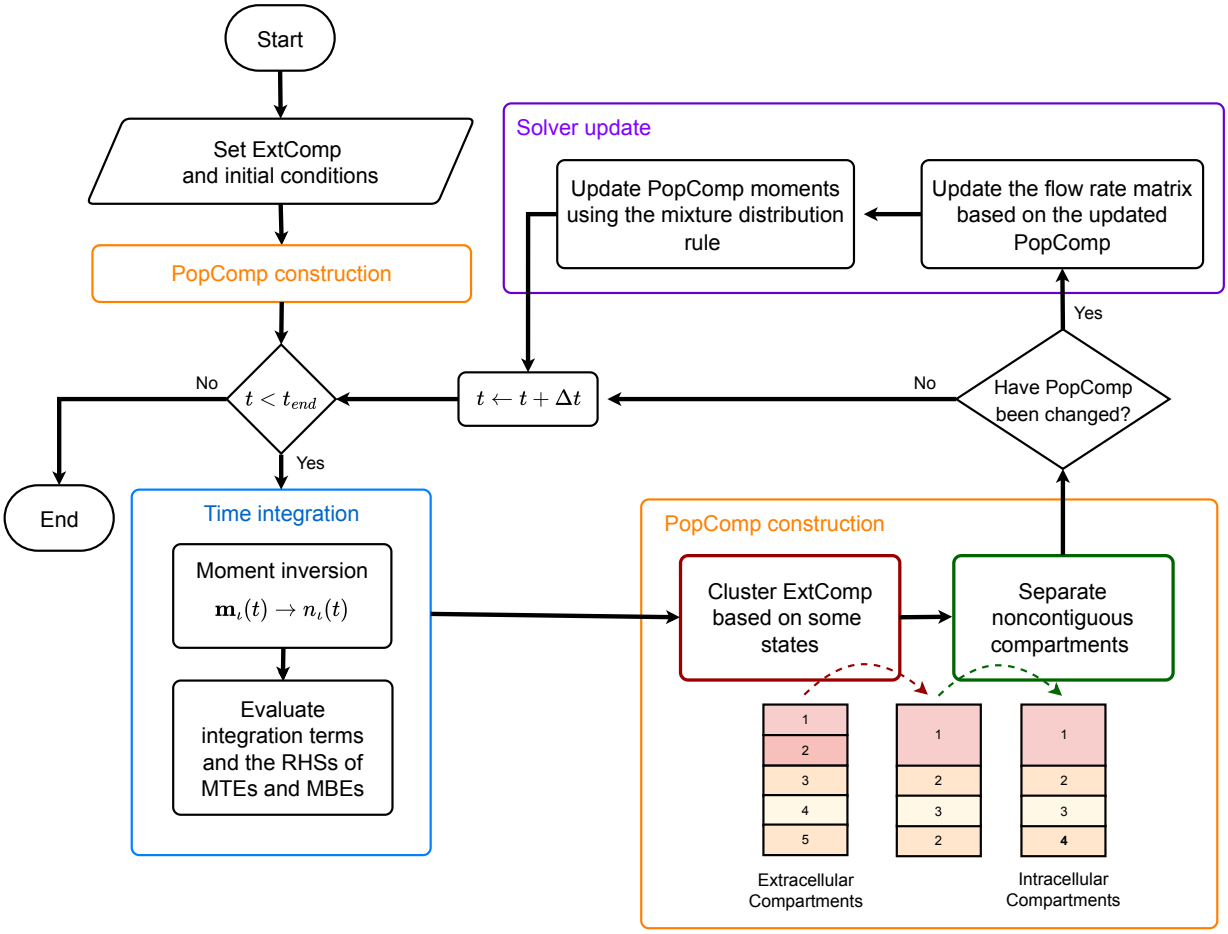


Figure 7.2: Diagram of the adaptive population compartmental scheme (APCS).

tions, the attractor may vary dynamically, continuously guiding the simulation trajectory toward itself.

As long as the adaptation rate is nonzero, the presence of gradients can perturb the population, promoting the emergence of intracellular heterogeneity.

Figure 7.2 shows the diagram of the APCS. In addition to the time-step solver for the adaptive time-step IVP solver, two processes were incorporated into the development of the APCS:

1. **PopComp Construction:** A PopComp map is generated based on the concentration gradients of extracellular species. If the map is derived from the gradient of a single species, only the following sub-steps are required:

   (a) **ExtComp clustering:** The concentration of substance i across different ExtComp compartments is represented in the data set $\mathbf{c}_i = [c_{i,1}, \ldots, c_{i,\mathrm{n_{EC}}}]$. The parameter ϵ_d represents the acceptable level of homogeneity and is used to group neighboring data points when the difference between them is smaller than ϵ_d. The PopComp is constructed by grouping ExtComp compartments according to the procedure outlined in Algorithm 3.

   The algorithm performs hierarchical clustering on compartments used on $\mathbf{c}_i$, aiming to group similar values based on the acceptable level of homogeneity, ϵ_d, and the averaged distance between the means of the clusters, μ_d. It begins by sorting $\mathbf{c}_i$ and their corresponding compartments, forming initial clusters based on ϵ_d, where neighboring data points, $\mathbf{c}_i$, are grouped together if their difference is smaller than the threshold. This initial step ensures that closely related values are grouped early in the process, even before further refinements are applied.

   Next, the algorithm calculates the average distance between the means of these preliminary clusters , μ_d. If this average distance is exceeded for a particular cluster, the algorithm recursively partitions this cluster into smaller subgroups (as of Algorithm 4). This is done by identifying the largest gap between sorted values within the cluster and splitting it at that point. The process continues iteratively until all clusters either satisfy the average distance criterion or are too small to divide further.

   Finally, the algorithm outputs popComp_list, a refined PopComp map for the system. This map captures the key structural features of $\mathbf{c}_i$, effectively providing an approximate representation of the primary patterns and relationships within the data. By focusing on local similarities and addressing larger gaps, the

approach ensures that only the main features of $\mathbf{c}_i$ are preserved, resulting in an approximate but meaningful representation of the data.

(b) **Noncontiguous-compartment separation:** The PopComp map generated from PopComp Construction does not guarantee that compartments will be contiguous, as multi-region clusters are typically formed, necessitating further division. However, by examining the connectivity of the compartments (i.e. via $\mathbf{Q}_{ext}$), spatially contiguous compartments can be created. This method is detailed in Tajsoleiman et al. [59].

PopComp Construction, as presented above, is based on the assumption of an attractor variable derived from the concentration gradients. If the system exhibits multiple attractors, the method remains effective; however, a technique known as Compartment Overlap is required to generate a single PopComp map. This method was used by Tajsoleiman et al. [59] for hydrodynamic compartmentalization.

2. **Solver update:** This step is required when the PopComp map (i.e., popComp_list) has changed from the previous time step, reflecting the change in the relationship between ExtComp and PopComp (i.e., the external concentrations and considered populations). The adaptive time-step IVP solver has to adjust the number of ODEs based on the number of PopComp compartments. Specifically, the flow rate between the new PopComp compartments and $\mathbf{m}_\iota$ for these compartments needs to be addressed.

   Since popComp_list and $\mathbf{Q}_{ext}$ are known, Algorithm 2 can be used to determine the flow rate between the new PopComp compartments, $\mathbf{Q}'_{pop}$. Here, $\mathbf{l}'_\iota$ denotes the set of Compartments j belonging to Compartment ι in the new PopComp map. The new set of moments, $\mathbf{m}'_\iota$, can be obtained by projecting the $\mathbf{m}_\iota$ from the old PopComp map back onto the ExtComp map, i.e., $\mathbf{m}_j = \mathbf{m}_\iota$ where $j \in \mathbf{l}_\iota$. Then, using the mixture distribution rule:

$$\mathbf{m}'_\iota = \sum_{j \in \mathbf{l}'_\iota} \alpha'_j \mathbf{m}_j \tag{7.3}$$

   where α'_j is the volume fraction of Compartment j that belongs in Compartment ι, i.e. $\alpha'_j = v_j / \sum_{j \in \mathbf{l}'_\iota} v_j$.

