A FRAMEWORK FOR DEVELOPMENT OF INTEGRATED AND COMPUTATIONALLY FEASIBLE MODELS OF LARGE-SCALE MAMMALIAN CELL BIOREACTORS

By

PARHAM FARZAN

A dissertation submitted to the

School of Graduate Studies

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Marianthi G. Ierapetritou

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2018

ABSTRACT OF THE DISSERTATION

A Framework for Development of Integrated and Computationally Feasible Models of Large-Scale Mammalian Cell Bioreactors

By

PARHAM FARZAN

Dissertation Director

Marianthi G. Ierapetritou

Industrialization of bioreactors has been achieved by integrating knowledge from several applying core concepts of chemical engineering, cellular and molecular biology, and biochemistry. Mathematical modeling has provided insight into biological and physical phenomena in bioreactors. Currently, cell culture models focus on the relevant portions of the cellular activities that control the production of protein of interest. Application of existing models for improvement of bioreactor operation is first investigated. Particular attention is paid to large scale mammalian cultures for their key role in production of complex therapeutic proteins and financial success of biotechnology industry. The attributes of such culture demand enhancement in modeling as their performance is sensitive to local gradients of chemical and physical stimuli. A framework for development of dynamic and computationally feasible models that take into account the interactions of hydrodynamics and cellular activities is presented. The system is modeled as a network of zones to capture spatial heterogeneity. For computational tractability the biophase evolves dynamically over continuous time while

hydrodynamics is defined over a discrete space. The discrete space consists of steady states of flow under predetermined operating conditions. Definition of states for operation through discretization of process parameters converts the problem into a dynamic transition problem. The decomposition of time and space provides the necessary formulation and modeling environment for coupling the model with nonlinear solvers and optimization of operation policy. Sensitivity of the cell culture model to operating conditions is a novel feature which allows studying the effects of hydrodynamics on growth, viability and productivity of organisms. Finally, the application of surrogate modeling for replacing the discrete space of hydrodynamics with a semi-continuous space is investigated.

Acknowledgement

I would like to gratefully acknowledge my advisor, Professor Marianthi G. Ierapetritou, for her supervision, insightful advice, kind encouragement, and endless support during my Ph.D. I would like to express my sincere appreciation for the opportunity to learn and engage in solving interesting problems. I would also like to thank the members of the committee for their valuable feedback and suggestions; Professor Ioannis Androulakis, Professor Rohit Ramachandran and Professor David Coit. I would like to acknowledge the School of Engineering for funding and support throughout my graduate research career at Rutgers.

Throughout my time at Rutgers, my fellow graduate students and colleagues have enriched my experience providing insight, encouragement, and most importantly, friendship: Ravendra, Jinjun, Zhaojia, Nikisha, Amanda, Nihar, Sebastian, Zilong, Xian, Shu, Shishir, Charles, Abhay, Lisia, Nirupaplava, Atharv, Ou, Siddharth and Praneeth.

Finally, I would like to acknowledge my family who have been always the source of my strength and courage. My parents and my brother have always motivated me to set higher goals and work hard towards them.

Table of Contents

List of Figures

List of Tables

List of Symbols

μ: Cellular growth rate

μ_d: Cellular death rate

$A_{c,c'}$: Contact area between neighbor compartments c and c'

$[Amm]$: Concentration of ammonia in the bioreactor

$\boldsymbol{C}_{\mathrm{i}}$: Composition of ith feeding step

$[DO]$: Concentration of dissolved oxygen in the bioreactor

$\mathrm{DO}_{\mathrm{cri}}$: Dissolved oxygen concentration criterion for setting aeration rate

DO_{eq}: Oxygen saturation concentration at 37°C

$DO_{metabolite}$: Correction factor for uptake and production rates of metabolites

$[Glc]$: Concentration of glucose in the bioreactor

$[Gln]$: Concentration of glutamine in the bioreactor

$k_{d,bubble_{op,c}}$: Rate of loss of viable cells under operating condition op in compartment c due to interaction with bubbles

$k_{d,shear_{op,c}}$: Rate of loss of viable cells under operating condition op in compartment c due to interaction with mechanical shear

$(k_L a)_{op}$: Overall volumetric mass transfer coefficient under operating condition op

$[Lac]$: Concentration of lactate in the bioreactor

$(\widehat{M}_{c,c'})_{op}$: Corrected mass flowrate from compartment c' to compartment c under operating condition op

$m_{metabolite}$: Rate of utilization of metabolite for maintenance

OUR: Oxygen uptake rate

$\mathrm{RSD_{cri}}$: Relative standard deviation of distribution of cells over compartments used as a criterion for setting agitation rate

$\mathrm{t_f}$: Duration of the operation

$\mathrm{T_i}$: Time of i^{th} feeding step

$U_{G\,op,q}$: Superficial gas velocity in computational cell q under operating condition op

V_c: Volume of the liquid phase in compartment c

v^{sed}: Sedimentation velocity of cells

X: Overall viable cell density in the bioreactor

X_c: Viable cell density in compartment c

1 Introduction

Success rates and blockbuster sales of biologics have made large pharmaceutical companies shift their attention to biopharmaceuticals. Generating global revenue of $163 billion, biopharmaceutical products account for one fifth of pharmaceutical market [1]. The revenue grows at 8% annually which is double that of conventional pharmaceutical products. Within a biopharmaceutical manufacturing process, a cell represents the unit operation for manufacturing the product of interest. Mammalian cells have become the platform of choice for producing complex therapeutic protein. They have a unique advantage over microbial and yeast -based production systems in that the cells possess the necessary cellular machinery to manufacture and secrete large proteins with the necessary post-translational modifications. Protein produced in mammalian cell cultures represent half of the annual revenue generated by biotechnology industry and the revenue growth is expected to be more than 15% annually [2]. The growing demand for therapeutic proteins from mammalian cells and increasing focus from regulatory bodies have confronted the industry with challenges of ensuring affordability, delivery, and consistent quality of products. The productivity of the cells have been improved due to metabolic engineering, systems biology, and synthetic biology. Metabolic engineering is most well-known and established methodology for engineering cell productivity. This is achieved by specifically manipulating the activity of reactions or genes. Systems biology sheds light on complex interactions of cellular components and functions to determine the limitations of the process. The systemic understanding of the process is then applied to make precise changes to maximize process productivity. Synthetic biology entails designing and building artificial biological systems for the cell to unravel functionality which was not previously available [3]. The average commercial scale titer for mammalian-expressed products has increased from 0.2 – 0.5 g/L in early 90's to 2.5 g/L in the recent years [4]. This has been

mainly achieved via improving vectors, host cell engineering, clone selection, gene amplification, and cell line screening.

Optimization of cell culture processes is achieved by balancing product yield and cell growth. The goal in operation of a bioreactor is to enhance growth, viability and productivity of organisms by adjusting their environment. The environment refers to physical and chemical stimuli acting on organisms. Physical properties of the flow include bubble size distribution, superficial gas velocity and dissipation rate of energy. Shear stress and energy dissipation rate can be used to quantify hydrodynamic forces acting on an organism. Some of the effects of mechanical stress are lysis, change in distribution and number of surface receptors, alteration in production of specific proteins, rate of DNA synthesis, rate of certain metabolic processes and induction of apoptosis [5]. Chemical stimuli are concentrations of metabolites (e.g. glucose, glutamine, lactate and ammonia), dissolved oxygen concentration, and pH. They may be extended to include concentrations of other supplements which have functions such as species transport enhancement, growth stimulation, shear protection and surface charge modification. Figure 1-1 presents the cause and effect diagram of bioreactor operation. The main operational parameters are impeller rotation speed, gas sparging flow rate, pH and temperature. Inefficient mixing creates spatial gradients in pH, dissolved oxygen (DO), carbon dioxide and metabolites concentrations, mechanical shear, volume fraction of gas phase and temperature. As organisms travel inside the reactor they are exposed to fluctuating environmental conditions which affect metabolism, yield and quality of product [6]. Addition and removal of gases are inevitable parts of operation of a large-scale fermenter. Due to low solubility in water, high densities of organisms quickly consume all the oxygen in a saturated culture and produce enough carbon dioxide that it can have inhibitory effects.

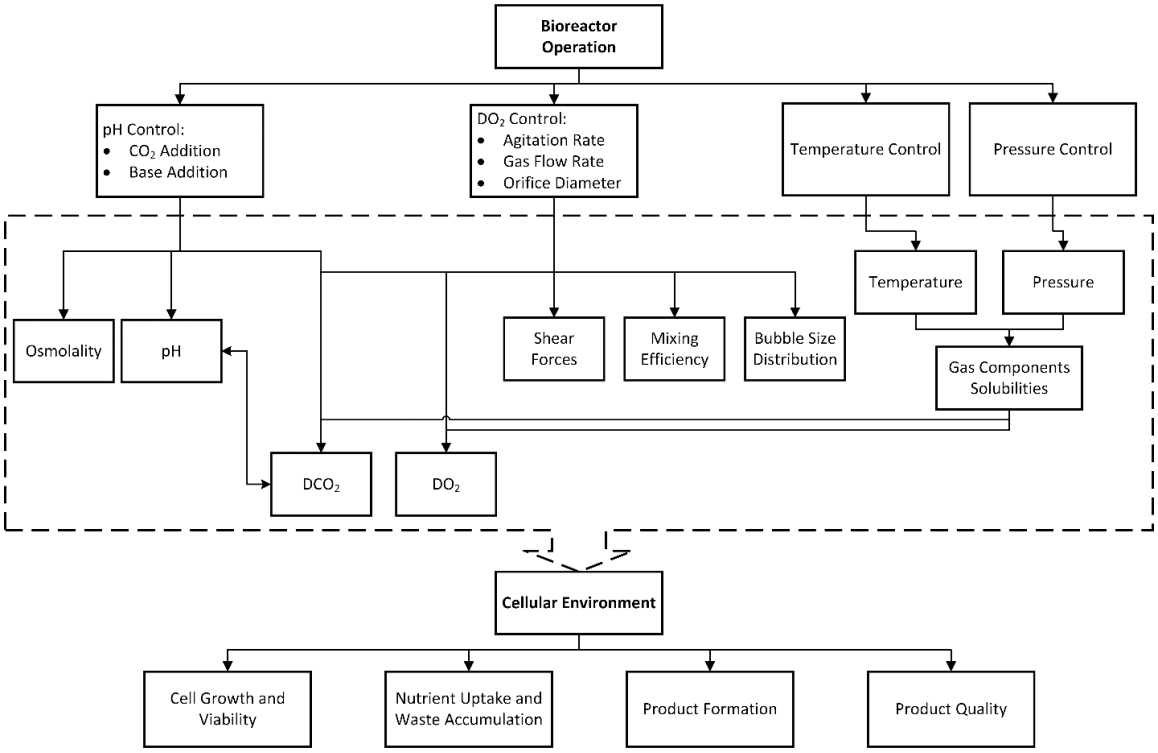

Figure 1-1: Cause and effect diagram for operation of a bioreactor

1.1 Decision-making in biopharmaceutical industry

In the absence of mature manufacturing technology platform, the industry relies on capacity expansion, operational agility and efficiency of supply chain to enhance overall production economics. The capacity of manufacturing sites has increased up to 200,000L due to utilization of large-scale bioreactors (as large as 25,000L) [7]. However scale-up methodologies are still based on overall characterization of the system characteristics such as aeration and agitation mechanisms. Despite growing revenue, the return on innovation, as the engine of the industry, has been declining. In its 2016 annual report, Deloitte Center for Health Solutions published data on Research and Development (R&D) performances of a group of 12 large-cap biopharmaceutical companies. The rate of return on R&D is calculated considering total cost of launching an asset and future net cash inflows an asset is forecast to deliver. The rate of

return has been showing a declining trend from 10.1% in 2010 to 3.7% in 2016 [8]. In response the industry has implemented strategic solutions to increase value of the drug pipeline and reduce cost of launching new assets. Some of these solutions are [9]:

- Replenishing pipeline through Mergers and Acquisitions;
- Development of explicit therapy area focus and expertise;
- Balancing in-house and outsourced activities based on the mission defined for the enterprise;
- Identification of priority customers and timely response to emerging markets;
- Reinvesting in R&D by funding development of new capabilities such as advanced analytics and capturing new tools and technologies, e.g. immuno-oncology and gene editing;
- Data exploitation through ensuring availability and accessibility of data to the right people.

1.2 Motivation and outline of the dissertation

Execution of strategic decisions relies on tactical level decisions [10]. For example strategic R&D management involves project selection, budgeting, and commercialization while tactical R&D addresses the scheduling and resource management necessary for accomplishment of the project. The "National Strategic Plan for Advanced Manufacturing" published by the Executive Office of the President in 2012 highlighted the importance of integrated decision-making for efficient use of data and knowledge. Integration of information and decision-making among functions that comprise the supply chain of a company is achieved by enterprise-wide modeling and optimization. Supply chain encompasses R&D, supply of material, manufacturing and product distribution. Enterprise-wide optimization demands integration of operational activities of planning, scheduling, and control [11]. The design of

control systems for bioprocesses is complicated by uncertainty, the dynamic and nonlinear nature of the process, unreliable online sensors and time delays [12]. Mathematical models can be used to organize data, determine interactions within complex systems, and provide better understanding of biological systems [13]. Modeling requires multi-scale integration to capture complex reactions and transport phenomena and demand many CPU hours due to heterogeneity and dynamics of systems. An established solution for maintaining computational feasibility is substituting the original detailed model with a reduced model. Development of reduced models has been approached either by reduction the order of the model based on evaluation of the significance of its functions in the solution profile or development of surrogate models using data generated after careful experimental design [14]. While the importance of computer aided process engineering has been extensively covered in the literature [15], advancements in biology have outpaced the development of mathematical modeling and analysis in biochemical engineering. Currently, research has focused on developing mechanistic tools to further understand some of the challenges in scale-up and operation of bioreactors which include reduction in productivity and growth and increase in byproduct formation.

Pistikopoulos and his colleagues [16, 17] devised a strategy to mathematically model animal cell culture processes (Figure 1-2). As for the most modeling efforts, a simulation model for a bioreactor should be developed to address specific questions and the model should incorporate details relevant to the questions. An acceptable range of accuracy should be specified in early stages of model development [18] since once its complete the model should be validated within the specified range. Nolan et al described six characteristics of an ideal model [19]. First, the model must be based on a set of inputs that include the initial metabolite concentration, cell density, and fixed set of parameters. Second, the kinetic rate expressions are only dependent

on known and measurable parameters, specifically metabolite concentration, cell densities, and process parameters. Next, the metabolic phenotype should dependent on the biochemical pathways. All the important cellular pathways should be included, like carbon and energy pathways. Additionally, the impact of process parameters on cell density, growth and productivity must be included. A more detailed model may also include cell-cell, cell-substratum, and cell-carrier interactions. These interactions affect gene expression and cell morphology [20]. Finally, the model must be capable of modeling the long term culture dynamics of fed batch culture. Although this idea of the ideal bioreactor model is not new, efforts for modeling physical and biological phenomena have remained separate. The prime objective of this work is to develop a modeling framework which captures the dynamics of the system and takes into account the interactions between cellular activities and hydrodynamics. Calculation of the dynamics of the multiphase flow inside the bioreactor involves solving a large system of partial differential equations including momentum transfer and continuity equations for gas and liquid phases [21]. Therefore, maintaining computational feasibility poses the main challenge in this direction. This is achieved through modeling the system as a network of zones which leads to decomposition of time and space. It reduces the computation while preserves the essential attributes of hydrodynamics.

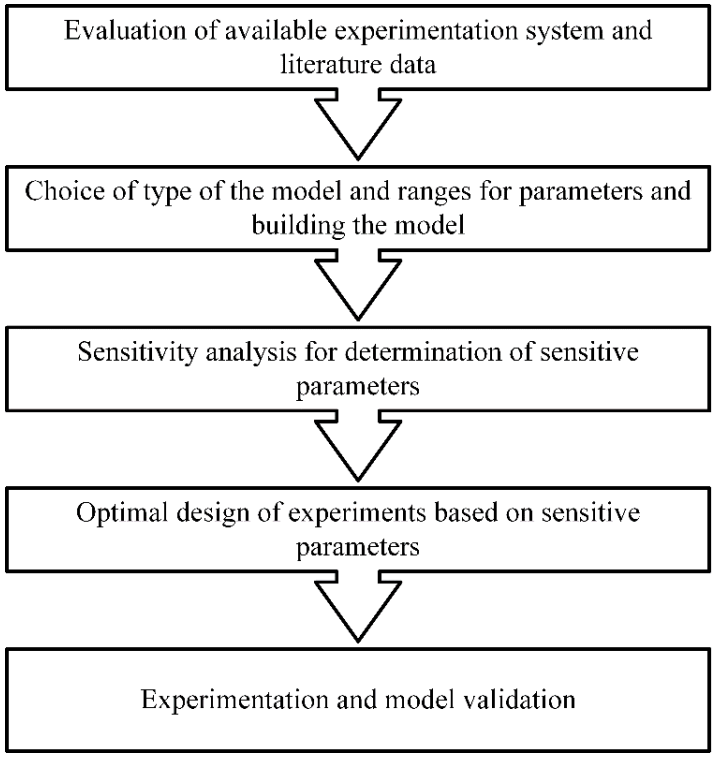

Figure 1-2: Systematic development of predictive mathematical models for animal cell cultures

In chapter 2 of this dissertation, a comprehensive review of modeling approaches for biological and physical phenomena in bioreactors is provided. Chapter 3 presents a modeling framework for addressing major computational challenges of integration of physiology and hydrodynamics. A case study has been conducted to explain the importance of integrated modeling in providing insight and improving bioreactor operation. Chapter 4 investigates application of surrogate modeling for exploitation of large datasets obtained from computer simulations of hydrodynamics. Surrogate modeling can improve the computational feasibility of the framework through minimizing the need for computer simulations. Finally, major conclusions of this dissertation as well as suggestions for future investigations are discussed in Chapter 5.

2 Bioreactor modeling

Bioreactor systems using mammalian cells for protein production can be categorized based on various criteria such as anchorage dependency, mobility of cells, means of aeration, and species transport [22]. Bioreactor systems can operate under different modes; batch, fed-batch, continuous, and perfusion [23]. Considerations regarding set up, operation and control of bioreactors can be found in the literature [24-28]. Independent of category and mode of operation, the phenomena happen in bioreactors can be divided into physiological or physical. For computational and experimental reasons, these two groups have been kept separated in bioreactor models. This chapter reviews methodologies developed for modeling phenomena in each category. The necessity for integration is discussed in the final section of this chapter.

2.1 Modeling of bio-phase

In order to maximize process yield, cell lines have been engineered along with the manufacturing process. Mathematical modeling of cellular functions represents the novel modality for optimizing and characterization of biological processes. This is achieved by simulating relevant portions of the cellular activities that control the production of protein of interest. Metabolic models are divided into two main groups based on whether cell mass composition is considered variable (structured) or fixed (unstructured). Unstructured and structured modeling techniques are reviewed in section 2.1.1. Metabolic models predict the response of a single cell to chemical environmental stimuli. This prediction can be regarded as the average response of cells in the culture. For capturing the heterogeneity among cells, population balance modeling has been implemented. Section 2.1.2 reviews population balance models developed for bioreactor systems.

2.1.1 Unstructured vs. structured models

Unstructured models utilize a reduced number of equations to capture metabolism macroscopically. Unstructured models do not have the capability to capture the effect of the growth condition on cellular composition. As a result they cannot accurately predict growth for a cell in transient condition. Different formulations for unstructured representation of growth and experimental methods for determination of their parameters have been discussed in books on bioprocessing [5, 29]. Table 2-1 reviews few of the papers in which those formulations and methods were used to predict growth rate of mammalian cells. Structured models divide cell mass into components and are based on a set of biochemical reactions, which create stoichiometric relations to represent metabolism of organism. The reactions contain both extracellular components and intracellular components. The flux of the intracellular metabolites can be determined by assuming a pseudo steady state inside the cell. Methods for structured modeling are reviewed in the rest of this section. Table 2-2 provides examples for applications of these methods.

Table 2-1: Unstructured models of mammalian cell cultures

Source	Cell type	Mode of operation	Components of the culture medium	Model predictive capabilities
[30]	Hybridoma	Batch Fed-batch	Glucose, glutamine, ammonia, lactate	Growth rate, death rate, antibody production rate, DNA and RNA synthesis rates, glucose and glutamine consumption rates, lactate and ammonia production rates
[31]	Mouse-mouse hybridoma	Batch Fed-batch	Glucose, glutamine, dissolved oxygen	Growth rate, death rate, antibody production rate, glucose, glutamine, dissolved oxygen consumption rates
[32]	Murine hybridoma	Batch Fed-batch	Glucose, glutamine, ammonia, lactate	Growth rate, death rate, description of cell cycle phases (G0, G1, S, G2/M), dependency of cell cycle transition on temperature, antibody production rate, glucose, glutamine consumption rates, lactate, ammonia production rates

[33]	CHO	Batch Fed-batch	Glucose, glutamine, ammonia, lactate	Growth rate, protein production rate, glucose, glutamine consumption rates, lactate, ammonia production rates
[34]	CHO	Batch Fed-batch	Glucose, lactate	Growth rate, glucose consumption rate, antibody production rate
[35]	CHO	Fed-batch	Glucose, glutamine, lactate, ammonia, and an unknown inhibitor	Utilization of a non-linear model predictive controller for controlling the glucose concentration.
[36]	Hybridoma	Perfusion	Glucose, glutamine, lactate, ammonia	Multivariable output-feedback control for the regulation of cell and glucose concentrations in perfusion cell cultures.
[37]	Hybridoma HB-58	Fed-batch	Glucose, glutamine, lactate, ammonia	Distinction between respiratory (low substrate uptake rate) and overflow metabolism (high substrate uptake rates).

[38]	CHO	Fed-batch	Glucose, glutamine, lactate, ammonia	Growth rate, death rate, antibody production rate, glucose and glutamine consumption rates, lactate and ammonia production rates

2.1.1.1 Metabolic flux analysis

Metabolic flux analysis (MFA) has become a standard tool for measuring the metabolic fluxes of cells at steady state and it has been applied to bioprocesses to gain novel insights not available by empirical means. For example using this methodology, Ahn et al [39] were able to quantify intracellular metabolic fluxes for different cellular growth phases. Europa et al [40] studied the distribution of intracellular fluxes under fed-batch and continuous operation modes. Intracellular fluxes are calculated by assuming the system is at steady state (i.e. pseudo-steady state) and conducting a mass balance around the intracellular metabolites. This gives rise to a system of linear algebraic equations. Since the number of unknown fluxes is larger than the number of balanced metabolites, for this linear algebraic problem to have a single solution, extracellular concentrations of metabolites are measured and the rates of consumption and secretion of components as well as the cellular growth are determined experimentally [41]. Extracellular measurements are required as inputs to estimate the steady state rates of reactions [42]. The biochemical network is simplified to such an extent that a few measured extracellular rates can determine all remaining fluxes in the model [43]. The parametric representation of the metabolic network can be hybridized with a non-parametric statistical regression model of cellular functions whose parameters are unknown to compensate the missing knowledge [41]. Regulatory constraints may also be used to compensate for lack of mass balance equations

[44]. Flux analysis can be used to determine level of involvement of certain pathways. It can be applied to identify metabolic restrictions on product formation and also understand the shift between catabolic and anabolic processes. Flux analysis can offer insight into the distribution of flux around metabolic branch points and be used for studying the interactions of pathways [45]. Defining enough constraints is not possible for large metabolic networks, so a "metabolic objective" (i.e. maximization of growth rate) is defined in the Flux Balance Analysis (FBA) methodology. There may be more than one distribution of flux that maximize the assumed objective [43]. For some cells like wild type E coli, FBA can predict their response successfully because they consume carbon nearly optimally [46]. To circumvent the definition of a specific objective, many approaches have been developed; e.g. minimization of metabolic adjustments (MOMA), regulatory on/off minimization (ROOM) and relative change (RELATCH). The idea is to take the solution obtained from FBA or experiments and adjust fluxes through minimal changes to regulatory and distribution patterns [44].

MFA and FBA can only be applied to early exponential growth phase or continuous culture since the cells are assumed in steady state. Furthermore, concentration dependencies cannot be determined which minimizes the capability of the system to create a dynamic response to culture perturbations. Prior to assuming metabolic fluxes are at steady states it has to be assumed that demands for nutrients by cells are met throughout the operation. In other words, demands and uptakes are equal [47, 48]. This is a major drawback since most industrial bioprocess is operated at fed-batch mode where cell population grows and cells experience constant fluctuations in nutrient, metabolite, and product levels.

2.1.1.2 Analysis of elementary modes (EM) and extreme pathways (EP)

Another group of quasi steady state (QSS) models are the macroscopic bioreaction models (MBMs). The first step is to treat the system like a black box and describe the macroscopic mass balances of the system by a set of elementary reactions. To prevent the system of becoming very large, especially for a system with complex compositions like the mammalian cultures, it is recommended to select few metabolites (e.g. glucose and glutamine) to be modeled by the macro-reactions. Then, intracellular fluxes can be estimated from extracellular production and uptake rates using the MFA methodology. Finally macro-reactions are generated through systematic reduction of the model. One macroscopic reaction is basically a large number of metabolic pathways which have been lumped together [45].

Metabolic reactions can be reduced to linearly independent equivalents by eliminating non-elementary intracellular reactions. This approach requires three conditions to be met; quasi steady state (QSS) assumption, feasibility and non-decomposability [49]. QSS assumption makes sure there is no net consumption or production of balanced metabolites. Feasibility ensures that only thermodynamically allowed fluxes are considered in the distribution of flux. Non-decomposability means that enzymes in one set of lumped pathways should not be a subset of enzymes of any other set. The first pathway based reduction technique is analysis of elementary flux mode (EFM). An elementary mode (EM) is a set of minimum number of reactions which represent a functional unit [50]. Physically, elementary flux analysis (EMA) represents all possibilities of a metabolic network leading to a product and may not be mathematically independent even though it is non-decomposable. Therefore, the macroscopic reactions may not be unique and require further examination. Second pathway analysis technique, analysis of extreme pathways (EPs), requires two additional criteria; network reconfiguration and systemic independence. By reconfiguring the network should all reversible reactions are presented as two irreversible reactions so intracellular reactions do not have a

negative flux. In addition, exchange of metabolite with the environment cannot be described with more than one reaction. Systemic independence makes a set of EPs a unique and minimal subset of EFMs that can represent all feasible steady-state flux distributions. The downside of this method is that some of the generated macro-reactions may have no physiological interpretation since the systematic independence is the only criterion. Because of systemic independence, EPs may exclude important pathways and doesn't incorporate all optimal pathways but quite often the results of EFM and EP coincide [49]. EFM and EP are not specifically modeling techniques however they are utilized as tools prior to model development in order to reduce the size of the model and also to minimize the complexity of the model. Practically by applying EFM and EP multiple biochemical reactions within multiple pathways can be compressed into a set of macroscopic reactions.

2.1.1.3 Dynamic structured modeling

Dynamic models represent an evolution of the pseudo-steady state models. Since metabolic activities are intrinsically non-stationary, dynamic models have many advantages over steady states models. They make it possible to study cyclic and oscillatory phenomena of cells. Moreover, unlike constraint based flux analysis methods, they are not limited to yield calculations and can provide more precise representation of cell productivity [51]. Dynamic flux balance can simulate dynamic metabolic behavior by combining steady state flux analysis with kinetic rate expression. The main challenge, however is to define an objective function (e.g. biomass yield) that is valid for the entire duration of the culture [43]. Flux Spectrum Approach (FSA) eliminates the necessity of such an objective. In this methodology the upper and lower bounds for unknown fluxes are specified which form an underdetermined model. Then dynamic fluxes are calculated using time-series data of a number of extracellular species [52].

Building a structured dynamic model which includes details of metabolic activities requires estimation of kinetic and regulatory parameters [53, 54]. To avoid this parameter estimation, data collected at steady states over a course of time can fit into a model. Phenotypic data are the results of interdependency of kinetic parameters so using them, circumvents quantifying kinetic details of each enzymatic reaction. One approach is to build an ensemble of models that can thermodynamically reach the same flux distribution and metabolic concentrations at different steady states. Then the size of the ensemble is reduced based on their responses to enzymatic perturbations [55]. The screening strategy of models to reach a single model should be robust enough to be independent of data used and the path chosen. The applicability of ensemble modeling can be limited by the accessibility or ability of calculation of reference steady state fluxes [55, 56]. Genetically structured models are only available for small metabolic networks so incorporation of regulation in dynamic models is not always possible. Another problem in developing dynamic models is parameter determination as many of them are insensitive to experimental data. A helpful assumption is that fast changing variables (e.g. intracellular variables) are in quasi steady state (QSS) so only slow variables (e.g. extracellular metabolites) need to be considered. This is a valid assumption since QSS often reaches within seconds or minutes [43]. This assumption makes QSS models capable of predicting the intracellular fluxes that are immediately redistributed in accordance with the change of extracellular variables. The order of dynamic models can be reduced by identifying the appropriate time domain for variables [51]. To implement system dynamics into MFA or FBA, network analysis can be combined with either kinetic relationships or extracellular time-series data. The first approach is preferred for genetic manipulation since reaction rates are available. The latter is used for process monitoring and control, which can be achieved by identification of transient intracellular components. For calculation of intracellular time-dependent fluxes, rates of consumption or secretion of extracellular metabolites are required as functions of time.

These rates should be determined using extracellular concentrations measured over time. Calculation of derivatives of concentrations is another practical problem. To resolve this issue, three approaches have been developed [43]. First, the time domain can be simply divided into distinct metabolic phases, for which average rates are calculated and intracellular fluxes can be calculated for each phase using steady state flux analysis [56]. Second, extracellular concentrations are represented as functions of time using one of the data smoothing techniques (moving average, spline fitting or polynomial fitting). Then fluxes are calculated by application of MFA or FBA for small time steps [52]. The third approach is Derivative-free Dynamic MFA (DMFA) where flux is described by a continuous piecewise linear function that is calculated directly from concentration data [57].

2.1.1.4 Cybernetic modeling

In a cybernetic model, a control system is defined for cells which determines the choice and extent of consumption of substrates [58]. This embedded controller compensates for the simple description of cellular activities and enables the model to predict the dynamic response of cells. Because of the huge number of reactions and complex mechanisms of gene regulatory and enzyme activities building a dynamic model is not easy [51]. Perhaps the easiest way to achieve this is by representing the metabolism of cells by small number of reactions. This unstructured approach toward metabolic network modeling is called lumped cybernetic modeling (LCM). To have a dynamic structured model, quasi steady state assumption still needs to be valid. By considering a metabolic objective a hybrid cybernetic model (HCM) is built which achieves and maintains the optimality through distribution of uptake flux among elementary modes. Cybernetic mechanism can also be a possible solution to unavailability of gene regulatory data.

Table 2-2: Structured models for mammalian cells

Source	Cell Line	Mode of Operation	Components of the Metabolic Network	Number of components	Method
[59]	Mouse-mouse hybridoma	Continuous	Glycolysis, TCA cycle, amino acid pathways	19 amino acids, 32 species	Flux analysis
[60, 61]	Mammalian	Fed-batch	MAb synthesis and secretion, glycosylation	6 species	Reaction kinetics
[53]	Hybridoma	-	Glycolysis, glutaminolysis, TCA cycle, fatty and amino acids metabolism, trans-membrane transport, product formation	50 species	Reaction kinetics
[62, 63]	Mouse-mouse hybridoma	Fed-batch	TCA cycle coupled with glucose and glutamine metabolism	20 amino acids	Stoichio-metric analysis of metabolic network

[40]	Mouse–mouse hybridoma	Fed-batch, Continuous	Glycolysis, glutaminolysis, TCA cycle, lactate production, amino acid catabolism, and synthesis of biomass and antibody	20 amino acids	Flux analysis
[64]	Hybridoma, CHO	Fed-batch, Continuous	TCA cycle, Glycerophospholipid metabolism, Pentose phosphate pathway, Fatty acid biosynthesis, Pyrimidine metabolism, Sphingolipid metabolism, Purine synthesis, Biosynthesis of steroids	228 metabolites	Flux analysis
[65]	CHO	Batch, Continuous	Glycolysis, glutaminolysis, TCA cycle, nucleotides synthesis	12 metabolites	Interval representation of flux analysis
[45]	Hybridoma, CHO	Fed-batch, perfusion	Glycolysis, TCA cycle, glutaminolysis	24 extra-cellular, 13 intracellular species, 18 amino acids	MFA/ DMFA, EFM, EP

[39]	CHO	Fed-batch	Glycolysis, pentose phosphate pathway, TCA cycle, anaplerotic and cataplerotic reactions, amino and fatty acids metabolism, lactate metabolism	77 metabolites	MFA
[66]	Myeloma	Batch, Fed-batch	Glycolysis, pentose phosphate pathway, glutaminolysis, TCA cycle, amino acids metabolism	11 species	Cybernetic
[67]	Myeloma	Batch, Fed-batch	Glycolytic pathway, pentose phosphate pathway, glutaminolytic pathway, TCA cycle	28 species	EM, HCM
[68]	Murine hybridoma	Batch, Fed-batch	TCA cycle, Embden–Meyerhoff fermentative pathway, glutaminolysis, and respiration	30 metabolites, viable and dead cells, MAb	Flux analysis

2.1.2 Single cell vs. population balance models

Single-cell model (SCM) was the first attempt to consider both abiotic and biotic phases in a bioreactor [69] and can predict the average response of a population. In fact, cells behave differently. These different ways of responding to signals from other cells and environment can be related to parameters such as number and distribution of surface receptors, physiological states and mechanical (viscoelastic) properties of cells [20]. Alternatively population balance models (PBM) can describe the distribution of internal cell states. These states are characterized by internal coordinates [70]. Cell age [71] or cell mass [72] are often used as the internal coordinates [73]. Intracellular concentrations has also been used as the internal coordinates [74]. Due to this inherent heterogeneity growth, nutrient uptake and production of the population can be described. Contrary to continuum models [75], which consider cells as a lumped, continuous and expanding phase, information about cell division and partitioning of material upon cell division can be incorporated into population balance models [76]. Moreover, gene regulation, i.e. how cells respond to environmental signals, can be incorporated into the population balance models [77]. For this purpose a stochastic internal coordinate can be introduced that shows weather the cell responds to environmental signals. This coordinate is determined by the strength of the signals. If the signal is neither weak nor strong the value of the coordinate is assigned probabilistically. Stochastic population models can also describe the fact that cells with identical initial state eventually behave differently, which is not possible using single cell models [77]. The generally followed formulation of cell population balances was presented by Ramkrishna [78] in which the population balance equation has the following form:

$$\frac{\partial f_1(\boldsymbol{z},t)}{\partial t} + \nabla.\bar{\dot{Z}}(\boldsymbol{z},\boldsymbol{c}) f_1(\boldsymbol{z},t) = -\sigma(\boldsymbol{z},\boldsymbol{c}) f_1(\boldsymbol{z},t) + 2\int_{\mathrm{B}} \sigma(\boldsymbol{z}',\boldsymbol{c}) p(\boldsymbol{z},\boldsymbol{z}',\boldsymbol{c}) f_1(\boldsymbol{z}',t)\, dv'$$

where:

$\boldsymbol{z} \equiv (z_1, z_2 \ldots z_n)$	cell's physiological state vector
$\boldsymbol{c} \equiv (c_1, c_2 \ldots c_m)$	concentration vector for the nutrient medium
$f_1(\boldsymbol{z},t)$	number density function
B	admissible state in n-dimensional space
$\bar{\dot{Z}}(\boldsymbol{z},\boldsymbol{c})$	growth rate
$\sigma(\boldsymbol{z},\boldsymbol{c})$	transition probability function
$p(\boldsymbol{z},\boldsymbol{z}',\boldsymbol{c})$	partitioning function

For calculation of growth an unstructured approach, for example Monod kinetics, can be used. This is based on the assumption that yield of cell mass production per unit mass of limiting nutrient is constant [74]. To consider intracellular processes, a structured SCM and PBM should be linked. In this case application of model reduction techniques to the SCM can help keeping the model mathematically tractable [79]. In this case assuming growth rate not dependent on the cell mass makes the coupling of SCM and PBM easier. This assumption has been validated for mammalian cells [79].

An alternative formulation for cell population balance models was developed by Nielsen et al. [80]. The distribution is described with the following equation:

$$\frac{\partial \Psi(\boldsymbol{X},t)}{\partial t} + \frac{\partial}{\partial \boldsymbol{X}}[R(\boldsymbol{X},t)\Psi(\boldsymbol{X},t)] = [\mu(\boldsymbol{X},t) - \mu(t)]\Psi(\boldsymbol{X},t)$$

where:

$\boldsymbol{X}$	vector of concentrations of intracellular components
$\Psi(\boldsymbol{X}, t)$	dimensionless discrete distribution function for the mass fraction of cells
$R(\boldsymbol{X}, t)$	net rate of formation of $\boldsymbol{X}$
$\mu(\boldsymbol{X}, t)$	specific growth rate of cells with composition $\boldsymbol{X}$
$\mu(t)$	average specific growth rate of the biomass

The internal coordinates are concentrations of cell components, which can function as a built-in tool for integration of metabolic network and population model. This feature can provide an option for avoiding iterating between SCM and PBM. If cell mass is considered as an internal coordinate, the number density function can be converted to mass fraction distribution function [80]. However, the internal coordinates should be chosen in such a way that the distribution can be validated by experimental data.

Development of PBM requires information at single-cell level for calculation of intrinsic physiological state (IPS) functions; growth and division rates and partitioning probability function [76]. Different methods have been developed to use data obtained under unbalanced [81, 82] or balanced [83] growth conditions and calculate IPS functions via solving an inverse problem. But none of them has advanced to a robust methodology with general applicability. Another hindrance is the complex numerical solutions required because of the nonlinear partial integral differential equations in PBMs [76]. Stamatakis provided a summary of approaches that had been developed to solve these models [84]. A stable distribution of a population can be described by an ensemble of SCMs that differ according to the initial conditions or key properties such as division size [85-88]. A hybrid modeling framework has been introduced in which the internal coordinates are separated into extensive and intensive variables [84].

Extensive variables form a population balance problem of a smaller size and intensive variables are captured by an ensemble or a continuum model. It was shown that under certain conditions this hybrid model is equivalent to the original PBM. Fernandes et al. provided a summary of cell population balance models [89]. Table 2-3 provides examples for population balance models developed for mammalian cell cultures.