   Now that $\mathbf{Q}_{ext}$, $\mathbf{c}_j$, $\mathbf{Q}'_{pop}$, and $\mathbf{m}'_\iota$ are available, the new time step for the IVP solver can be executed.

This procedure is repeated until the final simulation time is reached.

Algorithm 3 Hierarchical Clustering Algorithm

Input: $\mathbf{c}_i$, extComp_list, ϵ_c
Output: popComp_list
Section 1: Initialization of lists and variables
Initialize empty lists:
initial_value_clusters $\leftarrow$ [ ]
initial_comp_clusters $\leftarrow$ [ ]
Initialize variables:
current_cluster_values $\leftarrow$ [$c_i[1]$]
current_cluster_comps $\leftarrow$ [extComp_list[1]]
Section 2: Grouping consecutive values based on ϵ_c
for $k = 2$ **to** len($\mathbf{c}_i$) **do**
  if $c_i[k] - c_i[k-1] < \epsilon_c$ **then**
    Add $c_i[k]$ and extComp_list[k] to current_cluster_values and current_cluster_comps
  else
    Append current_cluster_values and current_cluster_comps to initial_value_clusters and initial_comp_clusters
    Reset:
    current_cluster_values $\leftarrow$ [$c_i[k]$]
    current_cluster_comps $\leftarrow$ [extComp_list[k]]
  end if
end for
Append the last current_cluster_values and current_cluster_comps to initial_value_clusters and initial_comp_clusters
Section 3: Compute cluster averages
Calculate the mean of each initial cluster:
cluster_averages $\leftarrow$ [mean(cluster) | cluster $\in$ **initial_value_clusters**]
Compute the average range between consecutive pairs of cluster_averages:
if len(cluster_averages) > 1 **then**
  $\mu_d \leftarrow$ mean([d_k | d_k = cluster_averages[k] $-$ cluster_averages[$k-1$] for $k = 2, \cdots,$ len(cluster_averages)])
else
  $\mu_d \leftarrow 0$
end if
Section 4: Refine clusters based on μ_d and ϵ_c
Initialize empty list:
popComp_list $\leftarrow$ []
for each value cluster $\mathbf{v}$, compartment group $\mathbf{l}$ in initial_value_clusters, initial_comp_clusters **do**
  if len($\mathbf{v}$) > 1 **and** $\mathbf{v}$[end] $- \mathbf{v}[1] > \mu_d$ **then**
    Partition $\mathbf{v}$ and $\mathbf{l}$ using μ_d:
    (par_values, par_comps)
      $\leftarrow$ cluster_partition($\mathbf{v}, \mathbf{l}, \mu_d, \epsilon_c$) {See Algorithm 4}
    Append par_comps to popComp_list
  else
    Append $\mathbf{l}$ to popComp_list
  end if
end for

return popComp_list

Algorithm 4 cluster_partition: Recursive Partitioning of Clusters

Input: $\mathbf{v}$, $\mathbf{l}$, μ_d, ϵ_c
Output: par_values, par_comps
if $\mathbf{v}[\text{end}] - \mathbf{v}[1] \leq \mu_d$ or $\mathbf{v}[\text{end}] - \mathbf{v}[1] < \epsilon_c$ **then**
  return $[\mathbf{v}]$, $[\mathbf{l}]$
end if
Compute pairwise distances:
$\mathbf{d} \leftarrow [d_k \mid d_k = \mathbf{v}[k] - \mathbf{v}[k-1] \text{ for } k = [2, \cdots, \text{len}(\mathbf{v})]$
Find the index of the largest gap:
$\text{max_dist_idx} \leftarrow \arg\max(\mathbf{d})$
Split the values and compartments into left and right groups:
$\mathbf{v}_{\text{left}} \leftarrow \mathbf{v}[1 : \text{max_dist_idx}]$, $\mathbf{v}_{\text{right}} \leftarrow \mathbf{v}[\text{max_dist_idx} + 1 : \text{end}]$
$\mathbf{l}_{\text{left}} \leftarrow \mathbf{l}[1 : \text{max_dist_idx}]$, $\mathbf{l}_{\text{right}} \leftarrow \mathbf{l}[\text{max_dist_idx} + 1 : \text{end}]$
Recursively partition the left and right groups:
$(\text{par_values}_{\text{left}}, \text{par_comps}_{\text{left}}) \leftarrow \text{cluster_partition}(\mathbf{v}_{\text{left}}, \mathbf{l}_{\text{left}}, \mu_d)$
$(\text{par_values}_{\text{right}}, \text{par_comps}_{\text{right}}) \leftarrow \text{cluster_partition}(\mathbf{v}_{\text{right}}, \mathbf{l}_{\text{right}}, \mu_d)$
Combine the results from the left and right partitions:
$\text{par_values} \leftarrow \text{par_values}_{\text{left}} + \text{par_values}_{\text{right}}$
$\text{par_comps} \leftarrow \text{par_comps}_{\text{left}} + \text{par_comps}_{\text{right}}$
return par_values, par_comps

All simulations were performed on a laptop configured with Ubuntu 64-bit, a 12$^{\text{th}}$ Gen Intel® Core™ i7-1260P processor (12 cores and 16 threads, operating between 400 MHz and 4.7 GHz), 16 GB of memory, a 512 GB SSD, and Intel® Iris® Xe integrated graphics.

7.3 Results and discussion

As observed in Sections 6.3.3 and 6.3.4, when the characteristic advection timescale, $\tilde{t}_{q,X,j}$, is shorter than the population adaptation timescale, $\tilde{t}_{\zeta,z,j}$, across all compartments, a single population ($n_{PC} = 1$) should be sufficient to represent the entire culture within the system. Then, for the scenario simulated as per the conditions in Section 6.3.4, we conducted a new simulation with two different transportation maps: one for external species (ExtComp) with $n_{EC} = 9$, and another specifically for microorganism populations (PopComp) with $n_{PC} = 1$, i.e. a single intracellular distribution is assumed for the entire

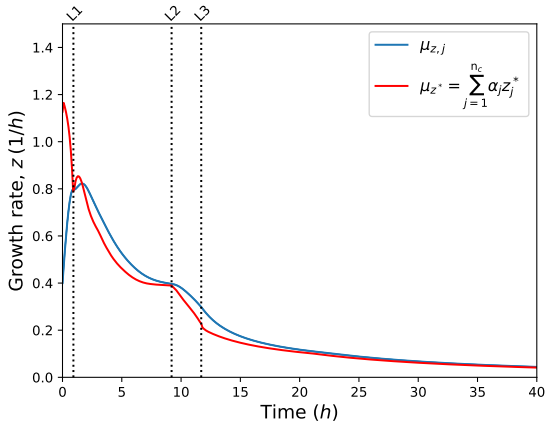


Figure 7.3: Growth rate over time. At each time point, the mean growth rate, μ_z, approaches the volume-averaged optimal growth rate, μ_{z^*} of the optimal rates, z_j^*, set by the external substance concentration in different compartments.

bioreactor population. The MTEs for PopComp are expressed as:

$$\frac{dm_{\beta,\iota}}{dt} = c_{X,\iota} \int_0^{z_{max}} \left(\beta z^{\beta-1} \sum_{j \in \mathbf{l}_\iota} \alpha_j \zeta_j + z^\beta \sum_{j \in \mathbf{l}_\iota} (\alpha_j r_{X,j}) \right) \tau_\iota dz \tag{7.4}$$

The MBEs, as described in the previous section, were applied to ExtComp, along with the initial conditions outlined in Table 6.1 and the model parameters listed in Table 6.2.