Table 2-3: Population balance models of mammalian cell cultures

Source	Cell Type	SCM	Phenomena Captured / Model Capabilities
[90]	CHO	structured	Description of the population based on caspase activity, Coupled extra- and intra-cellular mass balance, MAb production, Inclusion of apoptosis based on caspase activity.
[47]	Mammalian	structured	Coupled extra- and intra-cellular mass balance, Description of the population based on cell mass and cell cycle stage.
[76]	Not specific	unstructured	Description of the population based on cell mass, Effects of extracellular substrates concentrations on population growth.
[91]	Not specific	unstructured	Description of the population based on the intracellular content of one substance, Capturing the stochasticity involved in cell division.
[92]	Mammalian	unstructured	Description of the population based on cell volume, DNA content and cell cycle phase Capturing the effect of contact inhibition on culture expansion.

[93]	AGE1.HNA AT, CHO-K1	unstructured	Description of the population based on cell volume, DNA content and cell cycle phase Ensemble modeling for integration of single-cell models Coupled extra- and intra-cellular mass balance.

2.2 Modeling of hydrodynamics

Expansion of cells for protein production requires agitation and aeration to maintain a homogenous distribution of cells throughout the bioreactor and make sure oxygen and nutrients reach the cells. These functions have been well studied leading to quantitative and qualitative relationships. Over the years, these models have further refined and been utilized to develop industrial manufacturing processes for therapeutic protein production. A model of hydrodynamics of a bioreactor must incorporate cells, growth medium, and bubbles simultaneously as a multiple phase system. The governing equations of the multiphase flow in stirred tank bioreactors have been extensively discussed in mechanical engineering literature [94] and built into commercially available software packages. However to precisely model phenomena such as turbulence [95], interface momentum transfer and mass transfer many parameters need to be estimated. In this section a review of the methods used for the estimation of parameters for gas-liquid mass transfer is provided. In a bioreactor system, oxygen has to transfer from a bubble to inside the cell. Among all the resistances on oxygen transfer the liquid film around the bubble has been recognized as the rate limiting resistance [96, 97]. This resistance can be characterized by the liquid phase mass transfer coefficient and the specific interfacial area. The volumetric gas-liquid mass transfer coefficient, which is the product of liquid phase mass transfer coefficient and specific interfacial area, can be estimated by

measuring dissolved oxygen concentrations over time [98] or by using correlations available in the literature [99-101]. The volumetric mass transfer may be treated as a black box and estimated using data reconciliation techniques [102]. The overall volumetric mass transfer coefficient can be expressed as a function of total specific power dissipated in the liquid [103]. Using these correlations the liquid mass transfer coefficient has been estimated to be 0.01 cm/s for small bubbles and 0.04 cm/s for large bubbles [29]. The average volumetric coefficient cannot help with the spotting of dead zones and oxygen depleted locations. Linek et al. [104] also argued that because of the errors in measuring volumetric mass transfer coefficient it is more favorable to measure or calculate liquid phase transfer coefficient directly. For this purpose several models have been developed which are briefly reviewed in this section. Cachaza et al. [105] compared the results of some of these models with the experimental data.

(a). The film model: In this model a film with a constant thickness and diffusivity is assumed in the liquid phase around the bubble, which is responsible for the entire resistance on mass transfer. In reality the film thickness and diffusivity may not be constant for a single bubble or a swarm of bubbles. Furthermore in the presence of turbulent eddies the role of diffusivity in mass transfer becomes less important. However for highly diffusive gases, large interface contact time or small boundary layer thickness film theory can provide an accurate estimate [106].

(b). Border diffusion layer model: In this steady state model the turbulent flow around a bubble is divided into four sections: the main turbulent stream, the turbulent boundary layer, the viscous sub-layer and the diffusion sub-layer. This division enables the model to account for the effects of both turbulent eddies and molecular diffusion on mass transfer [107]. Azbel [106] provided an empirical correlation for estimation of the diffusion sub-layer thickness

which has also been used for calculation of the heat transfer coefficient in multiphase systems [108].

(c). Higbie penetration model: This penetration model assumes a residence time for a fluid element at the interface in which the mass transfer occurs [109]. The residence time is short and doesn't allow reaching the steady state. The limiting step in mass transfer is surface refreshment, which is done by turbulence. Each turbulent eddy spends approximately the same amount of time at the interface. The interface contact time can be represented as a function of eddy length and fluctuation velocity [110]. Kawase et al. [111] derived an equation using Higbie's penetration theory to express the liquid mass transfer coefficient as a function of energy dissipation rate (ε) and Schmidt number. The penetration theory has gained acceptance and has been used for modeling the mass transfer phenomena in bioreactors [21, 112].

(d). Surface renewal theory: In this model instead of assuming equal contact times for eddies, the contact time is calculated using a statistical probability function. The surface elements are divided into groups based on their times of exposure to the gas phase. In each group the fraction of surface elements which is replaced by fresh surface is assumed to be equal to the mean rate of production of fresh surface(s) [113]. Lamont et al. [114] showed that the mean rate of surface production could be expressed as a function of energy dissipation rate and liquid viscosity. The constants of this function can be calculated using experimental data [115].

(e). Slip velocity model: In this model small and large bubbles are assumed to behave differently [116]. Small bubbles (diameter less than 1 mm) are treated as rigid spheres with immobile surface and slipless interface and the interface limits mass transfer. While large

bubbles (diameter greater than 2.5 mm) have mobile and ellipsoidal surface and eddy turbulence has greater impact on their mass transfer [107]. The equations for calculation of liquid mass transfer coefficients for both groups can be found in Linek et al. [103].

(f). Eddy turbulence model: The model is based on the idea that small scale eddies of the turbulent field are responsible for the surface renewal and the effect of gross mean flow of fluid relative to the bubbles is negligible. The range for the scales of these turbulent motions is from the smallest viscous motions to the inertial motions. These motions are much smaller in scale than the gas bubbles and they attack the surface of bubbles from every direction. As a result, the size of the gas bubble is not a very critical parameter for the estimation of liquid mass transfer coefficient [114]. Linek et al. [117] provided the equation for calculation of the liquid mass transfer coefficient together with the suggested values for its constant.

The models mentioned above are used for calculation of the liquid mass transfer coefficient. To obtain the volumetric coefficient an estimate of the specific interfacial area should be obtained. Correlations for calculation of the specific interface area for Newtonian and non-Newtonian liquids can be found in the literature [118, 119]. Some of these correlations require calculation of the gas hold up which can be done by using available experimentally derived relationships [120]. A better estimation of the interface area of bubbles may be obtained by calculation of the size distribution of bubbles using population balance methodology [121, 122]. In this way it is possible to take into account the effects of bubbles coalescence and breakup, phase change, pressure change and reactions on the interfacial area. However obtaining the size distribution of bubbles does not lead to a precise calculation of specific interfacial area. Most bubbles do not keep the spherical shape as they grow [112], also swarms formed by bubbles cause extra resistance on mass transfer.

2.3 Integration of physiology and hydrodynamics

Reliable mechanistic bioreactor models facilitate improvement in equipment utilization, medium design, and feeding strategy, and shed light on some of the challenges of scale-up including productivity and byproduct formation [48]. Throughout the operation of a large-scale bioreactor extracellular chemical and physical stimuli show spatial and temporal variations. Chemical stimuli include concentrations of nutrients, supplements, and waste. Each chemical stimulus has specific functions such as species transport enhancement, growth stimulation, shear protection, surface charge modification. Physical stimuli include all hydrodynamic forces experienced by the cells in the bioreactor. The experimental observations reported in the literature indicate that these extracellular parameters affect growth, viability and productivity of cell populations [123-127]. For example the impacts of aeration on dynamics of the culture is not limited to providing cells with oxygen. Aeration through gas sparging directly affects viable cell density. Cells attached to bubbles become trapped in the foam layer and die. Also energy dissipated due to bubble rupture is two or three order of magnitudes higher than what cells can tolerate and is one of the causes of cell loss [123]. However, the review of modeling approaches provided in section 2.1 shows that it is common to reduce extracellular stimuli to concentrations of few metabolites. Assuming spatial homogeneity and neglecting the effects of physical environmental parameters have been the basis of most of culture models. This has been due to computational and experimental challenges and led to lack of quantitative representations of the extents of effects of chemical and mechanical stress parameters such as pH, dissolved oxygen, mechanical shear and gas holdup. Exclusion of hydrodynamics in a bioreactor model, results in lumping the effects of different causes of cell loss in one parameter, e.g. growth rate. The performance of the model deteriorates when process conditions change as the parameters of the model are highly dependent on strain, cultivation medium and

fermentation conditions [128]. Consequently existing models don't possess satisfactory predictive power especially when used outside their calibration ranges and literature data can only be used for qualitative studies. In addition to narrow confidence intervals, often achieving a satisfactory fit of experimental data requires considering extra terms or assumptions, which are theoretically difficult to explain. Overall, despite the practical and commercial applications of animal cells there are only few reports on their kinetics of growth and production. More specifically there is no literature report on the kinetic parameters for Chinese hamster ovary (CHO) cell batch culture related to mAb production [34].

The specific attributes of large scale mammalian cell cultures, e.g. high level of spatial heterogeneity and sensitivity of organisms to physical environmental stimuli, demand a modeling framework which captures both biology and hydrodynamics of the system and also their interactions. A model of a stirred tank bioreactor should consist of multiple components in order to represent fluid flow, mass transfer, gas-liquid interaction, cell metabolism and cell source variability. The state and input variables of the model can be divided into four groups. The first group includes the operational parameters such as agitation, aeration rate and temperature; the second group consists of environmental parameters such as pH, concentrations of chemicals inside the bioreactor (i.e. constituents of the cultivation medium, dissolved oxygen and carbon dioxide); and the third group comprises of biochemical (intracellular) parameters that can be used to describe the metabolic state of cells including cell mass composition, enzymes and proteins. Finally the fourth group includes macro-biological parameters to take into account contamination, degeneration, aggregation, and/or mutation [129]. The dynamics of cell and environment interaction should be modeled based on understanding of the response times of biological and physical mechanisms. In other words, a measure of rates of different mechanisms makes it possible to determine the bottleneck

mechanism. Prokop compared characteristic times of physical and biological mechanisms [20]. Characteristic time is a measure of the time needed by the mechanism to smooth out a change to a certain extent. For example, physical mechanisms of mixing, oxygen transfer to the liquid phase, and diffusion are faster compared to biological mechanisms of oxygen and substrate consumption. Analysis of characteristic times of mechanisms and identification of bottleneck mechanisms are very important in modeling and control of bioreactors. From this evaluation one can decide which mechanism to be further investigated. The next chapter provides a framework which captures the interactions of physiology and hydrodynamics. The framework is devised to take into account the inherent dynamics of the system, maintain computational tractability, and couple the model with optimization solvers. This is an improvement on the current state of lumped-parameter bioreactor modeling.

3 Development of dynamic, integrated and computationally feasible bioreactor model

The objective is to improve the reliability of bioreactor models by capturing the effects of hydrodynamics on its performance. Computational Fluid Dynamics (CFD) simulation is the established approach for solving the system of equations formed from conservation laws and thermodynamics numerically and calculation of necessary attributes of flow. Due to the inherent dynamics of biological processes, the hydrodynamics of the system has to be solved dynamically. For solving CFD simulations dynamically, time is discretized using step sizes in the range of 0.01 to 0.1 seconds [98, 105, 130-132] even for a small reactor. Wang et al. reported CPU time of 12 seconds per CPU for each time step [130]. Kerdouss et al. reported 0.5 to 2 seconds per CPU for each iteration [98]. Considering the reported CPU times and the fact that bioreactors are usually operated in fed-batch mode for up to two weeks lead to the conclusion that dynamic CFD simulation of the entire operation is computationally infeasible. To tackle this problem, a two-step approach is taken [133]. This is based on the assumption that the effects of metabolic activities on hydrodynamics are negligible [48, 134]. First the steady state of the two phase flow inside the bioreactor under defined operating conditions is calculated. Then the problem is solved dynamically to obtain the evolution of biophase over time. In the dynamic phase of the simulation only species conservation is considered. Further improvement in computational feasibility in order to simulate fed-batch operation and study effects of composition and schedule of feeding on the performance of bioreactor is achieved through development of a compartmental model. The idea of this methodology is division of reactor into zones well-mixed enough not to contain segregated regions and calculation of fluxes between the zones based on expected flow patterns resulting from CFD simulations or experimental data [48, 135-137]. Compartmental modeling facilitates time-space decomposition which reduces computation significantly. Therefore it has been widely used for

modeling the hydrodynamics of stirred tanks [136-139]. A number of admissible states for operation are defined by discretizing process parameters of impeller rotation speed, gas sparging flow rate and operating volume. Data on fully developed steady state flow under all pre-defined operation states are obtained and stored in flow matrices. Flow matrices contain information on inlet and outlet fluxes, gas volume fraction, dissipation rate of mechanical energy, and gas superficial velocity of compartments and characterize the flow inside the reactor under specific operating conditions. Change in operating conditions is simulated by replacing the flow matrices of the current state of operation with those associated with the new state. This is based on the assumption that the time for the flow to reach a new steady state as a result of change in operational conditions is negligible compared to the total processing time, i.e. the settling time for hydrodynamics is small. It should be noted that compartmental modeling provides an approximation of the solution and as such it predicts more homogenous distribution for species since each compartment is homogeneous. It also does not account for diffusive mass transfer. On the other hand, it makes it possible to take into account hydrodynamics in dynamic analysis of reactor performance and couple the model with optimization solvers. The rest of this chapter explains the road map to develop the integrated model and use it for improvement of bioreactor operation. The integration of hydrodynamics with metabolism involves capturing the effects of dissolved oxygen (DO) concentration, bubbles and turbulent eddies on metabolic activities and viability of cells. In this chapter first the steps of development of the integrated model are explained. The final section of this chapter presents a case study for application of this framework.

3.1 Development of CFD simulations

The flow of a single phase, gas or liquid, is described by conservation laws of mass, momentum, energy, charge, etc. If thermodynamic, transport and chemical properties of a

component need to be specified these field equations may be accompanied by the constitutive equations of state, stress, chemical reactions, etc. These equations are built into commercial CFD packages such as ANSYS® Fluent®. Fluent® solves the system of equations numerically using finite volume methodology which discretizes the space into computational cells. Derivation of the field and constitutive equations of multi-phase flow inside a bioreactor requires consideration of local characteristics. This is not straightforward due to the presence of interfacial surface. The difficulty is the result of unknown motions of multiple deformable interfaces, variables' fluctuations due to turbulence and moving interfaces and discontinuity of properties at the interface. It has been concluded that by obtaining mean values of flow properties through proper averaging local instantaneous fluctuations are eliminated. Three methodologies have been introduced for averaging: Eulerian, Lagrangian and Boltzmann statistical averaging. The Eulerian approach takes time and space coordinates as independent variables and other variables are expressed with respect to them. In the Lagrangian description, particle coordinates replace spatial coordinates, which gives clear advantage to this method if the behavior of individual particles is of interest. On the other hand if the focus is group behavior of particles the Eulerian approach is preferred [94]. Tracking individual bubbles imposes extra computational cost and would improve model's predictive power only if there was sufficient knowledge on interactions between individual bubbles and the liquid phase; i.e. growth, breakage and agglomeration of bubbles and energy dissipation due to bubble rupture. Therefore gas and liquid phases are considered as continuums and Eulerian averaging is used in this simulation. The Eulerian Multiphase model implanted in Fluent creates sets of momentum and continuity equations for each phase and couples them through exchanging pressure and interphase coefficients [140]. Turbulence of flow is captured using the k-ε viscosity model which has been widely used for stirred tanks [21]. It has been recognized as a

robust and computationally economic model which gives reasonably accurate results for a wide range of turbulent flows [141]. A k-ε model consists of two transport equations; one for turbulent kinetic energy (k) and one for energy dissipation rate (ε). Motion of impeller is captured using multiple reference frame (MRF). To implement the MRF model the geometry is broken up into stationary and moving zones. The MRF model approximates the flow in the moving zone around impeller by freezing the motion of the moving part in a specific position and observing the instantaneous flow field. To use flow variables of one zone for calculation of fluxes at the boundary of the adjacent zone, a local reference frame transformation is performed at the interface between cell zones. In the absence of large-scale transient effects due to weak impeller-wall interactions, the MRF approach provides a reasonable approximation of the flow [141].

3.2 Development of the integrated model

The state of operation is defined by operating volume, impeller rotation speed, and gas sparging flowrate. For every admissible state a separate CFD simulation is developed to obtain the steady state flow under associated operating conditions. After the solution is converged, the flow information of computational cells are extracted. This information contains spatial coordinates, turbulent energy dissipation rate, volume fraction of gas phase and mass flowrates to neighbor computational cells. Compartments are formed through agglomeration of computational cells. A compartment can be regarded as a fixed volume of liquid in space whose flow properties are determined by the operating conditions. Inlet and outlet fluxes, gas volume fraction, dissipation rate of mechanical energy, and gas superficial velocity of compartments are calculated using data obtained from CFD simulations. The surface between two neighbor compartments is composed of faces of computational cells. Mass flowrates across these faces

are summed to calculate flowrates between neighbor compartments (Figure 3-1). Fluent reports the rate of mass flow as a positive value if the flow direction is the same as the face area normal direction. A negative value means the flow direction and the face area normal directions are opposite. When exporting mass flowrates across faces, attention has to be paid to the direction of the face area normal. In Equation 3-1, $\dot{m}_{qq'}$ is the mass flowrate from computational cell q' to q and $\dot{M}_{ij}$ is the flowrate from compartment j to i under operating condition op.