While the external concentration profiles of glucose, acetate, and biomass were similar to those shown in Figures 6.13 and 6.14, the growth rate profile of the population, as depicted in Figure 7.3, exhibited a similar mean growth rate, μ_z, but lacked the heterogeneity observed in the results presented in Figure 6.16. The loss of information of extracellular gradients in the model is evident from the replacement of the compartmental rate of change of the internal state of cells, ζ_j, with the volume-averaged rate of change, $\sum_{j \in \mathbf{l}_\iota} \alpha_j \zeta_j$, in

Equation (7.4), yielding:

$$\begin{aligned}\sum_{j\in\mathbf{l}_\iota}\alpha_j\zeta_j &= \sum_{j\in\mathbf{l}_\iota}\left(\frac{1}{t_A}+z\right)\left(\alpha_j z_j^*-z\right)\\ &= \left(\frac{1}{t_A}+z\right)\left(\sum_{j\in\mathbf{l}_\iota}\alpha_j z_j^*-z\right)\\ &= \left(\frac{1}{t_A}+z\right)\left(\mu_{z^*}-z\right)\end{aligned} \tag{7.5}$$

As shown in Equation (7.5), the current model leads the cell population to experience only the volume-averaged optimal growth rate, μ_{z^*}, as a single attractor, rather than treating the compartment-specific optimal growth rates, z_j^*, as multiple distinct attractors. Therefore, the population will behave as if experiencing a well mixed system, e.g. the lab-scale system discussed in Section 6.3.2, where no gradients are present, and no heterogeneity is expected.

Hence, internal heterogeneity becomes apparent only when $\mathrm{n_{PC}} \geq 2$ since at least two extracellular compartments are needed to inform the intracellular moment balance about extracellular gradients. To minimize the required $\mathrm{n_{PC}}$ while maintaining accurate representation of heterogeneity, the Adaptive PopComp Scheme (APCS) was developed to balance model accuracy with computational efficiency (CPU time). The MTEs for each $\mathbf{l}_\iota$ in PopComp for this study are expressed as:

$$\begin{aligned}\frac{dm_{\beta,\iota}}{dt} = c_{X,\iota}\int_0^{z_{max}}&\left(\beta z^{\beta-1}\sum_{j\in\mathbf{l}_\iota}\alpha_j\zeta_j + z^\beta\sum_{j\in\mathbf{l}_\iota}(\alpha_j r_{X,j})\right)\tau_\iota dz\\ &+\frac{1}{v_\iota}\left(\sum_{\substack{\iota'=1\\ \iota'\neq\iota}}^{\mathrm{n_{PC}}} q_{\iota'\to\iota}m_{\iota',\boldsymbol{\beta}} - m_{\boldsymbol{\beta},\iota}\sum_{\substack{\iota'=1\\ \iota'\neq\iota}}^{\mathrm{n_{PC}}} q_{\iota\to\iota'}\right)\end{aligned} \tag{7.6}$$

As shown in Section 7.2.2, if the PopComp evaluated for the next time step differs from the current one, then the following quantities need to be updated: $\mathbf{l}_\iota$, $q_{\iota\to\iota'}$, and $m_{\beta,\iota}$. The PopComp map was dynamically reconstructed based on the gradient of the optimal growth rate, z_j^* (considered the attractor for the system), and the acceptable level of homogeneity, ϵ_d, for the APCS was set to 0.005 h^{-1}.

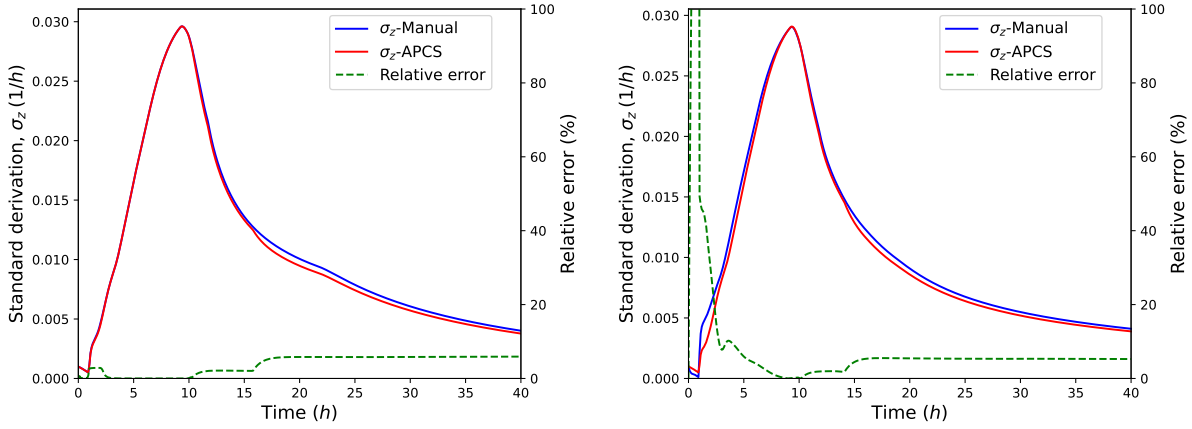
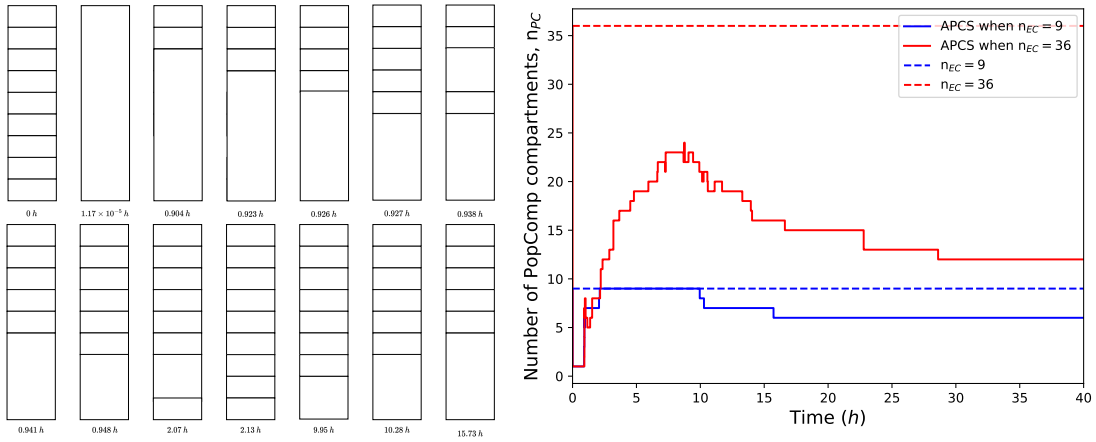


Figure 7.4: Standard deviation (σ_z) and relative error comparing the implementation of APCS with the manual PopComp map fixed at $n_{PC} = n_{EC}$ for (a) $n_{EC} = 9$ and (b) $n_{EC} = 36$.

The simulation results for the standard deviation (σ_z) and the relative error, comparing the scenarios with the implementation of the APCS to those using a fixed manual PopComp map at $n_{PC} = 9$, are presented in Figure 7.4a. APCS dynamically adjusted the PopComp map, as depicted in Figure 7.5. The results indicate that simulations with APCS yielded σ_z values that closely matched those obtained with the manual PopComp map. Throughout the simulation, the relative error remained below 10%.

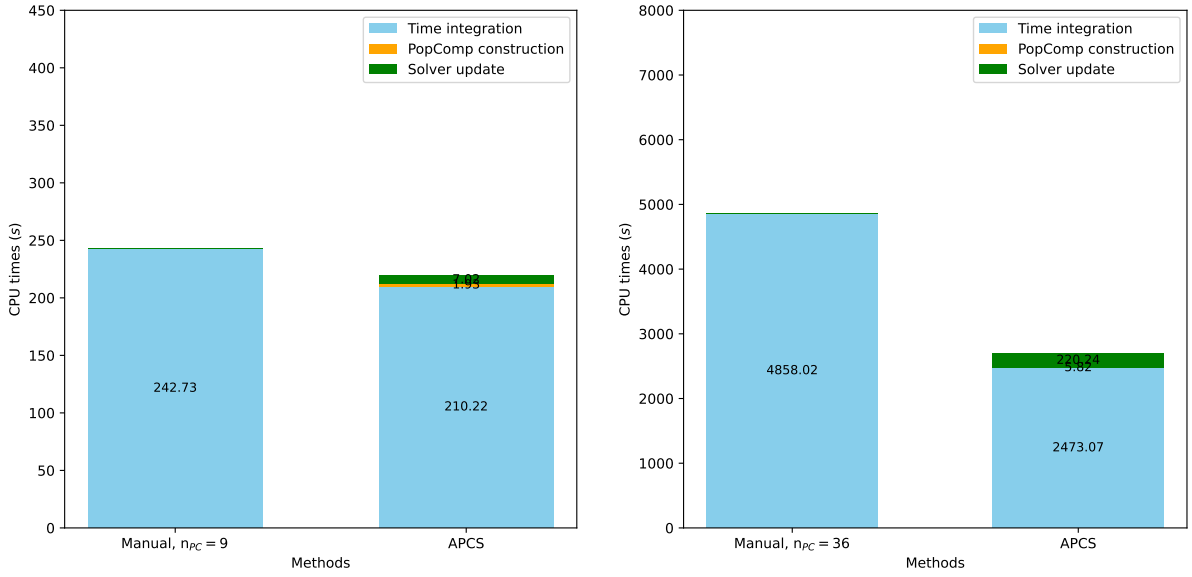


(a) (b)

Figure 7.5: (a) Evolution of the PopComp map and (b) number of PopComp compartments in the map when the APCS was implemented for $n_{EC} = 9$ and $n_{EC} = 36$.

The CPU times required to simulate the large-scale bioreactor using different methods—APCS, the manual PopComp map fixed at $n_{PC} = 1$, and the manual PopComp map fixed at $n_{PC} = 9$—are shown in Figure 7.6a. As expected, implementing APCS resulted in CPU times that were between those of the manual PopComp maps, fixed at $n_{PC} = 1$ and $n_{PC} = 9$. The contribution of PopComp construction and solver updates to the total CPU time was minimal, accounting for only 4%. While APCS achieved a modest 9.71% reduction in CPU time compared to the manual PopComp map fixed at $n_{PC} = 9$, this limited improvement was due to the comparable number of PopComp compartments required relative to the number of ExtComp compartments throughout the simulation (see Figure 7.5b for $n_{EC} = 9$).

On the other hand, much higher CPU savings can be obtained if different CFD-based compartmentalizations are implemented. For example, when the compartmental models in Figure 6.7 were horizontally divided across the baffles, resulting in $n_{EC} = 36$, the APCS method significantly reduced CPU time by approximately 44.43% compared to the manual PopComp map fixed at $n_{PC} = 36$, as shown in Figure 7.6b. This larger time saving was attributed to a lower number of PopComp compartments required relative to the number of ExtComp compartments throughout the simulation (see Figure 7.5b for $n_{EC} = 36$). Moreover, APCS maintained low relative errors during the simulation, as illustrated in Figure 7.4b. The higher relative error observed at the start of the simulation was due to the small expected value of σ_z.



(a) (b)

Figure 7.6: Comparison of CPU time required to simulate the large-scale bioreactor with different methods: the APCS and the manual PopComp map fixed at $n_{PC} = n_{EC}$ for (a) $n_{EC} = 9$ and (b) $n_{EC} = 36$. The scope of each process is shown in Figure 7.2.

7.4 Conclusion

This chapter introduced a novel approach for optimizing the simulation of large-scale bioreactors by incorporating the Adaptive PopComp Scheme (APCS), which enables the model to account for both intracellular heterogeneity and concentration gradients. APCS enhances the computational framework by dynamically adjusting compartmentalization based on real-time system conditions.

By allocating computational resources to regions where gradients and heterogeneity are most pronounced, the model achieves higher efficiency without compromising accuracy. Simulation results demonstrate that APCS maintains an acceptable level of standard deviation (σ_z) while significantly improving computational performance, as reflected by the reduction in CPU time.

Through addressing the trade-off between predictive power and computational cost, APCS makes the model well-suited for large-scale industrial applications that demand

high efficiency, such as model-based real-time control or off-line optimization involving repetitive simulations.

Although this study applied APCS to a univariate number density function ($n_z = 1$), the approach is inherently adaptable and can be extended to simulations involving multiple internal states, resulting in multidimensional number density functions ($n_z \geq 2$). This scalability paves the way for efficiently modeling more complex bioprocesses, offering deeper insight into the behavior and optimization of large-scale bioreactor systems.

By effectively balancing computational efficiency and predictive accuracy, this work lays a robust foundation for developing scalable modeling tools in bioprocess engineering. The insights gained here can support the design and control of next-generation bioreactors, ultimately enabling more efficient, accurate, and intelligent industrial bioprocesses.

Chapter 8

Conclusions and Future Directions

8.1 Conclusions

This thesis presents and validates advanced modeling approaches for improved-accuracy predictions of concentration gradients and microbial population heterogeneity in large-scale bioprocesses. By integrating CFD-derived compartmental models (CMs) with structured metabolic models and population balance frameworks, the developed methods provide more in-depth insights into how environmental gradients influence microbial behavior. The developed integration bioprocess model significantly improved the predictive accuracy for large-scale bioreactor simulations. Throughout this research, a systematic exploration of modeling strategies has led to several key insights that enhance our understanding of concentration gradients and microbial heterogeneity in bioprocesses.

A compartmentalization scheme was developed based on CFD simulations to characterize mixing patterns within a 30 m^3 bioreactor. The study highlighted that conventional compartmentalization approaches may introduce a *short-cutting* effect, potentially leading to inaccurate mixing time predictions. This finding demonstrates the necessity of a manually refined framework to correct these discrepancies and enhance predictive accuracy. Recently, Vasile et al. [109] introduced higher-dimensional approximations, including one-dimensional plug-flow compartments, which could eliminate the need for manual refinements. Further analysis in Section 6.3.4 revealed that manually partitioning the mixing zone cross-sectionally at the height of the impellers and midway between them effectively captured the main circulation within the compartments. This approach yielded mixing time estimates closely matching values reported in the literature. These insights underscore the importance of accurately representing hydrodynamics in bioreactors to improve

modeling reliability and performance.

A key focus of this thesis was to couple the hydrodynamic phenomena to metabolic phenomena occurring both inside and outside the cells. Two different scenarios for metabolic behavior were considered: i) the intracellular concentrations are assumed to reach instantaneous equilibrium with their extracellular environments at all time, and ii) the intracellular concentrations are not in equilibrium with their extracellular environment.

When instantaneous equilibration was assumed, a constraint-based metabolic model was coupled with a CFD-based CM. Since the constraint-based model was based on a linear programming (LP) problem, the binary search tree metabolic modeling (BSTMM) method was developed to rapidly calculate the active set of constraints at any given moment. This approach efficiently integrates dynamic flux balance analysis models (DFBA) with CMs, enabling large-scale simulations of microbial metabolism. The modeling framework was shown to accurately capture diauxic growth behavior in a computational efficient way, allowing hours of bioreactor operation to be simulated in minutes on standard desktop hardware. These findings underscore the potential for the presented modeling framework for real-time applications in model predictive control and bioreactor design optimization. To validate this approach, a comparison between model predictions to data reported for the fermentation of *E. coli* was conducted in both a small scale and a large scale industrial bioreactors. While acceptable fitting was obtained for both bioreactor scales, it was found that the kinetic parameters used for the large-scale bioreactor in this study had to be different from the ones used for the small-scale bioreactor. This finding led to the hypothesis that intracellular heterogeneity may contribute to variations in parameter values as per case ii) above, where instantaneous equilibration between intracellular and extracellular environments is not assumed. Subsequently, an investigation into if accounting for this heterogeneity would permit fitting the data at both scales with the same set of kinetic parameters values. This motivated the investigation of population balance models (PBMs) to account for intracellular heterogeneity.

To address intracellular heterogeneity, a coupled PBM and CM were implemented to account for microbial variability arising from environmental gradients. In parallel, the study highlighted the importance of accurately reconstructing the number density function, particularly when solving a structured metabolic model coupled with PBM. For this task, a novel solution of the method of moments was developed based on an assumed truncated normal distribution (MM-TND) of the intracellular population. This approach was found to provide a more conceptually accurate representation of the number density function compared to a reported quadrature method of moments (QMOM), an extended quadrature method of moments (EQMOM, assuming a normal distribution), and the bin method approach. The relative advantages of the proposed MM-TND method are as follows.