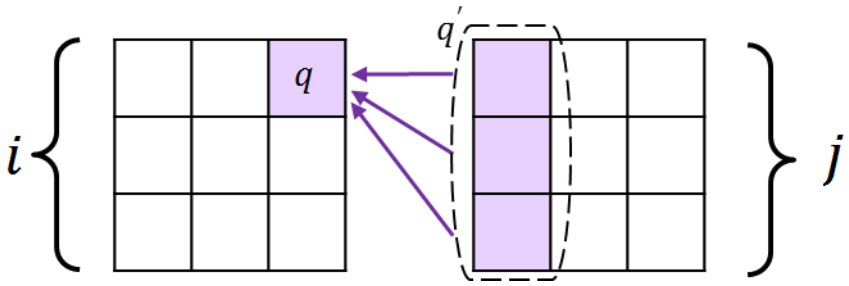

Figure 3-1: Mass flowrate between two neighbor compartments is calculated using flow properties of computational cells at the boundary

$$(\dot{M}_{ij})_{op} = \sum_{q\epsilon i}\sum_{q'\epsilon j}(\dot{m}_{qq'})_{op} \quad (\dot{m}_{qq'} > 0) \tag{3-1}$$

Net flow between two neighbor compartments is not necessarily zero, but net flow for each compartment should be zero in order to maintain mass conservation. Fluent does not report flowrates for faces at the boundaries. Boundary faces are located at surface of meshing zones. Although very small, these missing values cause mass imbalance and introduce errors. These errors are smoothed out by making the smallest possible changes to calculated flowrates. It is assumed that the calculated values for mass flowrate from compartment j to i, $\dot{M}_{ij}$, has error ε_{ij}. The sum of squared errors is minimized by solving the problem explained by Equations 3-

2 to 3-4. The second constraint makes sure that zero elements will remain zero, i.e. flowrates between non-neighbor compartments remain zero.

$$\underset{\varepsilon_{ij}}{minimize} \sum_i \sum_j {\varepsilon_{ij}}^2 \tag{3-2}$$

Subjected to:

$$\sum_i \dot{M}_{ij} - \varepsilon_{ij} = \sum_i \dot{M}_{ji} - \varepsilon_{ji} \tag{3-3}$$

$$-\dot{M}_{ij} \leq \varepsilon_{ij} \leq \dot{M}_{ij} \tag{3-4}$$

After obtaining corrected mass flowrates, $\widehat{\dot{M}}_{i,j}$, a system of ordinary differential equations (ODEs) is formed to calculate changes in viable cell densities in compartments, X_i, due to the liquid flow (Equation 3-5). In Equation 3-5, V_i is the volume of the liquid phase in compartment i.

$$\begin{bmatrix} \frac{dX_1}{dt} \\ \vdots \\ \frac{dX_c}{dt} \\ \vdots \\ \frac{dX_N}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{\sum_{i=2}^{N} \widehat{M}_{i,1}}{\rho_l V_1} & \cdots & \frac{\widehat{M}_{1,c}}{\rho_l V_1} & \cdots & \frac{\widehat{M}_{1,N}}{\rho_l V_1} \\ \vdots & \ddots & \vdots & & \vdots \\ \frac{\widehat{M}_{c,1}}{\rho_l V_c} & \cdots & -\frac{\sum_{i=1}^{N} \widehat{M}_{i,c(i\neq c)}}{\rho_l V_c} & \cdots & \frac{\widehat{M}_{c,N}}{\rho_l V_c} \\ \vdots & & \vdots & \ddots & \vdots \\ \frac{\widehat{M}_{N,1}}{\rho_l V_N} & \cdots & \frac{\widehat{M}_{N,c}}{\rho_l V_N} & \cdots & -\frac{\sum_{i=1}^{N-1} \widehat{M}_{i,N}}{\rho_l V_N} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_c \\ \vdots \\ X_N \end{bmatrix} \tag{3-5}$$

The impact of agitation on homogeneity of distribution of cells between compartments is understood if sedimentation of cells is included in the model. In Equation 3-6, v^{sed} is sedimentation velocity which is estimated as $r^2/4$ (mm/h) and r is cell radius in μm [142]. Cell radius is estimated assuming spherical shape, density equal to water's and average mass of

1.1165×10^{-6} mg [47] for cells. c' and c'' are compartments below and above c. $A_{c,c'}$ is the contact area between two neighbor compartments.

$$\frac{dX_c}{dt} = \frac{v^{sed}}{V_c}\left(A_{c'',c}.X_{c"} - A_{c,c'}.X_c\right) \quad (3\text{-}6)$$

It has been reported that power input greater than 22,500 $W.m^{-3}$ causes cell damage [123, 143]. Power input is calculated as the product of density of the liquid phase (kg/m^3) and turbulent energy dissipation rate (m^2/s^3). Turbulent energy dissipation rate is the rate of absorption of kinetic energy which breaks large sized eddies to smaller eddies until it is converted to heat by viscous forces [144]. The rate of cell damage of 3.4 $\%.min^{-1}$ has been reported for cells in high shear regions [123]. Volume fraction of high shear region in compartments is calculated using the data obtained from CFD simulations. Since cells are assumed to be homogenously distributed inside compartments, the volume fraction of the high shear region is equal to the fraction of cells exposed to shear beyond the tolerable threshold. Therefore the rate of loss of viable cells under operating condition op and in compartment c is calculated using Equation 3-7.

$$k_{d,shear_{op,c}} = 0.034.\, volume\ fraction\ of\ high\ shear\ region_{op,c} \quad (3\text{-}7)$$

Interaction with bubbles has also been reported as one of the sources of cell loss. Cells attach to bubbles, rise with them to the surface, become trapped in the foam layer and perish. Also maximum energy dissipated due to bubble rupture is two or three order of magnitude higher than the tolerable threshold for cells [123]. No value has been reported in the literature for rate of cell loss due to interaction with bubbles. One motive for integrated modeling is to capture

uncertainty where it occurs. Therefore, the rate of cell loss due to interaction with bubbles is estimated by assuming an interaction vicinity around bubbles. It is assumed that fraction of cells in a compartment which are in the vicinity of bubbles are lost over the average lifespan of a bubble in the compartment. The volume of bubble vicinity is calculated using the reported average bubble diameter of 0.00289 m [145] and considering a critical distance from surface of bubbles. The critical distance is assumed to be 0.5 mm. Number of bubbles in compartment c under operating condition op is estimated using local gas holdup obtained from CFD simulations and the average bubble diameter (Equation 3-8). Average bubble lifespan in compartment c under operating condition op is estimated from equation 3-9 using local gas holdup and air sparging flowrate. Equation 3-10 shows the estimated rate of loss of cells due to interactions with bubbles under operating condition op in compartment c.

$$number\ of\ bubbles_{op,c} = \frac{volume\ fraction\ of\ gas\ phase_{op,c}.volume_c}{bubble\ volume} \tag{3-8}$$

$$bubble\ lifespan_{op,c} = \rho_{air}\frac{volume\ fraction\ of\ gas\ phase_{op,c}.volume_c}{air\ sparging\ flowrate} \tag{3-9}$$

$$k_{d,bubble_{op,c}} = \frac{\ln(1-\frac{volume\ of\ the\ interaction\ vicinity \times number\ of\ bubbles_{op,c}}{V_c})}{bubble\ lifespan_{op,c}} \tag{3-10}$$

Viable cell density (VCD) in compartment c is calculated using Equation 3-11. μ and μ_d are metabolic rates of growth and death. They are calculated as functions of metabolites' concentrations using the metabolic model. c' and c'' are compartments below and above c. N is the total number of compartments. It can be seen that the hydrodynamics part of Equation 3-11 is a linear system of ODEs of rank N. The discretization of space provided by

compartmental modeling allows application of proper orthogonal decomposition (POD) for development of a reduced order model (ROM) [146]. In order to achieve this, first the full rank system is solved to create time-series snapshots of distribution of cells over compartments under a specific operating condition. Then a set of orthonormal bases are generated through eigen-decomposition [147]. Bases with no significant impact on the solution profile are truncated to obtain the reduced rank model. A system with $N = 16$ was used to evaluate the performance of the ROM developed with this methodology. The ROM showed satisfactory performance while the reduction in the rank of the system was small. The impact of the initial condition used for generating snapshots became apparent when fewer basis functions were used for approximation. For large reactors with greater number of compartments, however, it is recommended to investigate application of this methodology.

$$\frac{dX_c}{dt} = \left(\mu - \mu_d - k_{d,shear}{}_{op,c} + k_{d,bubble}{}_{op,c} - \frac{v^{sed}.A_{c,c'}}{V_c} - \frac{\sum_{i=1}^{N}(\widehat{M}_{i,c})_{op}}{\rho_l.V_c}\right)X_c + \frac{v^{sed}.A_{c'',c}}{V_c}X_{c''} + \frac{\sum_{i=1}^{N}(\widehat{M}_{c,i})_{op}.X_i}{\rho_l.V_c} \quad (3\text{-}11)$$

Contrary to cells, metabolites are assumed to be homogenously distributed at all time. This follows the fact that for the small geometry considered in this case study, fast diffusion dominates mass transfer. For larger bioreactors, local diffusive mass transfer has lower importance relative to convection so the flux matrix may be used for calculation of distribution of metabolites [48]. The dissolved oxygen (DO) concentration is also assumed to be homogenously distributed. For calculation of DO concentration, mass transfer and cellular uptake are considered. The ratio of rates of oxygen uptake to carbon dioxide production by cells has been reported to vary within a narrow range around 1 [148]. So oxygen uptake rate (OUR) is assumed to be 0.35 pmol.cell^{-1}.hr^{-1} which has been reported for carbon dioxide

production [149]. The overall volumetric mass transfer coefficient under operating condition op, $(k_L a)_{op}$, is calculated using Equation 3-12 in which $U_{G_{op,q}}$ is superficial gas velocity (m/s) for computational cell q under operating condition op. α and β are constant and have been experimentally estimated to be 0.412 and 0.809, respectively [105]. Volumetric mass transfer coefficient is the product of liquid phase mass transfer coefficient; k_L (m/s) and specific interfacial area; a ($m^2.m^{-3}$). Mass transfer stops after reaching the saturation concentration at 37°C. $3.43x10^{-5}$ has been reported for saturation mass fraction of oxygen [150]. The unstructured model is assumed to predict metabolites' uptake and production rates for an oxygen saturated culture. Experimental data are available which report these rates at different concentrations of dissolved oxygen [151]. The reported rates are divided by the rate obtained at saturated level of oxygen to calculate correction factors for different concentrations of DO (Figure 3-2). To take into account the effects of mass transfer mechanism on biological activities, values for metabolites' uptake and production rates predicted by the unstructured model are multiplied by correction factors. The direct effects of DO concentration on cellular rates of growth and death have not been reported. In this study, to penalize depletion of DO when solving for optimal aeration policy, it is assumed that growth stops when DO is depleted.

$$(k_L a)_{op} = \frac{\sum_q \alpha {U_{G_{op,q}}}^{\beta}.gas\ volume\ fraction_{op,q}.volume_q}{\sum_q volume_q} \quad (3\text{-}12)$$

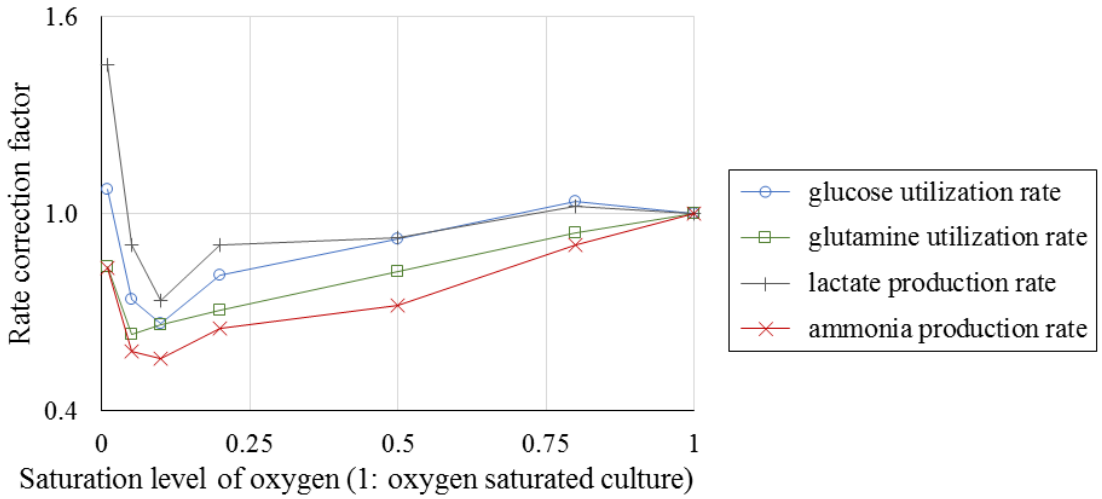

Figure 3-2: Effects of DO concentration on uptake and production rates of metabolites

The behavior of the bio-phase is predicted through incorporation of an unstructured metabolic model and consideration of the effects of environmental parameters on viable cell density. For the purpose of this study, a metabolic model from the literature is adapted to represent cellular growth and death rates as functions of concentrations of metabolites. As discussed, due to lumped nature of unstructured models, estimated values of their parameters have narrow confidence intervals. Moreover, these values have to be re-estimated when the model is used to predict the dynamics of a different bioreactor system. An important group of parameters are threshold metabolites concentrations that separate growth and death rates into different regimes. Threshold concentrations are determined by observing how viable cell density reacts to concentrations of metabolites. Capturing the dynamic behavior of the system sometimes requires considering multiple phases for cellular growth and death during which cells react differently to environmental stimuli. Xing et al. assumed the death phase begins after viable cell density declines by 10% from its peak value [38]. It was assumed that cellular growth didn't happen during this phase. The integration with hydrodynamics incorporates additional

sources of cell loss into the model which impacts viable cell density profile. Therefore, integration with hydrodynamics demands parameters of the metabolic model to be re-estimated. In the final section of this chapter, a metabolic model is adapted to explain integration of physical and biological processes for a case study. The presented results presented are merely meant to demonstrate the capabilities of the modeling framework.

3.3 Coupling the model with nonlinear solvers

Maximization of bioreactor yield is achieved through manipulation of process parameters based on the determined optimal operating policy. Reduction of the order of the model through compartmental modeling provides the formulation and modeling environment necessary for coupling the model with nonlinear or mixed-integer nonlinear programming solvers. There are two basic categories of optimization algorithms that use reduced models that are differentiated based on their dependencies on the original detailed model for calculation of gradients [14]. For the proposed formulation most of the evaluations and construction of the reduced models are done before starting the optimization since minimal communication between the optimization algorithm and the original model is preferred to reduce the computational complexity. Advanced nonlinear optimization algorithms are capable of handling numerous decision variables and constraints. However, the inherent dynamics of this problem makes it challenging. The classical approach to dynamic optimization problems takes advantage of Pontryagin's maximum principle and maximizes the control Hamiltonian over the set of all admissible controls [152]. The application of this approach becomes difficult for larger systems with state constraints [153] so direct approaches based on parameterization of variables have been preferred [154, 155]. The method of collocation has been proposed for parameterization of variables [156, 157]. Stiffness of the system of ordinary differential equations (ODE) is evaluated for different initial values. Stiffness ratio is calculated using the eigenvalues of the

Jacobian matrix at different points in time [158]. It is observed that the order of magnitude of stiffness ratio varies between 6 and 28 throughout the integration. Therefore, an approach based on discretization of time using fixed step sizes does not provide good approximation of the solution unless the step size is very small, i.e. less than 10^{-9} hour. Instead, integration is carried out using appropriate solvers for stiff, nonlinear ODE's and the problem is formulated for application of the interior point method [159]. Equations 3-13 to 3-16 represent the optimization problem. As discussed, there is no reliable literature report on the kinetic parameters of mAb production for Chinese hamster ovary (CHO) cell batch culture. Therefore, instead of final protein concentration, the time-integrated value of biomass is the subject of maximization. T_i and $\mathbf{C}_i$ are time and composition of i^{th} feeding. In addition to initial nutrients' concentrations, schedule and composition of feeding, it also finds near optimal criteria for setting aeration and agitation rates. Aeration is stopped or started based on DO level, i.e. if DO concentration falls below the criterion aeration starts, otherwise it stops. For adjustment of impeller rotation speed, a measure of homogeneity is defined based on relative standard deviation (RSD) of distribution of cells over compartments. If the value of RSD is greater than the criterion, the impeller speed increases to improve homogeneity of distribution of cells.

$$\underset{\boldsymbol{C}_0, T_i, \boldsymbol{C}_i, RSD_{cri}, DO_{cri}}{\text{maximize}} \int_0^{t_f} \text{biomass}\, dt \tag{3-13}$$

Subjected to:

Process Model: Equation 6

Control Bounds:

$$0 < \cdots < T_{i-1} < T_i < T_{i+1} < \cdots < t_f \tag{3-14}$$

$$\boldsymbol{C}_{i_{min}} \leq \boldsymbol{C}_i \leq \boldsymbol{C}_{i_{max}} \tag{3-15}$$

$$0\% \leq RSD_{cri}, DO_{cri} \leq 100\% \tag{3-16}$$

3.4 Case study

In this section, the framework explained previously is implemented to develop the bioreactor model and understand interactions of system components.

3.4.1 Model development

A 3 L bioreactor with Rushton impeller and orifice sparger is considered. Figure 3-3 shows the geometry of the reactor. To develop the discrete space for admissible operating conditions agitation rates of 150, 225 and 300 RPM and two options for aeration are considered; spraging air at 0.01 vvm (gas volume flow per unit of liquid volume per minute) and no aeration. The liquid fill levels of 130, 155, 180 and 205 mm are considered. Meshing for the highest operating volume results in 363,175 computational cells, 947,988 faces and 226,971 nodes. Feeding is modeled as change in the fill level. So overall three steps of feeding is allowed throughout the operation. The duration of operation is two weeks. Table 3-1 and Figures 3-4 to 3-6 show physical properties of the flow under different operating conditions calculated from CFD simulations. As expected, increase in the impeller rotation speed results in higher rate of dissipation of mechanical energy and liquid velocity. It also increases local gas holdup as seen in Figure 3-6. This is in conformance with the experimental observations reported in the literature [160]. However the overall gas holdup is affected by the thickness of the foam layer at the surface of the liquid phase.

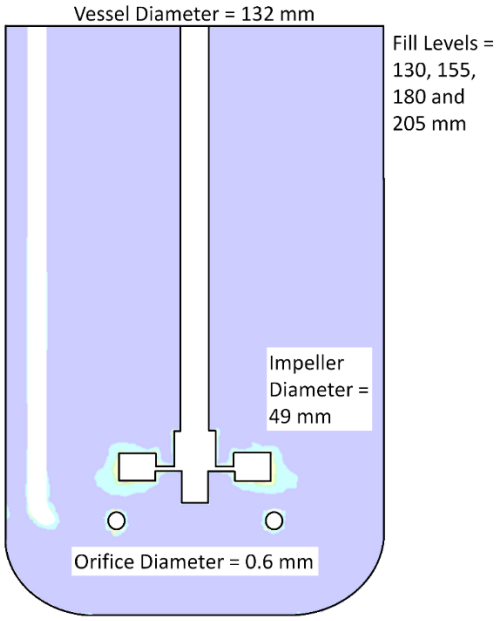

Figure 3-3: Geometry of the vessel considered for the CFD simulations

Table 3-1: Physical properties calculated from CFD simulations

Fill Level (mm)	Impeller Speed (RPM)	No Aeration	Aerated System		
		Power Input ($W.m^{-3}$)	Power Input ($W.m^{-3}$)	Overall Gas Holdup	Volumetric Mass Transfer (hr^{-1})
130	150	3.8	4.0	0.10	4.0
	225	11.0	12.0	0.18	4.6
	300	24.5	26.3	0.14	11.4
155	150	3.2	3.4	0.09	3.8
	225	9.5	10.4	0.25	4.2
	300	21.6	24.8	0.16	6.9
180	150	2.7	3.2	0.09	3.5
	225	8.4	9.4	0.14	4.3
	300	19.2	22.0	0.22	5.9
205	150	2.3	3.9	0.08	3.0
	225	7.0	7.9	0.13	3.5
	300	16.1	17.2	0.18	5.3

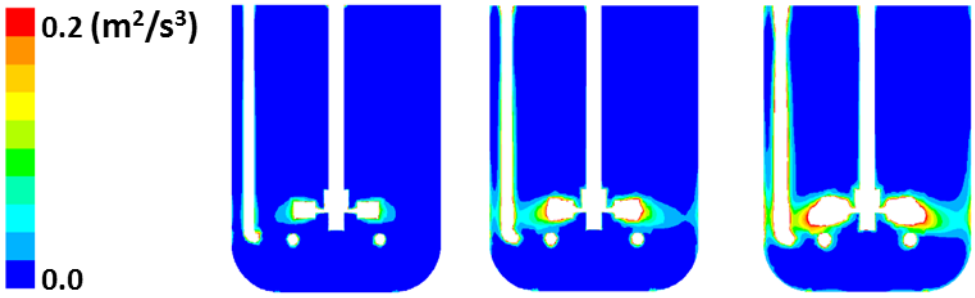

Figure 3-4: Dissipation rate of mechanical energy for liquid fill level of 180 mm and aeration rate of 0.01 vvm, from left to right impeller speeds are 150, 225 and 300 RPM

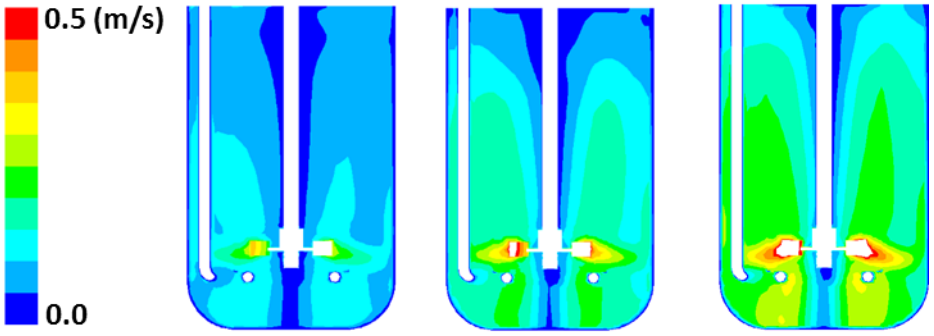

Figure 3-5: Velocity of liquid phase for liquid fill level of 205 mm and aeration rate of 0.01 vvm, from left to right impeller speeds are 150, 225 and 300 RPM

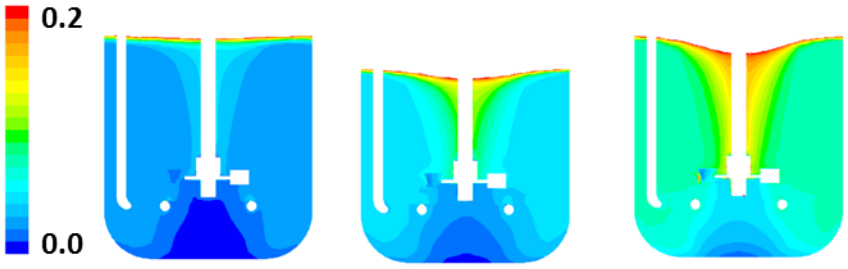

Figure 3-6: Gas volume fraction for liquid fill level of 155 mm and aeration rate of 0.01 vvm, from left to right impeller speeds are 150, 225 and 300 RPM

Computational cells are agglomerated to form compartments. The configuration of compartments is shown in Figure 3-7. The data obtained from CFD simulations are used to calculate flow properties of compartments under all states of operation. These properties include mass flowrates between compartments, volume fraction of region in which mechanical shear is higher than the tolerable threshold for cells, estimated number of bubbles, average time a bubble spends in a compartment, and overall volumetric gas-liquid mass transfer coefficient. The unstructured model developed by Xing et al. [38] is adapted to predict behavior of the bio-phase. The metabolic model captures the effects of concentrations of glucose, glutamine, lactate, and ammonia on cellular rates of growth and death in a CHO culture. The rate of utilization of glutamine for essential metabolic functions, i.e. maintenance, is calculated from Equation 3-17. The values of model parameters are shown in Table 3-2.