1. Provides more accurate moment calculations than QMOM when the integrands in the moment equations are discontinuous over the integration domain.

2. Respects lower and upper physical bounds, unlike EQMOM with assumed normal distribution.

3. Up to four orders of magnitude faster than the bin method in one-dimensional cases.

This MM-TND based approach correctly captures population heterogeneity through the standard deviation, demonstrating that intracellular heterogeneity emerges when the adaptation rate is comparable to the advection rate, or when their characteristic timescales are similar. Conversely, when adaptation occurs significantly faster or slower than advection, the population tends to exhibit reduced heterogeneity.

Simulation results confirmed that intracellular heterogeneity significantly influences metabolic regulation, leading to distinct variations in growth and substrate consumption across different bioreactor zones. A key finding was that the proposed MM-TND based population balance model coupled to the CM model predicted better the data reported for both a small-scale and a large-scale bioreactor as compared to our earlier work where intracellular heterogeneity was neglected. These findings highlight the importance of considering intracellular heterogeneity when modeling large-scale bioprocesses.

Lastly, an adaptive population compartmental scheme (APCS) was developed and used in combination with the MM-TND method to enhance computational efficiency without compromising model accuracy. The APCS automates the assignment of compartments for the cell population (PopComp) based on an attractor variable derived from concentration gradients, allowing the compartment structure to dynamically adapt throughout the simulation. The APCS approach is particularly accurate for cases where the characteristic advection timescale is shorter than the population adaptation timescale, i.e. when the advection rate exceeds the population adaptation rate. A condition typically observed in large-scale bioreactors. APCS dynamically adjusts the compartmental structure during simulations, optimizing the balance between computational cost and predictive accuracy. Its implementation demonstrated substantial reductions in computational expense while preserving accuracy, positioning APCS as a practical tool for efficiently simulating large-scale bioprocesses.

In summary, this thesis presents modeling approaches that incorporate the effects of concentration gradients and population heterogeneity to enhance the accuracy and predictive capabilities of large-scale bioreactor simulations.

8.2 Recommendations for Future Work

While the developed models demonstrated significant improvements in accuracy and efficiency, several areas warrant further research:

1. Further development of CFD-derived higher-dimensional CMs and dynamic compartmentalization for large-scale bioreactors is needed. In addition to introducing one-dimensional plug-flow compartments [109], Vasile et al. also explored the dynamic compartmentalization of vortex shedding. This prior work provides a foundation for studying the effects of macro-instabilities in large-scale bioreactors, facilitating the simulation of concentration oscillations caused by stable periodic flows.

2. A most detailed metabolic model can be used (i.e., dynamic flux balance model) with the segregated model (i.e., population balance model). The combination of these two approaches warrants further investigation, as it should, in theory, account for all cellular responses and population heterogeneity while minimizing the number of required parameters. Recently, Busschaert et al. [63] proposed integrating flux balance analysis with a population balance model to describe cell size distribution under quasi-steady state growth conditions. The current work represents a crucial first step toward developing a dynamic version of the model and integrating it with CMs for studying large-scale bioreactors.

3. While this study applied APCS to a univariate number density function, the concept is highly adaptable and could be extended to simulations involving multiple internal states, leading to multidimensional number density functions. Such implementation could provide further understanding about the cell population in terms of nutrient consumption across the bioreactor.

4. In view of their computational efficiency, the models developed in this research can be integrated with real-time model predictive control and optimization frameworks for industrial applications. To the author's knowledge, the coupling of a combined population balance model and CM—or CFD-based approaches—with real-time control and optimization strategies has not yet been explored and warrants further investigation. It is believed that such an approach will enable deeper control over heterogeneity within the cell population, potentially leading to more precise and efficient bioprocess management. Additionally, the proposed models could be used for the offline optimization of bioreactor operating conditions and design.

The aforementioned research directions remain highly relevant at the time of this study's publication. By exploring these areas, future research can further enhance the applicability of computational models in optimizing large-scale bioprocesses, ultimately contributing to more sustainable and efficient biomanufacturing practices.

References

[1] Ittisak Promma, Marc G Aucoin, Nasser Mohieddin Abukhdeir, and Hector Budman. A coupled metabolic flux/compartmental hydrodynamic model for large-scale aerated bioreactors. *Computers & Chemical Engineering*, 189:108806, 2024.

[2] Frank R Schmidt. Optimization and scale up of industrial fermentation processes. *Applied microbiology and biotechnology*, 68:425–435, 2005.

[3] Mamoru Ishii. Two-fluid model for two-phase flow. *Multiphase science and technology*, 5(1-4), 1990.

[4] Giorgio Besagni, Nicolò Varallo, and Riccardo Mereu. Computational fluid dynamics modelling of two-phase bubble columns: A comprehensive review. *Fluids*, 8(3):91, 2023.

[5] G Larsson, M Törnkvist, E Ståhl Wernersson, Ch Trägårdh, H Noorman, and S O Enfors. Substrate gradients in bioreactors: origin and consequences. *Bioprocess Engineering*, 14:281–289, 1996.

[6] Jérôme Morchain, Jean-Christophe Gabelle, and Arnaud Cockx. A coupled population balance model and cfd approach for the simulation of mixing issues in lab-scale and industrial bioreactors. *AIChE Journal*, 60(1):27–40, 2014.

[7] FB Campos and PLC Lage. A numerical method for solving the transient multidimensional population balance equation using an euler–lagrange formulation. *Chemical Engineering Science*, 58(12):2725–2744, 2003.

[8] Dadan Darmana, Niels G Deen, and JAM Kuipers. Parallelization of an euler–lagrange model using mixed domain decomposition and a mirror domain technique: Application to dispersed gas–liquid two-phase flow. *Journal of computational physics*, 220(1):216–248, 2006.

[9] Vivek V Buwa, Dhanannjay S Deo, and Vivek V Ranade. Eulerian–lagrangian simulations of unsteady gas–liquid flows in bubble columns. *International journal of multiphase flow*, 32(7):864–885, 2006.

[10] S Besbes, Mahmoud El Hajem, H Ben Aissia, Jean-Yves Champagne, and J Jay. Piv measurements and eulerian–lagrangian simulations of the unsteady gas–liquid flow in a needle sparger rectangular bubble column. *Chemical Engineering Science*, 126:560–572, 2015.

[11] Cyril W Hirt and Billy D Nichols. Volume of fluid (vof) method for the dynamics of free boundaries. *Journal of computational physics*, 39(1):201–225, 1981.

[12] Denis Gueyffier, Jie Li, Ali Nadim, Ruben Scardovelli, and Stéphane Zaleski. Volume-of-fluid interface tracking with smoothed surface stress methods for three-dimensional flows. *Journal of Computational physics*, 152(2):423–456, 1999.

[13] Suraj S Deshpande, Lakshman Anumolu, and Mario F Trujillo. Evaluating the performance of the two-phase flow solver interfoam. *Computational science & discovery*, 5(1):014016, 2012.

[14] Stéphane Popinet. Numerical models of surface tension. *Annual Review of Fluid Mechanics*, 50(1):49–75, 2018.

[15] Vishal Kumar Singh, Ioscani Jiménez del Val, Jarka Glassey, and Fatemeh Kavousi. Integration approaches to model bioreactor hydrodynamics and cellular kinetics for advancing bioprocess optimisation. *Bioengineering*, 11(6):546, 2024.

[16] Thomas Richard Donnelly. A topology-based method for the compartmentalization of multiphysics flows. Master's thesis, University of Waterloo, 2020.

[17] Nicolas Jourdan, Thibaut Neveux, Olivier Potier, Mohamed Kanniche, Jim Wicks, Ingmar Nopens, Usman Rehman, and Yann Le Moullec. Compartmental modelling in chemical engineering: A critical review. *Chemical Engineering Science*, 210:115196, 2019.

[18] Maxime Pigou and Jérôme Morchain. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. *Chemical Engineering Science*, 126:267–282, 2015.

[19] AG Fredrickson, RD Megee Iii, and HM Tsuchiya. Mathematical models for fermentation processes. In *Advances in applied microbiology*, volume 13, pages 419–465. Elsevier, 1970.

[20] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophysical journal*, 83(3):1331–1340, 2002.

[21] Efstratios N Pistikopoulos, Nikolaos A Diangelakis, and Richard Oberdieck. *Multi-parametric optimization and control*. John Wiley & Sons, 2020.

[22] B Xu, M Jahic, G Blomsten, and S-O Enfors. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with escherichia coli. *Applied microbiology and biotechnology*, 51:564–571, 1999.

[23] Gisela Nadal-Rey, Dale D McClure, John M Kavanagh, Sjef Cornelissen, David F Fletcher, and Krist V Gernaey. Understanding gradients in industrial bioreactors. *Biotechnology Advances*, 46:107660, 2021.

[24] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R Williamson, and Terence Hwa. Overflow metabolism in escherichia coli results from efficient proteome allocation. *Nature*, 528(7580):99–104, 2015.

[25] Andrew J Roe, Conor O'Byrne, Debra McLaggan, and Ian R Booth. Inhibition of escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology*, 148(7):2215–2222, 2002.

[26] Bo Xu, Mehmedalija Jahic, and Sven-Olof Enfors. Modeling of overflow metabolism in batch and fed-batch cultures of escherichiacoli. *Biotechnology progress*, 15(1):81–90, 1999.

[27] Thomas Schweder, Elke Krüger, Bo Xu, Britta Jürgen, Gustav Blomsten, Sven-Olof Enfors, and Michael Hecker. Monitoring of genes that respond to process-related stress in large-scale bioprocesses. *Biotechnology and bioengineering*, 65(2):151–159, 1999.

[28] Alvaro R Lara, Lidia Leal, Noemí Flores, Guillermo Gosset, Francisco Bolívar, and Octavio T Ramírez. Transcriptional and metabolic response of recombinant escherichia coli to spatial dissolved oxygen tension gradients simulated in a scale-down system. *Biotechnology and bioengineering*, 93(2):372–385, 2006.

[29] Michael Löffler, Joana Danica Simen, Günter Jäger, Karin Schäferhoff, Andreas Freund, and Ralf Takors. Engineering e. coli for large-scale production–strategies considering atp expenses and transcriptional responses. *Metabolic engineering*, 38:73–85, 2016.

[30] Christopher J Hewitt, Gerhard Nebe-Von Caron, Britta Axelsson, Caroline M McFarlane, and Alvin W Nienow. Studies related to the scale-up of high-cell-density e. coli fed-batch fermentations using multiparameter flow cytometry: Effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnology and bioengineering*, 70(4):381–390, 2000.

[31] Helen Onyeaka, Alvin W Nienow, and Christopher J Hewitt. Further studies related to the scale-up of high cell density escherichia coli fed-batch fermentations: the additional effect of a changing microenvironment when using aqueous ammonia to control ph. *Biotechnology and bioengineering*, 84(4):474–484, 2003.

[32] Peter Neubauer and Stefan Junne. Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Current opinion in biotechnology*, 21(1):114–121, 2010.

[33] Aysun Adan, Günel Alizada, Yağmur Kiraz, Yusuf Baran, and Ayten Nalbant. Flow cytometry: basic principles and applications. *Critical reviews in biotechnology*, 37(2):163–176, 2017.

[34] Guan Heng Yeoh and Jiyuan Tu. *Computational techniques for multiphase flows*. Butterworth-Heinemann, 2019.

[35] R Issa, JY Luo, and D Gosman. Prediction of impeller induced flows in mixing vessels using multiple frames of reference. 4th european conf. on mixing. In *IMechE Symposium Series*, 1994.

[36] Isabelle Naude, Catherine Xuereb, and Joél Bertrand. Direct prediction of the flows induced by a propeller in an agitated vessel using an unstructured mesh. *The Canadian Journal of Chemical Engineering*, 76(3):631–640, 1998.

[37] Harshal Patil, Ajey Kumar Patel, Harish J Pant, and A Venu Vinod. Cfd simulation model for mixing tank using multiple reference frame (mrf) impeller rotation. *ISH Journal of Hydraulic Engineering*, 27(2):200–209, 2021.

[38] JY Luo. Full flow field computation of mixing in baffled stirred vessels. *Trans. I. Chem. E.*, 71:342–344, 1993.

[39] André Bakker, Richard D LaRoche, Min-Hua Wang, and Richard V Calabrese. Sliding mesh simulation of laminar flow in stirred reactors. *Chemical Engineering Research and Design*, 75(1):42–44, 1997.

[40] A Dewan, V Buwa, and F Durst. Performance optimizations of grid disc impellers for mixing of single-phase flows in a stirred vessel. *Chemical Engineering Research and Design*, 84(8):691–702, 2006.

[41] OpenFOAM Foundation. C++ source code guide for openfoam v12, 2024. Accessed: 2024-11-13.

[42] Ansys. Fluent theory guide, 2024. Accessed: 2024-11-13.

[43] Mehrzad Tabatabaian. *CFD module*. Mercury Learning and Information, 2015.

[44] Alírio E Rodrigues. Residence time distribution (rtd) revisited. *Chemical engineering science*, 230:116188, 2021.

[45] An Chi Guo, Timothy Jewison, Michael Wilson, Yifeng Liu, Craig Knox, Yannick Djoumbou, Patrick Lo, Rupasri Mandal, Ram Krishnamurthy, and David S Wishart. Ecmdb: the e. coli metabolome database. *Nucleic acids research*, 41(D1):D625–D630, 2012.

[46] Gregory Stephanopoulos. Metabolic fluxes and metabolic engineering. *Metabolic engineering*, 1(1):1–11, 1999.

[47] Maciek R Antoniewicz. Methods and advances in metabolic flux analysis: a mini-review. *Journal of Industrial Microbiology and Biotechnology*, 42(3):317–325, 2015.

[48] Maciek R Antoniewicz. A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. *Metabolic engineering*, 63:2–12, 2021.

[49] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3):344–349, 2010.

[50] Amit Varma and Bernhard O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Applied and environmental microbiology*, 60(10):3724–3731, 1994.

[51] Honghao Zheng, Luis Ricardez-Sandoval, and Hector Budman. Robust estimation and economic predictive control for dynamic metabolic flux systems under probabilistic uncertainty. *Computers & Chemical Engineering*, 140:106918, 2020.

[52] Doraiswami Ramkrishna. *Population balances: Theory and applications to particulate systems in engineering*. Elsevier, 2000.

[53] Hugh M Hulburt and Stanley Katz. Some problems in particle technology: A statistical mechanical formulation. *Chemical engineering science*, 19(8):555–574, 1964.

[54] Peter Vrábel, Rob GJM van der Lans, Freija N van der Schot, Karel Ch AM Luyben, Bo Xu, and Sven-Olof Enfors. Cma: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chemical engineering journal*, 84(3):463–474, 2001.

[55] Peter Vrábel, Rob GJM van der Lans, Karel Ch AM Luyben, Lotte Boon, and Alvin W Nienow. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements. *Chemical Engineering Science*, 55(23):5881–5896, 2000.

[56] Maxime Pigou, Jérôme Morchain, Pascal Fede, Marie-Isabelle Penet, and Geoffrey Laronze. An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors. *Chemical Engineering Science*, 171:218–232, 2017.

[57] Gisela Nadal-Rey, Dale D McClure, John M Kavanagh, Benny Cassells, Sjef Cornelissen, David F Fletcher, and Krist V Gernaey. Development of dynamic compartment models for industrial aerobic fed-batch fermentation processes. *Chemical Engineering Journal*, 420:130402, 2021.

[58] Bernhard Sonnleitner and Othmar Käppeli. Growth of saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. *Biotechnology and bioengineering*, 28(6):927–937, 1986.