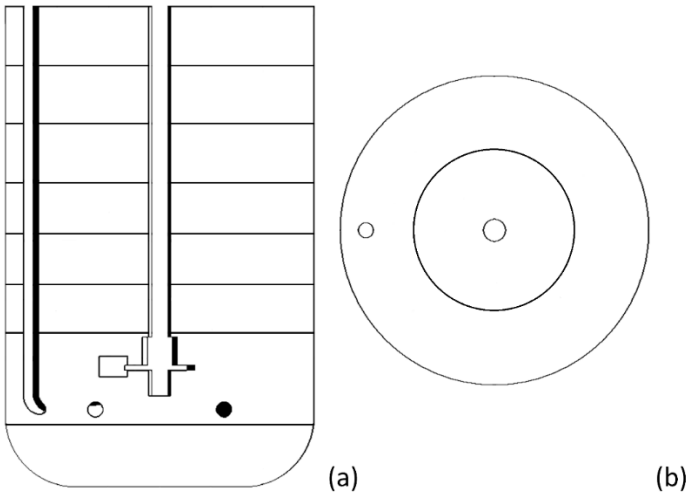

Figure 3-7: Arrangement of compartments; side (a) and top (b) views

Table 3-2: Unstructured model parameters

Parameter	Value	Unit
m_{Glc}	$69.2e-12$	$mmol.cell^{-1}.hr^{-1}$
a_1	$3.2e-12$	$mmol.cell^{-1}.hr^{-1}$
a_2	2.1	mM
μ_{max}	0.029	hr^{-1}
$\mu_{d_{max}}$	0.016	hr^{-1}
K_{Glc}	0.084	mM
K_{Gln}	0.047	mM
KI_{Lac}	43	mM
KI_{Amm}	6.51	mM
KD_{Lac}	45.8	mM
KD_{Amm}	6.51	mM
d_{Gln}	$7.2e-3$	hr^{-1}
$Y_{X/Glc}$	$1.69e\ 8$	$cell.mmol^{-1}$
$Y_{X/Gln}$	$9.74e\ 8$	$cell.mmol^{-1}$
$Y_{Lac/Glc}$	1.23	$mmol.mmol^{-1}$
$Y_{Amm/Gln}$	0.67	$mmol.mmol^{-1}$
DO_{eq}	1.0699	mM
OUR	$3.5e-10$	$mmol.cell^{-1}.hr^{-1}$

$$m_{Gln} = \frac{a_1\,[Gln]}{a_2+[Gln]} \tag{3-17}$$

Cellular rates of growth and death are calculated knowing the Monod constants ($K_{metabolite}$) and using Equations 3-18 and 3-19, respectively. Growth and death rates are then used in Equation 3-11 to calculate viable cell densities in compartments.

$$\mu = \mu_{max} \frac{[Glc]}{K_{Glc}+[Glc]} \frac{[Gln]}{K_{Gln}+[Gln]} \frac{KI_{Lac}}{KI_{Lac}+[Lac]} \frac{KI_{Amm}}{KI_{Amm}+[Amm]} \quad (3\text{-}18)$$

$$\mu_d = \mu_{d\,max} \frac{[Lac]}{KD_{Lac}+[Lac]} \frac{[Amm]}{KD_{Amm}+[Amm]} \quad (3\text{-}19)$$

The estimated value for overall volumetric mass transfer coefficient under present operation conditions is used in Equation 3-20 to calculate concentration of DO, where X is the overall viable cell density.

$$\frac{\mathrm{d}[DO]}{\mathrm{dt}} = (k_L a)_{op}\big(DO_{eq} - [DO]\big) - X.OUR \quad (3\text{-}20)$$

Concentrations of metabolites are calculated from Equations 3-21 to 3-24 by knowing the values of yield parameters (Y's). The impact of DO on uptake and production rates ($DO_{metabolite}$) is estimated through spline interpolation of experimental data shown in Figure 3-2. The model also captures chemical degradation of glutamine.

$$\frac{d[Glc]}{dt} = -DO_{Glc}\left(\frac{\mu-\mu_d}{Y_{X/Glc}} + m_{Glc}\right)X \quad (3\text{-}21)$$

$$\frac{d[Gln]}{dt} = -DO_{Gln}\left(\frac{\mu-\mu_d}{Y_{X/Gln}} + m_{Gln}\right)X - d_{Gln}\,[Gln] \quad (3\text{-}22)$$

$$\frac{d[Lac]}{dt} = DO_{Lac}.Y_{Lac/Glc}\left(\frac{\mu-\mu_d}{Y_{X/Glc}} + m_{Glc}\right)X \quad (3\text{-}23)$$

$$\frac{d[Amm]}{dt} = DO_{Amm}.Y_{Amm/Gln}\ \frac{\mu-\mu_d}{Y_{X/Gln}}X \qquad (3\text{-}24)$$

3.4.2 Results

Definition of states for operation, creates a finite set of admissible actions for every point in time. Depending on the present operational conditions it is possible to increase or decrease impeller rotation speed, stop or start air sparging and feed or not to feed. It is also possible to continue under current conditions. Changing the state of operation is limited to once in every 2 hours. The concentrations of glucose and glutamine, initially and in feed, are constrained to be under 100 and 10 mM, respectively. Solving the optimization problem described in Equations 3-13 to 3-16, suggests feeding at 210, 255 and 305 hours after the start of operation. Nutrients concentrations stay at their upper bounds due to limited number of feeding steps. The criterion for aeration is 85% of saturated DO concentration (DO_{cri} = 85), i.e. aeration starts if DO concentration falls below this value or otherwise stops. The obtained criterion for agitation is 0.04% (RSD_{cri} = 0.04). Impeller rotation speed is increased if the RSD of distribution of cells over compartments is greater than agitation criterion. In the opposite case, where RSD has smaller value, lower agitation rate is selected for the next 2 hours. Although the existence of multiple local optima cannot be ruled out, the solution shows improvement in the yield of operation through manipulation of feeding schedule and composition, agitation and aeration rates. Figure 3-8 compares the reactor operated under the near optimal policy with a reactor fed with the same composition at 84, 168, 252 hours after the start of operation and is constantly aerated and agitated at 300 RPM, i.e. the criteria for aeration and agitation are 100% DO saturation and RSD of 0%, respectively. Schedule of feeding should be determined with consideration of capacity of the reactor and duration of operation. Early addition of feed, despite increasing cellular population, causes accumulation of lactate and ammonia in the

system, which further inhibits growth. Early utilization of reactor's capacity also risks nutrition depletion especially in late stages of operation when larger population of cells demand more nutrients. Late feeding, on the other hand, results in poor utilization of nutrients and reactor's capacity. Figure 3-8 shows through manipulation of feeding schedule, even with limited number of feeding steps, cells are provided with enough nutrients to maintain growth throughout the operation.

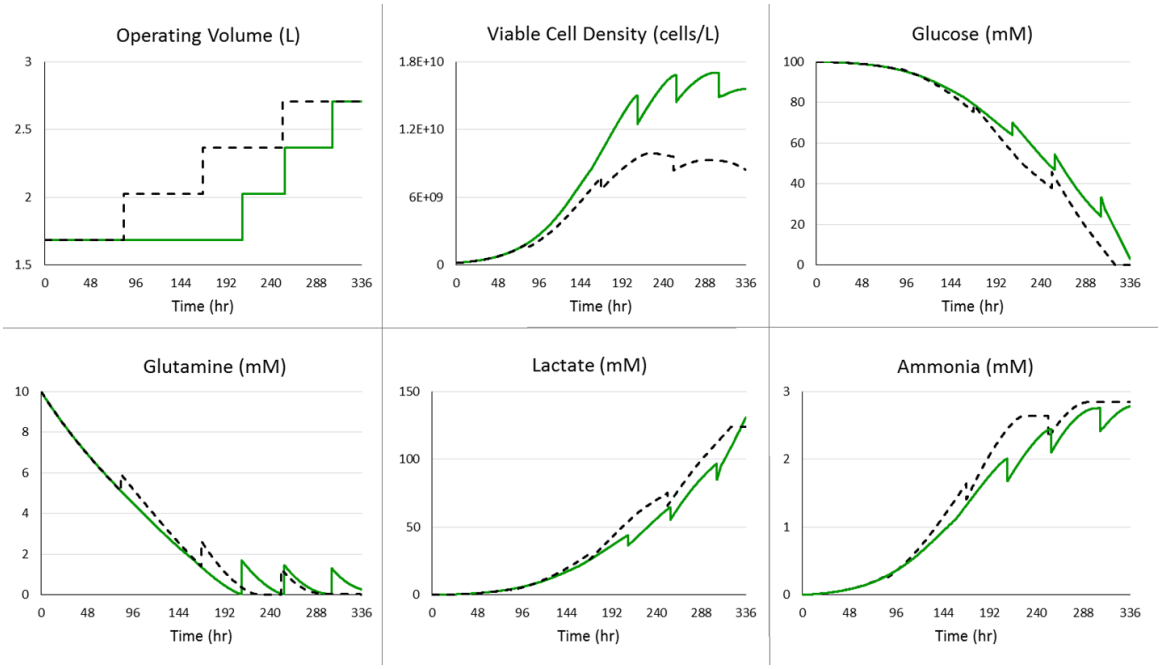

Figure 3-8: Effects of feeding schedule on the performance of bioreactor; near optimal policy (green), uniform feeding schedule (black)

Inclusion of mass transfer mechanism in the model leads to improvement in aeration and agitation policies. Figures 3-9 and 3-10 demonstrate the effects of aeration and agitation criteria while feeding schedule and composition are at their near optimal values. DO concentration drops quickly when aeration stops because of fast consumption by cells. Stopping and starting aeration according to the near-optimal policy prevent loss of viable cells due to unnecessary

aeration. Figure 3-9 shows the near optimal policies obtained for agitation and aeration of the bioreactor result in low DO concentrations in the middle and also toward the end of operation when cell population is large (green curves). To evaluate the performance of the solver, the solution is compared with polices which result in higher agitation rates and more aeration throughout the operation. Figure 3-9 shows the reactor operated under DO_{cri} = 95 and RSD_{cri} = 0.03 (yellow curves). These values for aeration and agitation criteria increase the total aeration time from 250 to 274 hours. Under RSD_{cri} = 0.04, impeller rotation speed stays at its lowest value until the last feeding step. The lower value for RSD_{cri} results in more agitation. More agitation and aeration together increase the gas holdup which consequently increases the rate of cell loss due to interactions with bubbles. The increase in cell loss due to interactions with bubbles is apparent in the VCD plot, particularly after the first feeding step. Figure 3-9 also shows the system operated under DO_{cri} = 85 and RSD_{cri} = 0.03 (red curves). Higher agitation rate expedites the mass transfer between gas and liquid phases. Without increasing DO_{cri}, higher rate of mass transfer results in more frequent stopping of aeration. The aeration time reduces to 232 hours. It causes more drops in DO concentration before the first feeding step which slows down the growth. DO concentration also affects rates of utilization and production of metabolites. The effect on ammonia concentration is relatively more visible since the rate correction factor varies over wider range (Figure 3-2). Comparing yellow and red curves in Figure 3-9 shows that although DO depletion slows down the growth before the first feeding step under DO_{cri} = 85 and RSD_{cri} = 0.03, it is compensated in later stages of the process. This is due to less inhibitory effects of metabolites as lower DO concentration reduces the production rates of metabolites. Figure 3-10 shows the effects of aeration policy on performance of the reactor while feeding schedule and composition and RSD_{cri} are at their near optimal values. More aeration increases the inhibitory effects of metabolites and interactions

with bubbles (dark blue curves). However, without increasing the impeller speed, gas holdup remains low and loss due to interactions with bubbles does not increase significantly. Less aeration results in DO depletion which slows down the growth (light blue curves). However it will be partially compensated in later stages of operation as it reduces the interaction with bubbles and inhibitory effects of metabolites.

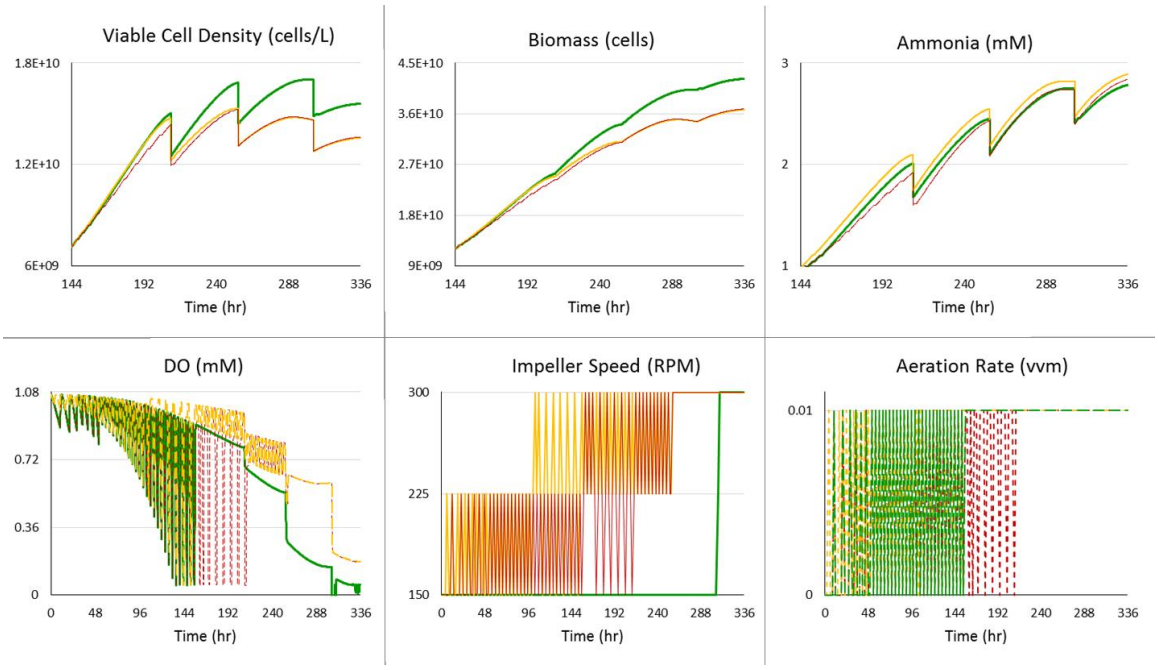

Figure 3-9: Effects of agitation and aeration policies on the performance of bioreactor; DO_{cri} *= 85 and* RSD_{cri} *= 0.04 (green),* DO_{cri} *= 95 and* RSD_{cri} *= 0.03 (yellow),* DO_{cri} *= 85 and* RSD_{cri} *= 0.03 (red)*

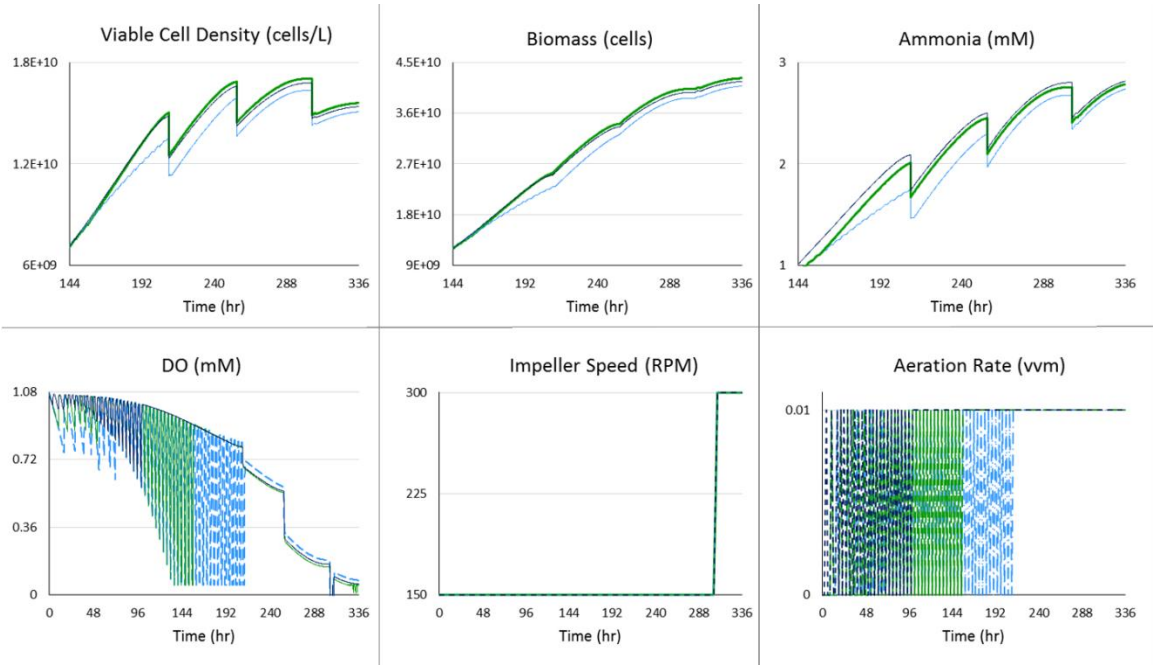

Figure 3-10: Effects of aeration policy on the performance of bioreactor; DO_{cri} = 85 (green), DO_{cri} = 95 (dark blue), DO_{cri} = 75 (light blue)

4 Development of surrogate models for hydrodynamics

To address major challenges of implementation of process models in biotechnology industry, i.e. lack of universality and narrow confidence intervals, the proposed framework for modeling cell cultures integrates biological and physical phenomena. The burden on dynamic solver is reduced by assuming the hydrodynamics at its fully developed state and implementation of compartmental modeling. A set of admissible operating states is defined by discretization of process parameters. Change in operation is followed by hydrodynamics switching to a new steady state that would be reached under new conditions. Operating condition is defined by process parameters of impeller rotation speed, gas sparging flow rate and operating volume. Data of the steady flows under considered operating conditions are obtained from computational fluid dynamics (CFD) simulations. The discrete space of admissible operating conditions can be expanded through finer discretization of process parameters. Alternatively, this chapter investigates exploitation of computer simulation data in order to replace the discrete space with a semi-continuous space. This is achieved through development of a surrogate model for hydrodynamics using data obtained from CFD simulations for computational cells. For CFD simulation of a 3L reactor, the space is discretized into 300 to 400 thousands computational cells. Therefore, the main challenge is to devise an efficient algorithm to explore the large sampling space. The rest of this chapter investigates application of kriging for prediction of velocity field as a function of liquid fill level under constant agitation and aeration rates.