[59] Tannaz Tajsoleiman, Robert Spann, Christian Bach, Krist V Gernaey, Jakob Kjøbsted Huusom, and Ulrich Krühne. A cfd based automatic method for compartment model development. *Computers & Chemical Engineering*, 123:236–245, 2019.

[60] Jérôme Morchain, Jean-Christophe Gabelle, and Arnaud Cockx. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. *AIChE Journal*, 59(2):369–379, 2013.

[61] Rita Lencastre Fernandes, Magnus Carlquist, Luisa Lundin, Anna-Lena Heins, Abhishek Dutta, Søren J Sørensen, Anker D Jensen, Ingmar Nopens, Anna Eliasson Lantz, and Krist V Gernaey. Cell mass and cell cycle dynamics of an asynchronous budding yeast population: experimental observations, flow cytometry data analysis, and multi-scale modeling. *Biotechnology and bioengineering*, 110(3):812–826, 2013.

[62] Vincent Quedeville, Hicham Ouazaite, Bastien Polizzi, Rodney O Fox, Philippe Villedieu, Pascal Fede, Fabien Létisse, and Jérôme Morchain. A two-dimensional population balance model for cell growth including multiple uptake systems. *Chemical Engineering Research and Design*, 132:966–981, 2018.

[63] Michiel Busschaert, Florence H Vermeire, and Steffen Waldherr. Modeling cell size distribution with heterogeneous flux balance analysis. *IEEE Control Systems Letters*, 7:1903–1908, 2023.

[64] Gary J Wells and W Harmon Ray. Methodology for modeling detailed imperfect mixing effects in complex reactors. *AIChE journal*, 51(5):1508–1520, 2005.

[65] Fabrizio Bezzo, S Macchietto, and CC Pantelides. General hybrid multizonal/cfd approach for bioreactor modeling. *AIChE Journal*, 49(8):2133–2148, 2003.

[66] Fabrizio Bezzo and Sandro Macchietto. A general methodology for hybrid multizonal/cfd models: Part ii. automatic zoning. *Computers & chemical engineering*, 28(4):513–525, 2004.

[67] Angélique Delafosse, Frank Delvigne, Marie-Laure Collignon, Michel Crine, Philippe Thonart, and Dominique Toye. Development of a compartment model based on cfd simulations for description of mixing in bioreactors. *BASE*, 2010.

[68] Alexandru Andrei Vasile, Marc G Aucoin, Hector Budman, and Nasser Mohieddin Abukhdeir. A flow alignment-informed method for compartmental modelling. *Computers & Chemical Engineering*, 185:108650, 2024.

[69] Xin Shen and Hector Budman. Set membership estimation with dynamic flux balance models. *Processes*, 9(10):1762, 2021.

[70] Petter Tøndel, Tor Arne Johansen, and Alberto Bemporad. Evaluation of piecewise affine control via binary search tree. *Automatica*, 39(5):945–950, 2003.

[71] Lucas Schulze, Douglas W Bertol, and Guilherme V Raffo. Fast computation of binary search tree for pwa functions representation using intersection classification. *Automatica*, 141:110217, 2022.

[72] Nikolaos V Mantzaris, Prodromos Daoutidis, and Friedrich Srienc. Numerical solution of multi-variable cell population balance models: I. finite difference methods. *Computers & Chemical Engineering*, 25(11-12):1411–1440, 2001.

[73] Daniel O'Sullivan and Stelios Rigopoulos. A conservative finite volume method for the population balance equation with aggregation, fragmentation, nucleation and growth. *Chemical Engineering Science*, 263:117925, 2022.

[74] Stelios Rigopoulos and Alan G Jones. Finite-element scheme for solution of the dynamic population balance equation. *AIChE Journal*, 49(5):1127–1139, 2003.

[75] Nikolaos V Mantzaris, Prodromos Daoutidis, and Friedrich Srienc. Numerical solution of multi-variable cell population balance models. iii. finite element methods. *Computers & Chemical Engineering*, 25(11-12):1463–1481, 2001.

[76] Nikolaos V Mantzaris, Prodromos Daoutidis, and Friedrich Srienc. Numerical solution of multi-variable cell population balance models: I. finite difference methods. *Computers & Chemical Engineering*, 25(11-12):1411–1440, 2001.

[77] Doraiswami Ramkrishna. Analysis of population balance—iv: The precise connection between monte carlo simulation and population balances. *Chemical Engineering Science*, 36(7):1203–1209, 1981.

[78] Hecham M Omar and Sohrab Rohani. Crystal population balance formulation and solution methods: a review. *Crystal Growth & Design*, 17(7):4028–4041, 2017.

[79] Mohsen Shiea, Antonio Buffo, Marco Vanni, and Daniele Marchisio. Numerical methods for the solution of population balance equations coupled with computational fluid dynamics. *Annual review of chemical and biomolecular engineering*, 11(1):339–366, 2020.

[80] Jitendra Kumar, Mirko Peglow, Gerald Warnecke, and Stefan Heinrich. The cell average technique for solving multi-dimensional aggregation population balance equations. *Computers & Chemical Engineering*, 32(8):1810–1830, 2008.

[81] Antonio Buffo and Ville Alopaeus. Solution of bivariate population balance equations with high-order moment-conserving method of classes. *Computers & Chemical Engineering*, 87:111–124, 2016.

[82] Antonio Buffo and Daniele L Marchisio. Modeling and simulation of turbulent polydisperse gas-liquid systems via the generalized population balance equation. *Reviews in chemical Engineering*, 30(1):73–126, 2014.

[83] Antonio Buffo, Daniele L Marchisio, Marco Vanni, and Peter Renze. Simulation of polydisperse multiphase systems using population balances and example application to bubbly flows. *Chemical engineering research and design*, 91(10):1859–1875, 2013.

[84] Michael Frenklach. Method of moments with interpolative closure. *Chemical Engineering Science*, 57(12):2229–2239, 2002.

[85] KW Lee. Change of particle size distribution during brownian coagulation. *Journal of Colloid and Interface Science*, 92(2):315–325, 1983.

[86] F Einar Kruis, Karl A Kusters, Sotiris E Pratsinis, and Brian Scarlett. A simple model for the evolution of the characteristics of aggregate particles undergoing coagulation and sintering. *Aerosol science and technology*, 19(4):514–526, 1993.

[87] Matteo Strumendo and Hamid Arastoopour. Solution of pbe by mom in finite size domains. *Chemical Engineering Science*, 63(10):2624–2640, 2008.

[88] Robert McGraw. Description of aerosol dynamics by the quadrature method of moments. *Aerosol science and Technology*, 27(2):255–265, 1997.

[89] Daniele L Marchisio, R Dennis Vigil, and Rodney O Fox. Quadrature method of moments for aggregation–breakage processes. *Journal of colloid and interface science*, 258(2):322–334, 2003.

[90] Daniele L Marchisio and Rodney O Fox. Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science*, 36(1):43–73, 2005.

[91] Roy G Gordon. Error bounds in equilibrium statistical mechanics. *Journal of Mathematical Physics*, 9(5):655–663, 1968.

[92] John C Wheeler. Modified moments and gaussian quadratures. *The Rocky Mountain Journal of Mathematics*, 4(2):287–296, 1974.

[93] C Yuan, Frédérique Laurent, and RO Fox. An extended quadrature method of moments for population balance equations. *Journal of Aerosol Science*, 51:1–23, 2012.

[94] Raymond Kan and Cesare Robotti. On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics*, 26(4):930–934, 2017.

[95] Christopher Greenshields. *OpenFOAM v10 User Guide*. The OpenFOAM Foundation, London, UK, 2022.

[96] JY Luo. Prediction of impeller induced flows in mixing vessels using multiple frames of reference. In *I. Chem. E. Symposium Series*, volume 136, pages 549–556, 1994.

[97] Mamoru Ishii and Novak Zuber. Drag coefficient and relative velocity in bubbly, droplet or particulate flows. *AIChE journal*, 25(5):843–855, 1979.

[98] Holger Marschall. *Towards the numerical simulation of multi-scale two-phase flows*. PhD thesis, Technische Universität München, 2011.