4.1 Model development

CFD simulations are developed for different fill levels and constant agitation and spraging rates. In order to combine data obtained for different fill levels, the impeller is considered as

reference and the z coordinates of computational cells are scaled. In Equation 4-1, c is the fraction of maximum operating capacity of the reactor which is calculated knowing the fill level. z' varies between -1 and 1 independent of the fill level.

$$z' = \begin{cases} \frac{z - z_{impeller}}{z_{impeller} - z_{bottom}} & z \leq z_{impeller} \\ \frac{z - z_{impeller}}{(z_{max} - z_{bottom})c - (z_{impeller} - z_{bottom})} & z > z_{impeller} \end{cases} \quad (4\text{-}1)$$

The inputs of the model are fill level and spatial coordinates; $\boldsymbol{x_s} = [c, r, a, z']$. The model is intended to predict the 3 elements of velocity in cylindrical coordinates for a given point in space. Therefore 3 sets of kriging models are built. Development of surrogate models involve manipulation of correlation matrices of the size of nxn which requires $O(n^4)$ computations. After combining data obtained for 4 fill levels the size of dataset (n) exceeds a million points. To select a subset of data points a sampling method based on Latin hypercube designs (LHDs) is implemented. After decomposing the complete data into disjoint equally spaced blocks, a subsample is obtained by collecting blocks according to a randomly generated LHD. This method is called LHD-based block bootstrap. The LHD-based block bootstrap takes into account the spatial dependency and therefore improves the estimation accuracy for correlation parameters in Gaussian process models. Additionally, because of the one-dimensional balance properties of LHDs, the block bootstrap subsamples are spread out more uniformly over the complete data and the resulting subsample represent the complete data more effectively. It is also shown that LHD can result in variance reduction in estimation compared with simple random samples [161]. The sampling steps are repeated to obtain multiple subsets of data points. The subsample sets are used to train kriging models. The average of predictions by all the models is considered as the final prediction.

Kriging is an interpolation method that features the observed data at all sample points. The kriging model consists of a linear regression term to capture global trend and a stochastic error term which follows a Gaussian process (Equation 4-2). Simultaneous selection of variables for the regression term and parameter estimation improves the predictive power of the surrogate. Most likely candidates for variables of the regression function are considered as model inputs. In this case, $\boldsymbol{X_s} = \left[c, r, a, z', r^2, a^2, z'^2, ra, rz', az'\right]$. Equations 4-3 to 4-5 explain calculation of Gaussian function with zero mean and variance σ^2 from the correlation function R.

$$\boldsymbol{y_s} = \boldsymbol{X_s}^T\boldsymbol{\beta} + Z(\boldsymbol{x_s}) \tag{4-2}$$

$$Z \sim (0, \sigma^2) \tag{4-3}$$

$$Cov[Z(\boldsymbol{x_s}), Z(\boldsymbol{x_s}')] = \sigma^2 R(\boldsymbol{x_s}, \boldsymbol{x_s}') \tag{4-4}$$

$$R(\boldsymbol{x_s}, \boldsymbol{x_s}') = \exp(-\textstyle\sum_{i=1}^{m} \theta_i \left|x_{s_i} - {x_s}'_i\right|^{p_i}) \tag{4-5}$$

Obtaining parameters $\boldsymbol{\theta}$ and $\boldsymbol{p}$ and simultaneously evaluating regression basis functions included in $\boldsymbol{X_s}$, is achieved through maximizing a penalized likelihood function, Q, with spatial dependence incorporated into the penalty term. The penalized likelihood function is formed using smoothly clipped absolute deviation (SCAD) penalty (Equations 4-6 and 4-7) in which $a > 2$ [162]. Inclusion of the penalty term increases the computational demand. For computational efficiency an iterative method based on local linear approximation of the penalty function is implemented (Equation 4-8). With careful initial guess and choice of regularization

parameter for the penalty term, one step of iteration has been shown to provide a statistically efficient approximation of the coefficients [163]. These approximations are called one-step sparse estimates (OSE) of the coefficients. To obtain the initial guess for model parameters least square estimates of $\boldsymbol{\beta}$ and σ^2 (Equations 4-9 and 4-10) are used to form the likelihood function l (Equation 4-11). Then the likelihood function l is maximized. Finally, the model predicts the variable of interest for given input $\boldsymbol{x}$ using Equation 4-12 in which $\boldsymbol{r}$ is the correlation matrix between observed and untried points.

$$p_{\boldsymbol{\lambda}}(\beta_j) = \begin{cases} \lambda|\beta_j| & |\beta_j| > \lambda \\ \lambda^2 + (a-1)^{-1}\left(a\lambda|\beta_j| - \frac{{\beta_j}^2}{2} - a\lambda^2 + \frac{\lambda^2}{2}\right) & \lambda < |\beta_j| \le a\lambda \\ (a+1)\frac{\lambda^2}{2} & |\beta_j| > a\lambda \end{cases} \tag{4-6}$$

$$Q(\boldsymbol{\theta}, \boldsymbol{\beta}) = l(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{p}) - n\sum_{j=1}^{p} p_{\boldsymbol{\lambda}}(\beta_j) \tag{4-7}$$

$$Q^*(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{-1}{2\sigma^2}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta}) - n\sum_{j=1}^{p} p_{\boldsymbol{\lambda}}'(\beta_j)|\beta_j| \tag{4-8}$$

$$\boldsymbol{\beta}(\boldsymbol{\theta}, \boldsymbol{p}) \cong ({\boldsymbol{X_s}}^T \boldsymbol{R}^{-1} \boldsymbol{X_s})^{-1} {\boldsymbol{X_s}}^T \boldsymbol{R}^{-1} \boldsymbol{y_s} \tag{4-9}$$

$$\sigma^2(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{p}) \cong \frac{1}{n}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta}) \tag{4-10}$$

$$l(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{p}) = \frac{-1}{2\sigma^2}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta}) - \frac{1}{2}|\boldsymbol{R}| - \frac{n}{2}\log(\sigma^2) \tag{4-11}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{r}^T(\boldsymbol{x_s}, \boldsymbol{x})\boldsymbol{R}^{-1}(\boldsymbol{y_s} - \boldsymbol{X_s}\boldsymbol{\beta}) \quad (4\text{-}12)$$

4.2 Results

Four CFD simulations are developed for four liquid fill levels and agitation and spraging rates of 150 RPM and 2 mL.min^{-1}. z_{bottom} is -30 cm and z_{max} is 100, 125, 150 and 175 cm in the four simulations. $z_{impeller}$ is 20 cm. The four CFD simulations collectively provide 1,472,523 data points. Table 4-1 shows the summary of input and outputs. The sampling steps are repeated to obtain 10 subsets of data points. Table 4-2 shows the sizes of subsample sets. The subsample sets are used to train 10 kriging models. Regularization parameter for the penalty term is selected via cross validation. Different values between 1 and 2 are assumed for λ and 7-fold cross validation have been performed. Minor sensitivity to λ has been observed for all elements of velocity, therefore λ is assumed 1.5.

Table 4-1: Summary of input and outputs exported from CFD simulation

	c	r	a	z′	V_r (m/hr)	V_a	V_z (m/hr)
Minimum	0.63	0.0	0.0	-1.0	0.1	0.0	-777.8
1st Quintile	0.76	0.4	0.2	-0.4	241.5	95.2	-23.8
Median	0.76	0.6	0.5	-0.3	335.6	180.2	10.9
Mean	0.82	0.5	0.5	-0.1	335.1	179.8	19.3
3rd Quintile	0.88	0.7	0.7	0.3	419.2	263.8	69.4
Maximum	1.00	1.0	1.0	1.0	1664.2	360.0	3409.0

Table 4-2: Sizes of subsample sets obtained by LHD-based block bootstrap sampling used for training surrogate

Kriging Model #	Subsample Size
1	1041
2	1606
3	1130
4	1118
5	800
6	894
7	1400
8	1541
9	1149
10	1024

OSE of coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are provided in the appendix. Two kriging models are developed for every subsample set using OSE of parameters and DACE toolbox. The final prediction of each method is the average of 10 predictions. 5 smaller subsample sets are generated using LHD-based block bootstrap to test the performances of both methods. These test sets have 366, 304, 284, 436 and 265 data points. Tables 4-3 to 4-5 compare the performances of the two methods for prediction of radial, angular and axial elements of velocity respectively. For better comparison penalized errors are also calculated. Penalized error is calculated using Equation 4-13 to penalize predictions which are out of acceptable ranges for the output, i.e. $[\alpha,\beta]$.

$$\delta = |\hat{y}_i - y_i| + (P-1)\left[\frac{\alpha - \hat{y}_i + |\alpha - \hat{y}_i|}{2}\right] + (P-1)\left[\frac{\hat{y}_i - \beta + |\hat{y}_i - \beta|}{2}\right] \quad (4\text{-}13)$$

It is noticed that mean squared error (MSE) and penalized MSE are equal in all cases for OSE, which means the predictions are within acceptable ranges. Figure 4-1 visualizes the velocity fields ($V = \sqrt{{V_r}^2 + {V_z}^2}$) predicted by the two methods and compares them with the result of CFD simulation. The performance of the kriging model in capturing the spatial trends is significantly improved by using the OSE method. Simultaneous parameter estimation and variable selection produces more meaningful predictions and improves the performance of the kriging models compared to the DACE toolbox. However, the margin of error is high. Therefore, more investigation is required before using surrogate models for hydrodynamics.

Table 4-3: Comparison of radial velocity predicted by OSE method and DACE toolbox

	OSE			DACE toolbox		
	Mean Relative Error	Square root of Mean Squared Error	Square root of Mean Squared of Penalized Error (0<y)	Mean Relative Error	Square root of Mean Squared Error	Square root of Mean Squared of Penalized Error (0<y)
Test Set 1	0.6	92.5	92.5	0.3	62.5	62.5
Test Set 2	0.5	108.3	108.3	0.4	94.9	117.1
Test Set 3	0.8	97.6	97.6	0.7	102.3	149.2
Test Set 4	1.6	88.6	88.6	0.4	70.6	87.2
Test Set 5	0.4	86.7	86.7	0.2	61.6	72.9
Average	0.8	94.7	94.7	0.4	78.4	97.8

Table 4-4: Comparison of angular velocity predicted by OSE method and DACE toolbox

	OSE			DACE toolbox		
	Mean Relative Error	Square root of Mean Squared Error	Square root of Mean Squared of Penalized Error (0<y<360)	Mean Relative Error	Square root of Mean Squared Error	Square root of Mean Squared of Penalized Error (0<y<360)
Test Set 1	8.6	93.5	93.5	8.0	70.3	83.4
Test Set 2	0.7	58.7	58.7	0.5	50.2	56.7
Test Set 3	0.6	65.2	65.2	0.7	79.6	144.2
Test Set 4	0.7	69.9	69.9	0.7	69.5	201.5
Test Set 5	1.0	71.7	71.7	0.6	46.1	78.2
Average	2.3	71.8	71.8	2.1	63.2	112.8

Table 4-5: Comparison of axial velocity predicted by OSE method and DACE toolbox

	OSE		DACE toolbox	
	Mean Relative Error	Square root of Mean Squared Error	Mean Relative Error	Square root of Mean Squared Error
Test Set 1	9.5	59.2	19.4	65.9
Test Set 2	2.1	115.2	2.6	124.2
Test Set 3	3.3	62.4	4.2	66.4
Test Set 4	1.7	85.8	2.3	94.0
Test Set 5	3.6	88.6	4.0	94.1
Average	4.0	82.2	6.5	88.9

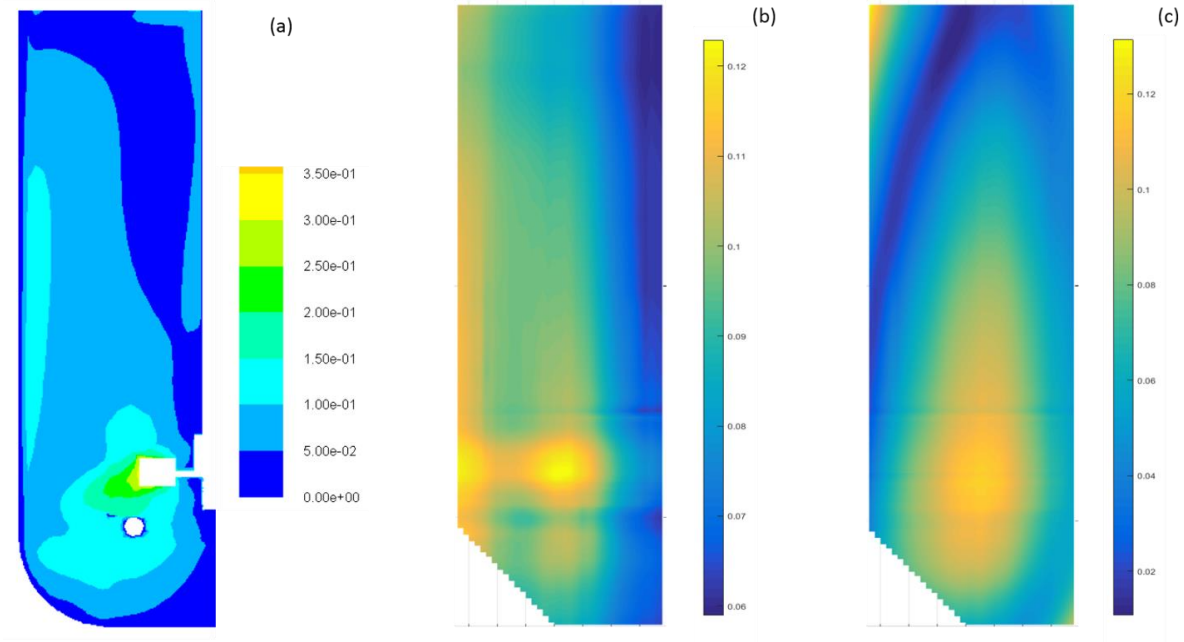

Figure 4-1: Velocity profile obtained from CFD simulation (a), OSE (b) and DACE toolbox (c) for liquid fill level of 205 cm and agitation and gas spraging rates of 150 RPM and 2 mL.min^{-1}

5 Concluding remarks

Mechanistic models can summarize the available process knowledge [164] and direct future knowledge generation efforts. However, moving toward implementation of model-based approaches in the biopharmaceutical industry has been limited [3]. Modeling of the biological processes represents the biggest challenge. For extremely simplified models significant variations and narrow ranges of validity are observed for the values of their parameters. This is partly because of the limited understanding of cellular mechanisms, which adversely affects the predictive capabilities of the models. A cell represents a complex set of interactions occurring with a hierarchy. This is further complicated by regulatory elements that interact within and throughout the hierarchy. On a cellular level, genes contain the essential information to run metabolic reactions. The information in the genes is first transcribed and then translated into functional units for reactions in the form of proteins and enzymes. Specific sets of reactions create metabolic pathways, which in turn are elements of metabolism. Regulation occurs within levels of metabolism and additionally it is interactive throughout the hierarchical process and therefore by analyzing gene expression data from a cell, it is difficult to predict the metabolic activity of a specific pathway. In order to understand cellular metabolism, genomic, transcriptomic and metabolomic data should be considered. This is however impractical for several reasons. The required set of data is not readily available to engineers optimizing and characterizing bioprocesses. That is attributed to high cost of running such analytical methods and also these types of assays have very limited throughput. Furthermore, additional statistical expertise is required just in order to analyze the data generated by these assays. Consequently, the most practical approach available is to examine the metabolism on a pathway level since this represents the net effect of all the higher hierarchical activities. This is a valid approach where the mathematical model is in agreement with the empirical data. The model is however

limited to assumptions and data set used to test the model. Currently, the data sets are limited to the cell type, strain, and/or clone. Additionally it is limited to the formulation of the media and process conditions. Currently these factors cannot be parameterized in order to fit a model, which can be universally applied. Novel techniques are constantly being developed which can be applied to bioprocesses to bridge this gap between where we are now versus what is required. These techniques are based on capturing information on the single-cell level. Additionally, they can improve the quantity and quality of the collected data. One such technique is called pulse response experimentation has been introduced as an effective way for obtaining kinetic information [165]. In this methodology perturbing the steady state of a chemostat is followed by fast sampling to characterize the changes in the process. Monitoring expression of individual genes has made addressing the physiological heterogeneity of cell populations possible [89]. Proper data analysis and modeling can channel the flow of single-cell level data and lead to understanding the cellular behavior at the population level. This is the reason that more effective communication between experimentalists and modelers has been called for in recent literature [13, 89, 166]. An algorithmic approach toward identification, estimation and specification of confidence intervals of parameters is required. Identifiability analysis helps with determining the parameters that can be estimated with high degree of confidence based on the available experimental data.

Another issue limiting the application of mechanistic models is lack of generality regarding the model parameters. Change of the process or operating conditions require re-estimation of model parameters as model outputs are insensitive to operating conditions. Affordability of computational power has increased the access of modeling software for a variety of uses. Commercially available computational fluid dynamics (CFD) modeling software are now available with the necessary capabilities to create representative simulations of all the

hydrodynamic conditions proceeding in the bioreactor. CFD represents the most advanced aspect of bioreactor modeling, since the underlying mathematics and engineering principles have been rigorously worked out. However the feasibility of those models even for laboratory scale bioreactors is still under question as a dynamic simulation of one minute of the operation may require hours of CPU time [98, 130]. The decomposition of time and space through compartmental modeling improves the computational feasibility significantly. Although an approximation, it facilitates integration of hydrodynamics and physiology of bioreactor which leads to sensitivity of bioreactor model to operating conditions. Quantification of the extent of effects of mechanical shear and bubble interactions on viability of cells is a valuable outcome which is achievable through integrated modeling, experimental design, and parameter estimation. Uncertainty and sensitivity analyses are required to understand variance of model outputs due to variability of different sources in its inputs. Integrated modeling results in better estimating cellular growth and death rates as functions of metabolites concentrations. It can address lack of universality of cell culture models, which is a major hindrance towards implementation of mechanistic models in biotechnology industry. Difficulty of validation of CFD simulations may be regarded as a drawback. An alternative to computer simulation can be development of compartmental models using experimental data directly. For measuring velocity fields the method of Particle Image Velocimetry (PIV) has been employed [137]. In this method, the movement of particles are measured between two light pulses. Then the subsections from each image frame are cross-correlated pixel by pixel. The correlation produces a signal peak, identifying the common particle displacement. It is capable of calculating all three components of velocity. For measuring local gas holdup, computed tomography (CT) has been used [160, 167].

Once a framework for integration of hydrodynamics and biological processes is in place, the components of the model can be selected based on the specific demands of the study. pH is an important process parameter in operation of a bioreactor since it impacts cell metabolism particularly glucose uptake and lactate production rates [168]. It is usually controlled via sparging CO_2 and base supplementation and is maintained above 7.0 [125]. Carbon dioxide is introduced into the liquid phase as the result of mass transfer and metabolic activities. The concentrations of carbon components in the liquid phase, i.e. aqueous carbon dioxide, carbonic acid, bicarbonate, and carbonate, are calculated knowing the gas-liquid mass transfer coefficient and equilibrium constants [169]. Similar to DO, the experimental data on utilization and production rates of metabolites at different values of pH have been published [151]. Incorporation of a buffer system into the model results in better representation of the system [133]. This is achieved at the cost of increasing the nonlinearity and rank of the ODE system.

Integration with hydrodynamics introduces new parameters to the cell culture model, i.e. sedimentation rate, tolerable shear threshold, rate of cell damage due to shear, bubble size distribution, interaction distance with bubbles, and mass transfer coefficients. The uncertainties in the values of these parameters can be included in dynamic analysis of the operation. Sensitivity of the solution profile to model parameters can be determined and uncertainty in sensitive model parameters can be considered while solving for optimal operating policy. Although the proposed approach cannot guarantee that the global solution is obtained since a local optimization algorithm is utilized, the convergence to the global optimal solution can be improved using initialization strategies in the interior point method. Mathematical models are very useful if they can generate prospective data which can lead to cost effective operations. Currently, in order to reach such a state in the model, in-depth understanding is needed in the following fields; science to understand the unit operation, engineering to develop design

equations for the unit operations, and mathematics and statistics to simulate the model and mine information from the model. In order to guide process development a novel approach is required in which data is integrated from multiple biological levels which includes molecular, cellular, and multicellular sources. Once the current barriers in molecular and cellular biology and genetics are cleared, vast amount of information is going to be readily available for use in characterizing bioprocesses. By integrating this new information, robust models can be created with increased predictability. Overall, the current focus should be on creating universal computational models that by providing specific parametric data about the process and the cells, prospective data can be generated which can create value to user. A well-posed model which represents the system well can be linked to proper optimization algorithms and recommend low-cost improvements for the operation of bioreactor. Furthermore, unit operation models with improved computational feasibility facilitate integration of control, scheduling, and planning as the leading step toward integrated decision-making.

Acknowledgement of previous publications

Several sections of this dissertation have been published or submitted for publication elsewhere. Following are acknowledged.

Chapter 2 has been published in full under the citation:

Farzan, P., B. Mistry, and M.G. Ierapetritou, *Review of the Important Challenges and Opportunities related to Modeling of Mammalian Cell Bioreactors.* AIChE Journal, 2017. **63**(2): p. 398–408.

Chapter 3 has been submitted to the Journal of Industrial & Engineering Chemistry Research in full under the title:

Farzan, P. and M.G. Ierapetritou, *A Framework for Development of Integrated and Computationally Feasible Models of Large-Scale Mammalian Cell Bioreactors*.

Appendix

Table A. 1: One-step sparse estimate of θ for radial velocity

Kriging Model #	θ_1	θ_2	θ_3	θ_4
1	3.5	3.4	1.4	1.7
2	9.9	74.4	4.4	93.4
3	105.5	1.8	1.3	1.6
4	3.0	10.6	1.3	2.9
5	1.1	2.3	0.2	3.5
6	12.5	1.0	0.9	9.7
7	4.0	48.6	9.8	5.6
8	15.1	7.0	0.7	3.2
9	0.7	2.8	0.0	0.5
10	0.3	1.1	0.1	0.5

Table A. 2: One-step sparse estimate of β for radial velocity

Kriging Model #	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_10	β_11
1	100.0	100.0	100.0	37.5	-50.9	15.5	100.0	-100.0	100.0	100.0	-100.0
2	100.0	100.0	100.0	100.0	-97.0	-36.9	5.1	70.5	100.0	100.0	-100.0
3	100.0	100.0	100.0	100.0	-37.2	69.6	-100.0	-100.0	53.5	-100.0	95.5
4	100.0	100.0	100.0	100.0	-4.4	45.0	61.4	-100.0	100.0	100.0	75.6
5	100.0	97.6	100.0	100.0	100.0	-18.2	100.0	-38.3	66.9	64.4	-100.0
6	100.0	100.0	100.0	100.0	100.0	100.0	-64.3	-46.9	36.3	-100.0	-100.0
7	100.0	100.0	100.0	100.0	-42.6	100.0	65.0	-100.0	100.0	63.3	100.0
8	100.0	100.0	100.0	100.0	5.9	4.9	100.0	-100.0	100.0	-100.0	100.0
9	100.0	100.0	100.0	-100.0	27.0	-51.7	-45.3	-100.0	100.0	100.0	-100.0
10	100.0	-19.6	100.0	-100.0	-5.5	-5.4	100.0	-100.0	66.7	-98.3	100.0

Table A. 3: One-step sparse estimate of θ for angular velocity

Kriging Model #	θ_1	θ_2	θ_3	θ_4
1	181.2	0.0	31.3	14.6
2	516.7	1000.0	0.0	0.0
3	30.1	0.0	0.2	0.0
4	0.0	25.3	2.8	2.6
5	0.0	0.0	41.6	9.8
6	0.0	0.0	27.3	9.5
7	8.6	0.0	0.0	0.0
8	0.0	0.0	21.5	11.8
9	4.5	0.0	0.0	0.0
10	0.0	0.0	25.0	14.5

Table A. 4: One-step sparse estimate of β for angular velocity

Kriging Model #	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_10	β_11
1	100.0	100.0	55.3	78.5	19.1	-51.6	-100.0	100.0	-100.0	49.5	57.8
2	100.0	-86.1	100.0	100.0	-100.0	-100.0	-20.8	29.7	100.0	11.0	93.1
3	100.0	100.0	-100.0	4.1	98.0	-10.1	-100.0	7.2	100.0	-100.0	-70.3
4	100.0	67.4	0.3	100.0	-28.5	75.9	-100.0	26.9	-92.5	-76.9	100.0
5	100.0	0.0	93.8	100.0	-15.1	-100.0	-74.6	100.0	-7.9	-95.7	43.0
6	100.0	57.0	98.8	80.5	2.7	-100.0	-100.0	-49.3	42.5	-60.2	-22.8
7	100.0	-31.7	100.0	100.0	-67.1	-31.5	-100.0	0.0	91.8	100.0	100.0
8	100.0	28.5	100.0	100.0	-50.4	-26.7	-100.0	63.5	-79.1	-13.0	100.0
9	94.3	100.0	-85.5	100.0	-100.0	-26.2	-29.8	-12.5	-100.0	59.7	39.2
10	100.0	43.2	100.0	72.4	-86.1	-74.2	-100.0	-20.2	62.4	100.0	100.0

Table A. 5: One-step sparse estimate of θ for axial velocity

Kriging Model #	θ_1	θ_2	θ_3	θ_4
1	0.5	4.6	3.1	1.1
2	23.0	100.0	2.9	0.8
3	0.0	4.5	6.7	3.0
4	0.8	10.4	3.7	2.9
5	0.9	22.1	1.3	4.2
6	7.9	2.4	4.3	1.3
7	20.8	27.4	8.1	0.3
8	10.7	17.1	0.2	1.5
9	89.1	12.3	0.4	0.5
10	0.2	0.0	0.0	0.0

Table A. 6: One-step sparse estimate of β for axial velocity

Kriging Model #	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_10	β_11
1	86.2	100.0	-100.0	-100.0	96.4	-48.3	73.2	18.6	-100.0	-5.1	-100.0
2	100.0	-100.0	51.9	100.0	-66.7	-27.3	-63.4	-96.2	-60.1	100.0	100.0
3	-45.4	100.0	-100.0	34.7	-100.0	100.0	-100.0	-26.1	100.0	100.0	33.3
4	-66.2	100.0	-100.0	7.6	57.6	-11.2	100.0	100.0	-100.0	-100.0	44.1
5	94.2	-39.5	-100.0	-53.4	-64.4	51.3	-100.0	-42.6	100.0	100.0	-78.9
6	-60.9	100.0	-76.6	-46.1	0.6	54.5	100.0	8.6	-100.0	-100.0	39.0
7	46.4	-18.0	-3.4	0.0	94.5	100.0	-32.4	-20.2	-100.0	-100.0	-91.3
8	29.8	-0.1	61.7	-100.0	-34.2	-17.5	-18.5	-6.2	76.1	-100.0	100.0
9	12.4	-14.6	-73.6	88.1	-86.1	100.0	-66.8	25.7	-100.0	34.6	100.0
10	-32.6	100.0	-89.2	100.0	100.0	-100.0	-25.6	77.2	-100.0	31.3	-100.0

Bibliography

1. Otto, R., A. Santagostino, and U. Schrader, *Rapid Growth in Biopharma Challenges and Opportunities*. 2014, McKinsey & Company.
2. Davidson, A. and S.S. Farid, *Innovation in Biopharmaceutical Manufacture*, in *BioProcess international*. 2014. p. 12-19.
3. Koutinas, M., et al., *Bioprocess Systems Engineering: Transferring Traditional Process Engineering Principles To Industrial Biotechnology.* Computational and Structural Biotechnology Journal, 2012. **3**(4): p. 1 - 9.
4. Rader, R.A. and E.S. Langer, *30 years of upstream productivity improvements.* BioProcess International, 2015. **13**(2): p. 10-15.
5. Shuler, M.L. and F. Kargi, *Bioprocess Engineering*. 2 ed. 2002: Prentice Hall.
6. Lara, A.R., et al., *Living with heterogeneities in bioreactors.* Molecular Biotechnology, 2006. **34**(3): p. 355-381.
7. Farid, S.S., *Process economics of industrial monoclonal antibody manufacture.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **848**(1): p. 8-18.
8. Terry, C. and N. Lesser, *Balancing the R&D equation*. 2017, Deloitte Center for Health Solutions.
9. Gyurjyan, G., et al., *Rethinking pharma productivity*. 2017, McKinsey & Company.
10. Varma, V.A., et al., *Enterprise-wide modeling & optimization - An overview of emerging research challenges and opportunities.* Computers & Chemical Engineering, 2007. **31**(5 - 6): p. 692 - 711.
11. Grossmann, I.E., *Advances in mathematical programming models for enterprise-wide optimization.* Computers & Chemical Engineering, 2012. **47**: p. 2 - 18.
12. Shimizu, K., *A tutorial review on bioprocess systems engineering.* Computers & Chemical Engineering, 1996. **20**(6 - 7): p. 915 - 941.
13. Bailey, J.E., *Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities.* Biotechnology Progress, 1998. **14**(1): p. 8-20.
14. Biegler, L.T., Y.D. Lang, and W.J. Lin, *Multi-scale optimization for process systems engineering.* Computers & Chemical Engineering, 2014. **60**: p. 17-30.

15. Puigjaner, L. and G. Heyen, *Computer Aided Process and Product Engineering (CAPE)*. 2006: Wiley.

16. Kontoravdi, C., E.N. Pistikopoulos, and A. Mantalaris, *Systematic development of predictive mathematical models for animal cell cultures.* Computers & Chemical Engineering, 2010. **34**(8): p. 1192-1198.

17. Kiparissides, A., et al., *'Closing the loop' in biological systems modeling - From the in silico to the in vitro.* Automatica, 2011. **47**(6): p. 1147 - 1155.

18. Sargent, R.G., *Verification and validation of simulation models.* Journal of Simulation, 2013. **7**(1): p. 12-24.

19. Nolan, R.P. and K. Lee, *Dynamic model of CHO cell metabolism.* Metabolic Engineering, 2011. **13**(1): p. 108-124.

20. Prokop, A., *Implications of Cell Biology in Animal Cell Biotechnology*, in *Animal Cell Bioreactors*, C.S. Ho and D.I.C. Wang, Editors. 1991, Butterworth-Heinemann.

21. Schmalzriedt, S., et al., *Integration of physiology and fluid dynamics*, in *Process Integration in Biochemical Engineering*. 2003. p. 19-68.

22. Kuystermans, D. and M. Al-Rubeai, *Bioreactor Systems for Producing Antibody from Mammalian Cells*, in *Cell Engineering, Vol 7: Antibody Expression and Production*, M. Al-Rubeai, Editor. 2011. p. 25-52.

23. Xie, L., W. Zhou, and D. Robinson, *Protein production by large-scale mammalian cell culture.* New Comprehensive Biochemistry, 2003. **38**: p. 605 - 623.

24. Spier, R.E., *Encyclopedia of Cell Technology*. Wiley Biotechnology Encyclopedias. 2000: Wiley-Interscience.

25. Panda, T., *Bioreactors: Analysis and Design*. 2011: Tata McGraw-Hill Education Private Limited.

26. Ho, C.S. and D.I.C. Wang, *Animal Cell Bioreactors*. Biotechnology Series, ed. J.E. Davies. 1991: Butterworth-Heinemann.

27. Mandenius, C.-F. and N.J. Titchener-Hooker, *Measurement, Monitoring, Modelling and Control of Bioprocesses*. Advances in Biochemical Engineering/Biotechnology. 2013: Springer.

28. Meyer, H.-P. and D. Schmidhalter, *Industrial Scale Suspension Culture of Living Cells*. 2014: Wiley.

29. Bailey, J. and D.F. Ollis, *Biochemical Engineering Fundamentals*. 2nd ed. 1986: McGraw-Hill.

30. Jang, J.D. and J.P. Barford, *An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody.* Biochemical Engineering Journal, 2000. **4**(2): p. 153 - 168.

31. Tatiraju, S., M. Soroush, and R. Mutharasan, *Multi-rate nonlinear state and parameter estimation in a bioreactor.* Biotechnology and Bioengineering, 1999. **63**(1): p. 22 - 32.

32. Munzer, D.G.G., et al., *An unstructured model of metabolic and temperature dependent cell cycle arrest in hybridoma batch and fed-batch cultures.* Biochemical Engineering Journal, 2015. **93**: p. 260-273.

33. Goudar, C.T., et al., *Logistic equations effectively model mammalian cell batch and fed-batch kinetics by logically constraining the fit.* Biotechnology Progress, 2005. **21**(4): p. 1109-1118.

34. Lopez-Meza, J., et al., *Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer CHO cells.* Cytotechnology, 2016. **68**(4): p. 1287-1300.

35. Craven, S., J. Whelan, and B. Glennon, *Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller.* Journal of Process Control, 2014. **24**(4): p. 344-357.

36. Sbarciog, M., D. Coutinho, and A. Vande Wouwer, *A simple output-feedback strategy for the control of perfused mammalian cell cultures.* Control Engineering Practice, 2014. **32**: p. 123-135.

37. Amribt, Z., H.X. Niu, and P. Bogaerts, *Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures.* Biochemical Engineering Journal, 2013. **70**: p. 196-209.

38. Xing, Z.Z., et al., *Modeling Kinetics of a Large-Scale Fed-Batch CHO Cell Culture by Markov Chain Monte Carlo Method.* Biotechnology Progress, 2010. **26**(1): p. 208-219.

39. Ahn, W.S. and M.R. Antoniewicz, *Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry.* Metabolic Engineering, 2011. **13**(5): p. 598-609.

40. Europa, A.F., et al., *Multiple steady states with distinct cellular metabolism in continuous culture of mammalian cells.* Biotechnology and Bioengineering, 2000. **67**(1): p. 25 - 34.

41. Carinhas, N., et al., *Systems biotechnology of animal cells: the road to prediction.* Trends in Biotechnology, 2012. **30**(7): p. 377 - 385.

42. Stephanopoulos, G.N., *Metabolic Flux Analysis*, in *Metabolic Engineering Principles and Methodologies*, G.N. Stephanopoulos, A.A. Aristidou, and J. Nielsen, Editors. 1998, Academic Press. p. 309 - 351.

43. Antoniewicz, M.R., *Dynamic metabolic flux analysis - tools for probing transient states of metabolic networks.* Current Opinion in Biotechnology, 2013. **24**(6): p. 973 - 978.

44. Long, C.P. and M.R. Antoniewicz, *Metabolic flux analysis of Escherichia coli knockouts: lessons from the Keio collection and future outlook.* Current Opinion in Biotechnology, 2014. **28**: p. 127 - 133.

45. Niu, H., et al., *Metabolic pathway analysis and reduction for mammalian cell cultures - Towards macroscopic modeling.* Chemical Engineering Science, 2013. **102**: p. 461 - 473.

46. Matsuoka, Y. and K. Shimizu, *Current status of ^{13}C-metabolic flux analysis and future perspectives.* Process Biochemistry, 2010. **45**(12): p. 1873 - 1881.

47. Sidoli, F.R., S.P. Asprey, and A. Mantalaris, *A Coupled Single Cell-Population-Balance Model for Mammalian Cell Cultures.* Industrial and Engineering Chemistry Research, 2006. **45**(16): p. 5801 - 5811.

48. Pigou, M. and J. Morchain, *Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models.* Chemical Engineering Science, 2015. **126**: p. 267-282.

49. Klamt, S. and J. Stelling, *Two approaches for metabolic pathway analysis?* Trends in Biotechnology, 2003. **21**(2): p. 64 - 69.

50. Orman, M.A., et al., *Pathway analysis of liver metabolism under stressed condition.* Journal of Theoretical Biology, 2011. **272**(1): p. 131-140.

51. Song, H.-S., F. DeVilbiss, and D. Ramkrishna, *Modeling metabolic systems: the need for dynamics.* Current Opinion in Chemical Engineering, 2013. **2**(4): p. 373 - 382.

52. Llaneras, F. and J. Pico, *A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient.* Bmc Bioinformatics, 2007. **8**.

53. Sanderson, C.S., J.P. Barford, and G.W. Barton, *A structured, dynamic model for animal cell culture systems.* Biochemical Engineering Journal, 1999. **3**(3): p. 203 - 211.

54. Smallbone, K., et al., *Towards a genome-scale kinetic model of cellular metabolism.* BMC Systems Biology, 2010. **4**: p. Article number 6.

55. Tran, L.M., M.L. Rizk, and J.C. Liao, *Ensemble Modeling of Metabolic Networks.* Biophysical Journal, 2008. **95**(12): p. 5606 - 5617.

56. Altamirano, C., et al., *Considerations on the lactate consumption by CHO cells in the presence of galactose.* Journal of Biotechnology, 2006. **125**(4): p. 547-556.