[99] Akio Tomiyama, Hidesada Tamai, Iztok Zun, and Shigeo Hosokawa. Transverse migration of single bubbles in simple shear flows. *Chemical Engineering Science*, 57(11):1849–1858, 2002.

[100] William E Ranz. Evaporation from drops-i and-ii. *Chem. Eng. Progr*, 48:141–146, 1952.

[101] Brian Cabral and Leith Casey Leedom. Imaging vector fields using line integral convolution. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 263–270, 1993.

[102] Robert S Laramee, Bruno Jobard, and Helwig Hauser. Image space based visualization of unsteady flow on surfaces. In *IEEE Visualization, 2003. VIS 2003.*, pages 131–138. IEEE, 2003.

[103] Johan Revstedt, Laszlo Fuchs, and C Trägárdh. Large eddy simulations of the turbulent flow in a stirred reactor. *Chemical Engineering Science*, 53(24):4041–4053, 1998.

[104] H Hartmann, JJ Derksen, and HEA Van den Akker. Macroinstability uncovered in a rushton turbine stirred tank by means of les. *AIChE journal*, 50(10):2383–2393, 2004.

[105] Cees Haringa, Ruben Vandewijer, and Robert F Mudde. Inter-compartment interaction in multi-impeller mixing. part ii. experiments, sliding mesh and large eddy simulations. *Chemical Engineering Research and Design*, 136:886–899, 2018.

[106] H Hartmann, JJ Derksen, and HEA Van den Akker. Mixing times in a turbulent stirred tank by means of les. *AIChE Journal*, 52(11):3696–3706, 2006.

[107] Alena Kukuková, Benjamin Noël, Suzanne M Kresta, and Joelle Aubin. Impact of sampling method and scale on the measurement of mixing and the coefficient of variance. *AIChE journal*, 54(12):3068–3083, 2008.

[108] Cees Haringa, Ruben Vandewijer, and Robert F Mudde. Inter-compartment interaction in multi-impeller mixing: Part i. experiments and multiple reference frame cfd. *Chemical Engineering Research and Design*, 136:870–885, 2018.

[109] Alexandru Andrei Vasile. *A High-Order, Flow-Alignment-Based Compartmental Modelling Method*. PhD thesis, University of Waterloo, 2025.

[110] P Vrábel, RGJM Van der Lans, YQ Cui, and K Ch AM Luyben. Compartment model approach: Mixing in large scale aerated reactors with multiple impellers. *Chemical Engineering Research and Design*, 77(4):291–302, 1999.

[111] William R Harcombe, William J Riehl, Ilija Dukovski, Brian R Granger, Alex Betts, Alex H Lang, Gracia Bonilla, Amrita Kar, Nicholas Leiby, Pankaj Mehta, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell reports*, 7(4):1104–1115, 2014.

[112] Michael A Henson and Timothy J Hanly. Dynamic flux balance analysis for synthetic microbial communities. *IET systems biology*, 8(5):214–229, 2014.

[113] Ryan P Nolan and Kyongbum Lee. Dynamic model of cho cell metabolism. *Metabolic engineering*, 13(1):108–124, 2011.

[114] Ali Sohrabi Shoabjareh, Fatollah Farhadi, and Mahmoud Reza Pishvaie. Simdfba: A framework for bioprocess simulation and development. *Computers & Chemical Engineering*, 169:108073, 2023.

[115] Alexandra Grancharova and Tor Arne Johansen. *Explicit nonlinear model predictive control: Theory and applications*, volume 429. Springer Science & Business Media, 2012.

[116] HY Lin, B Mathiszik, B Xu, S-O Enfors, and P Neubauer. Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of escherichia coli. *Biotechnology and bioengineering*, 73(5):347–357, 2001.

[117] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris,

Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[118] Fuchang Gao and Lixing Han. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1):259–277, 2012.

[119] Christophe H Schilling, Jeremy S Edwards, David Letscher, and Bernhard Ø Palsson. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and bioengineering*, 71(4):286–306, 2000.

[120] John E Cronan Jr and David Laporte. Tricarboxylic acid cycle and glyoxylate bypass. *EcoSal Plus*, 1(2):10–1128, 2005.

[121] Min-Kyu Oh and James C Liao. Gene expression profiling by dna microarrays and metabolic fluxes in escherichiacoli. *Biotechnology progress*, 16(2):278–286, 2000.

[122] Stéphanie Portet. A primer on model selection using the akaike information criterion. *Infectious Disease Modelling*, 5:111–128, 2020.

[123] I Houcine, E Plasari, R David, and J Villermaux. Feedstream jet intermittency phenomenon in a continuous stirred tank reactor. *Chemical Engineering Journal*, 72(1):19–29, 1999.

[124] S-O Enfors, M Jahic, A Rozkov, B Xu, M Hecker, B Jürgen, E Krüger, T Schweder, G Hamer, D O'beirne, et al. Physiological responses to mixing in large scale bioreactors. *Journal of biotechnology*, 85(2):175–185, 2001.

[125] Dale E Seborg, Thomas F Edgar, Duncan A Mellichamp, and Francis J Doyle III. *Process dynamics and control*, chapter Frequency Response Analysis and Control System Design, pages 244–261. John Wiley & Sons, 2016.

[126] P Hasal, I Fort, and J Kratena. Force effects of the macro-instability of flow pattern on radial baffles in a stirred vessel with pitched-blade and rushton turbine impellers. *Chemical Engineering Research and Design*, 82(9):1268–1281, 2004.

[127] Gary L Kleman and William R Strohl. Acetate metabolism by escherichia coli in high-cell-density fermentation. *Applied and environmental microbiology*, 60(11):3952–3958, 1994.

[128] Vanessa Zavatti, Hector Budman, Raymond L Legge, and Melih Tamer. Evaluation of flow cytometry and chemometric models for monitoring and predicting antigen production at full-scale. *Biochemical Engineering Journal*, 175:108136, 2021.

[129] Abhishek Mishra, Sarah Steinbach, Ibrahim M Tamer, and Hector Budman. Effect of oxidative stress on antigen productivity in b. pertussis cultures. *Biochemical Engineering Journal*, 210:109421, 2024.

[130] Amit Varma and Bernhard O Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/technology*, 12(10):994–998, 1994.

[131] George Stephanopoulos, Aristos A Aristidou, and Jens Nielsen. *Metabolic engineering: principles and methodologies*, chapter Review of Cellular Metabolism, pages 22–27. Elsevier, 1998.

[132] James Ahrens, Berk Geveci, and Charles Law. *Visualization Handbook*, chapter ParaView: An End-User Tool for Large Data Visualization, pages 717–731. Elsevier Inc., Burlington, MA, USA, 2005.

[133] Dirk Wilhelm. Rotating flow simulations with openfoam. *International Journal of Aeronautical Science & Aerospace Research*, 1(001), 2015.

[134] Jannike Solsvik and Hugo A Jakobsen. The foundation of the population balance equation: a review. *Journal of Dispersion Science and Technology*, 36(4):510–520, 2015.

[135] Shahida Anusha Siddiqui, Zeki Erol, Jerina Rugji, Fulya Taşçı, Hatice Ahu Kahraman, Valeria Toppi, Laura Musa, Giacomo Di Giacinto, Nur Alim Bahmid, Mohammad Mehdizadeh, et al. An overview of fermentation in the food industry-looking back from a new perspective. *Bioresources and Bioprocessing*, 10(1):85, 2023.

[136] Sören Werner, Stephan C Kaiser, Matthias Kraume, and Dieter Eibl. Computational fluid dynamics as a modern tool for engineering characterization of bioreactors. *Pharmaceutical Bioprocessing*, 2(1):85–99, 2014.

[137] Hugo A Jakobsen. Chemical reactor modeling. *Multiphase Reactive Flows*, 2008.

[138] Niels Gerbrand Deen, Tron Solberg, and Bjørn Helge Hjertager. Large eddy simulation of the gas–liquid flow in a square cross-sectioned bubble column. *Chemical engineering science*, 56(21-22):6341–6349, 2001.