57. Leighty, R.W. and M.R. Antoniewicz, *Dynamic metabolic flux analysis (DMFA): A framework for determining fluxes at metabolic non-steady state.* Metabolic Engineering, 2011. **13**(6): p. 745-755.

58. Kompala, D.S., D. Ramkrishna, and G.T. Tsao, *Cybernetic modeling of microbial growth on multiple substrates.* Biotechnology and Bioengineering, 1984. **26**(11): p. 1272 - 1281.

59. Gambhir, A., et al., *Analysis of cellular metabolism of hybridoma cells at distinct physiological states.* Journal of Bioscience and Bioengineering, 2003. **95**(4): p. 317 - 327.

60. Kontoravdi, C., et al., *Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring.* Computers & Chemical Engineering, 2007. **31**(5 - 6): p. 392 - 400.

61. Bibila, T.A. and M.C. Flickinger, *Use of a structured kinetic model of antibody synthesis and secretion for optimization of antibody production systems: I. Steady-state analysis.* Biotechnology and Bioengineering, 1992. **39**(3): p. 251 - 261.

62. Xie, L. and D.I.C. Wang, *Material balance studies on animal cell metabolism using a stoichiometrically based reaction network.* Biotechnology and Bioengineering, 1996. **52**(5): p. 579 - 590.

63. Xie, L. and D.I.C. Wang, *Applications of improved stoichiometric model in medium design and fed-batch cultivation of animal cells in bioreactor.* Cytotechnology, 1994. **15**(1 - 3): p. 17 - 29.

64. Quek, L.-E., et al., *Metabolic flux analysis in mammalian cell culture.* Metabolic Engineering, 2010. **12**(2): p. 161 - 171.

65. Llaneras, F. and J. Pico, *An interval approach for dealing with flux distributions and elementary modes activity patterns.* Journal of Theoretical Biology, 2007. **246**(2): p. 290-308.

66. Liu, Y.H., et al., *A cybernetic model to describe the dynamics of myeloma cell cultivations.* Applied Mathematics and Computation, 2008. **205**(1): p. 84-97.

67. Geng, J., et al., *Application of hybrid cybernetic model in simulating myeloma cell culture co-consuming glucose and glutamine with mixed consumption patterns.* Process Biochemistry, 2013. **48**(5-6): p. 955-964.

68. Dorka, P., et al., *Metabolic flux-based modeling of mAb production during batch and fed-batch operations.* Bioprocess and Biosystems Engineering, 2009. **32**(2): p. 183-196.

69. Shuler, M.L., *Single-cell models: promise and limitations.* Journal of Biotechnology, 1999. **71**(1-3): p. 225-228.

70. Ramkrishna, D., *Population Balances Theory and Applications to Particulate Systems in Engineering*. 2000: Academic Press.

71. Hjortso, M.A. and J. Nielsen, *Population balance models of autonomous microbial oscillations.* Journal of Biotechnology, 1995. **42**(3): p. 255 - 269.

72. Mantzaris, N.V., et al., *Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration.* Journal of Biotechnology, 1999. **71**(1-3): p. 157-174.

73. Henson, M.A., D. Muller, and M. Reuss, *Cell population modelling of yeast glycolytic oscillations.* Biochemical Journal, 2002. **368**: p. 433 - 446.

74. Nishimura, Y. and J.E. Bailey, *Bacterial population dynamics in batch and continuous-flow microbial reactors.* AIChE Journal, 1981. **27**(1): p. 73 - 81.

75. Fredrickson, A.G., *Formulation of structured growth models.* Biotechnology and Bioengineering, 1976. **18**(10): p. 1481-1486.

76. Mantzaris, N.V. and P. Daoutidis, *Cell population balance modeling and control in continuous bioreactors.* Journal of Process Control, 2004. **14**(7): p. 775 - 784.

77. Ramkrishna, D. and M.R. Singh, *Population balance modeling: Current status and future prospects.* Annual Review of Chemical and Biomolecular Engineering, 2014. **5**: p. 123 - 146.

78. Ramkrishna, D., *Statistical Models of Cell Populations* Advances in Biochemical Engineering 1979. **11**: p. 1 - 47.

79. Stamatakis, M., *Cell population balance and hybrid modeling of population dynamics for a single gene with feedback.* Computers & Chemical Engineering, 2013. **53**: p. 25-34.

80. Nielsen, J. and J. Villadsen, *Modelling of microbial kinetics.* Chemical Engineering Science, 1992. **47**(17 - 18): p. 4225 - 4270.

81. Ramkrishna, D., *Toward a self-similar theory of microbial populations.* Biotechnology and Bioengineering, 1994. **43**(2): p. 138 - 148.

82. Ramkrishna, D. and J. Schell, *On self-similar growth.* Journal of Biotechnology, 1999. **71**(1-3): p. 255-258.

83. Spetsieris, K. and K. Zygourakis, *Single-cell behavior and population heterogeneity: Solving an inverse problem to compute the intrinsic physiological state functions.* Journal of Biotechnology, 2012. **158**(3): p. 80-90.

84. Stamatakis, M., *Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches.* Chemical Engineering Science, 2010. **65**(2): p. 1008-1015.

85. Shuler, M.L. and M.M. Domach, *Mathematical models of the growth of individual cells*, in *Foundations of Biochemical Engineering*, H.W. Blanch, E.T. Papoutsakis, and G. Stephanopolous, Editors. 1983, American Chemical Society: Washington, D.C. p. 93 - 133.

86. Ataai, M.M. and M.L. Shuler, *Simulation of CFSTR through development of a mathematical model for anaerobic growth of Escherichia coli cell population.* Biotechnology and Bioengineering, 1985. **27**(7): p. 1051 - 1055.

87. Domach, M.M. and M.L. Shuler, *A finite representation model for an asynchronous culture of E. coli.* Biotechnology and Bioengineering, 1984. **26**(8): p. 877 - 884.

88. Kim, B.G. and M.L. Shuler, *A structured, segregated model for genetically modified Escherichia coli cells and its use for prediction of plasmid stability.* Biotechnology and Bioengineering, 1990. **36**(6): p. 581 - 592.

89. Fernandes, R.L., et al., *Experimental methods and modeling techniques for description of cell population heterogeneity.* Biotechnology Advances, 2011. **29**(6): p. 575-599.

90. Meshram, M., et al., *Modeling the coupled extracellular and intracellular environments in mammalian cell culture.* Metabolic Engineering, 2013. **19**: p. 57 - 68.

91. Mantzaris, N.V., *Stochastic and deterministic simulations of heterogeneous cell population dynamics.* Journal of Theoretical Biology, 2006. **241**(3): p. 690-706.

92. Fadda, S., A. Cincotti, and G. Cao, *A novel population balance model to investigate the kinetics of in vitro cell proliferation: Part I. model development.* Biotechnology and Bioengineering, 2012. **109**(3): p. 772-781.

93. Jandt, U., et al., *Synchronized Mammalian Cell Culture: Part II-Population Ensemble Modeling and Analysis for Development of Reproducible Processes.* Biotechnology Progress, 2015. **31**(1): p. 175-185.

94. Ishii, M. and T. Hibiki, *Thermo-Fluid Dynamics of Two-Phase Flow*. 2011: Springer.

95. Davies, J.T., *Turbulence Phenomena*. 1972: Academic Press, Inc.

96. Moo-Young, M. and H.W. Blanch, *Design Of Biochemical Reactors Mass Transfer Criteria For Simple And Complex Systems.* Advances in Biochemical Engineering, 1981. **19**: p. 1-69.

97. Kawase, Y. and M. Mooyoung, *Mathematical-Models for Design Of Bioreactors - Applications Of Kolmogorov Theory of Isotropic Turbulence.* Chemical Engineering Journal and the Biochemical Engineering Journal, 1990. **43**(1): p. B19-B41.

98. Kerdouss, F., et al., *Two-phase mass transfer coefficient prediction in stirred vessel with a CFD model.* Computers & Chemical Engineering, 2008. **32**(8): p. 1943-1955.

99. Yoshimoto, M., et al., *Gas-liquid interfacial area, bubble size and liquid-phase mass transfer coefficient in a three-phase external loop airlift bubble column.* Chemical and Biochemical Engineering Quarterly, 2007. **21**(4): p. 365-372.

100. Riet, K.v.t., *Mass transfer in fermentation.* Trends in biotechnology, 1983. **1**(4): p. 113 - 119.

101. Wang, K.B. and L.T. Fan, *Mass transfer in bubble columns packed with motionless mixers.* Chemical Engineering Science, 1978. **33**(7): p. 945 - 952.

102. Patel, N. and J. Thibault, *Data Reconciliation Using Neural Networks for the Determination of K(L)a*, in *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, M.D.C. Nicoletti and L.C. Jain, Editors. 2009. p. 197-214.

103. Linek, V., M. Kordac, and T. Moucha, *Mechanism of mass transfer from bubbles in dispersions - Part II: Mass transfer coefficients in stirred gas-liquid reactor and bubble column.* Chemical Engineering and Processing, 2005. **44**(1): p. 121-130.

104. Linek, V., T. Moucha, and M. Kordac, *Mechanism of mass transfer from bubbles in dispersions Part I. Danckwerts' plot method with sulphite solutions in the presence of viscosity and surface tension changing agents.* Chemical Engineering and Processing, 2005. **44**(3): p. 353-361.

105. Cachaza, E.M., et al., *Simultaneous Computational Fluid Dynamics (CFD) simulation of the hydrodynamics and mass transfer in a partially aerated bubble column.* Industrial and Engineering Chemistry Research, 2009. **48**(18): p. 8685-8696.

106. Azbel, D., *Two Phase Flows in Chemical Engineering*. 2009: Cambridge University Press.

107. Kadic, E. and T.J. Heindel, *An Introduction to Bioreactor Hydrodynamics and Gas-Liquid Mass Transfer*. 1st ed. 2014: Wiley.

108. Kumar, S. and L.S. Fan, *Heat-Transfer Characteristics in Viscous Gas-Liquid and Gas-Liquid-Solid Systems.* Aiche Journal, 1994. **40**(5): p. 745-755.

109. Higbie, R., *The rate of absorption of a pure gas into a still liquid during short periods of exposure.* Transactions of the American Institute of Chemical Engineers, 1935. **31**: p. 364–389.

110. Kawase, Y., B. Halard, and M. Moo-Young, *Theoretical prediction of volumetric mass transfer coefficients in bubble columns for Newtonian and non-Newtonian fluids.* Chemical Engineering Science, 1987. **42**(7): p. 1609-1617.

111. Kawase, Y., B. Halard, and M. Mooyoung, *Liquid-Phase Mass-Transfer Coefficients in Bioreactors.* Biotechnology and Bioengineering, 1992. **39**(11): p. 1133-1140.

112. Gimbun, J., C.D. Rielly, and Z.K. Nagy, *Modelling of mass transfer in gas–liquid stirred tanks agitated by Rushton turbine and CD-6 impeller: A scale-up study.* Chemical Engineering Research and Design, 2009. **87**(4): p. 437 - 451.

113. Danckwerts, P.V., *Significance of Liquid-Film Coefficients in Gas Absorption.* Industrial & Engineering Chemistry, 1951. **43**(6): p. 1460–1467.

114. Lamont, J.C. and D.S. Scott, *An eddy cell model of mass transfer into the surface of a turbulent liquid.* AIChE Journal, 1970. **16**(4): p. 513–519.

115. Bruning, S. and D. Weuster-Botz, *CFD analysis of interphase mass transfer and energy dissipation in a milliliter-scale stirred-tank reactor for filamentous microorganisms.* Chemical Engineering Research & Design, 2014. **92**(2): p. 240-248.

116. Calderbank, P.H. and M.B. Moo-Young, *The continuous phase heat and mass-transfer properties of dispersions.* Chemical Engineering Science, 1961. **16**(1-2): p. 39-54.

117. Linek, V., et al., *Gas-liquid mass transfer coefficient in stirred tanks interpreted through models of idealized eddy structure of turbulence in the bubble vicinity.* Chemical Engineering and Processing, 2004. **43**(12): p. 1511-1517.

118. Godbole, S.P., et al., *Hydrodynamics and Mass-Transfer in Non-Newtonian Solutions in a Bubble Column.* Aiche Journal, 1984. **30**(2): p. 213-220.

119. Akita, K. and F. Yoshida, *Bubble size, interfacial area, and liquid-phase mass transfer coefficient in bubble columns.* Ind. Eng. Chem., Process Des. Develop., 1974. **13**(1): p. 84 - 91.

120. Pangarkar, V.G., *Design of Multiphase Reactors*. 2015: Wiley.

121. Bauer, M. and G. Eigenberger, *A concept for multi-scale modeling of bubble columns and loop reactors.* Chemical Engineering Science, 1999. **54**(21): p. 5109-5117.

122. Chen, P., J. Sanyal, and M.P. Dudukovic, *Numerical simulation of bubble columns flows: effect of different breakup and coalescence closures.* Chemical Engineering Science, 2005. **60**(4): p. 1085-1101.

123. Chalmers, J., *Animal cell culture, effects of agitation and aeration on cell adaption*, in *Encyclopedia of Cell Technology*, R.E. Spier, Editor. 2000, Wiley-Interscience.

124. Alves, P., et al., *Influence of the Culture System upon Growth and Productivity of Animal Cells in Stirred Tanks*, in *Animal Cell Technology: Developments towards the 21st Century*, E.C. Beuvery, W.P. Zeijlemaker, and J.B. Griffiths, Editors. 1995, Springer. p. 105-109.

125. Ivarsson, M., et al., *Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation.* Journal of Biotechnology, 2014. **188**: p. 88-96.

126. Nienow, A.W., et al., *Scale-down studies for assessing the impact of different stress parameters on growth and product quality during animal cell culture.* Chemical Engineering Research & Design, 2013. **91**(11): p. 2265-2274.

127. Bibila, T.A. and D.K. Robinson, *In pursuit of the optimal fed-batch process for monoclonal antibody production.* Biotechnology progress, 1995. **11**(1): p. 1 -13.

128. Rocha, I., *Model-based strategies for computer-aided operation of recombinant E. coli fermentation*. 2003, Universidade do Minho.

129. Nicoletti, M.C., L.C. Jain, and R.C. Giordano, *Computational Intelligence Techniques as Tools for Bioprocess Modelling, Optimization, Supervision and Control*, in *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, M.D.C. Nicoletti and L.C. Jain, Editors. 2009. p. 1-23.

130. Wang, H.N., et al., *CFD modeling of hydrodynamic characteristics of a gas-liquid two-phase stirred tank.* Applied Mathematical Modelling, 2014. **38**(1): p. 63-92.

131. Azargoshasb, H., et al., *Three-phase CFD simulation coupled with population balance equations of anaerobic syntrophic acidogenesis and methanogenesis reactions in a continouos stirred bioreactor.* Journal of Industrial and Engineering Chemistry, 2015. **27**: p. 207-217.

132. Micale, G., et al., *CFD simulation of particle distribution in stirred vessels.* Chemical Engineering Research & Design, 2000. **78**(A3): p. 435-444.

133. Farzan, P. and M.G. Ierapetritou, *Integrated Modeling to Capture the Interaction of Physiology and Fluid Dynamics in Biopharmaceutical Bioreactors.* Computers & Chemical Engineering, 2017. **97**: p. 271–282.

134. Bezzo, F., S. Macchietto, and C.C. Pantelides, *A general methodology for hybrid multizonal/CFD models: Part I. Theoretical framework.* Computers & Chemical Engineering, 2004. **28**(4): p. 501 - 511.

135. Vrabel, P., et al., *Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements.* Chemical Engineering Science, 2000. **55**(23): p. 5881-5896.

136. Kagoshima, M. and R. Mann, *Development of a networks-of-zones fluid mixing model for an unbaffled stirred vessel used for precipitation.* Chemical Engineering Science, 2006. **61**(9): p. 2852-2863.

137. Delafosse, A., et al., *CFD-based compartment model for description of mixing in bioreactors.* Chemical Engineering Science, 2014. **106**: p. 76-85.

138. Vrabel, P., et al., *Compartment model approach: Mixing in large scale aerated reactors with multiple impellers.* Chemical Engineering Research & Design, 1999. **77**(A4): p. 291-302.

139. Bashiri, H., et al., *Compartmental modelling of turbulent fluid flow for the scale-up of stirred tanks.* Canadian Journal of Chemical Engineering, 2014. **92**(6): p. 1070-1081.

140. *ANSYS Fluent Theory Guide*. Release 15.0 ed. 2013: ANSYS Inc.

141. *ANSYS Fluent User's Guide*. Release 15.0 ed. 2013: ANSYS, Inc.

142. Adams, R.L.P., *Cell Culture for Biochemists*. 1990: Elsevier.

143. Kaiser, S.C., et al., *CFD for Characterizing Standard and Single-use Stirred Cell Culture Bioreactors*, in *Computational Fluid Dynamics Technologies and Applications*, I.V. Minin and O.V. Minin, Editors. 2011, InTech. p. 97-122.

144. Sarkar, J., et al., *CFD of mixing of multi-phase flow in a bioreactor using population balance model* Biotechnology Progress, 2016. **32**(3): p. 613–628.

145. Alves, S.S., et al., *Bubble size in aerated stirred tanks.* Chemical Engineering Journal, 2002. **89**(1-3): p. 109-117.

146. Chatterjee, A., *An introduction to the proper orthogonal decomposition.* Current Science, 2000. **78**(7): p. 808-817.

147. Chen, H., D.L. Reuss, and V. Sick, *On the use and interpretation of proper orthogonal decomposition of in-cylinder engine flows.* Measurement Science and Technology, 2012. **23**(8).

148. Xiu, Z.-L., W.-D. Deckwer, and A.-P. Zeng, *Estimation of rates of oxygen uptake and carbon dioxide evolution of animal cell culture using material and energy balances.* Cytotechnology, 1999. **29**(3): p. 159-166.

149. Mostafa, S.S. and X.J. Gu, *Strategies for improved dCO(2) removal in large-scale fed-batch cultures.* Biotechnology Progress, 2003. **19**(1): p. 45-51.

150. Kolev, N.I., *Solubility of O2, N2, H2 and CO2 in water*, in *Multiphase Flow Dynamics 4 Turbulence, Gas Adsorption and Release, Diesel Fuel Properties*, N.I. Kolev, Editor. 2012. p. 209 - 239.

151. Ozturk, S.S. and B.O. Palsson, *Growth, metabolic, and antibody-production kinetics of hybridoma cell-culture .2. Effects of serum concentration, dissolved-oxygen concentration, and medium ph in a batch reactor.* Biotechnology Progress, 1991. **7**(6): p. 481-494.

152. Bryson, J.A.E. and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. 1975: CRC Press.

153. Flores-Tlacuahuac, A., S.T. Moreno, and L.T. Biegler, *Global optimization of highly nonlinear dynamic systems.* Industrial & Engineering Chemistry Research, 2008. **47**(8): p. 2643-2655.

154. Mahadevan, R. and F.J. Doyle, *On-line optimization of recombinant product in a fed-batch bioreactor.* Biotechnology Progress, 2003. **19**(2): p. 639-646.

155. Banga, J.R., et al., *Dynamic Optimization of Bioreactors: A Review.* Proceedings of the Indian National Science Academy, 2003. **69**(3 & 4): p. 257 - 265.

156. Cuthrell, J.E. and L.T. Biegler, *Simultaneous-Optimization and Solution Methods for Batch Reactor Control Profiles.* Computers & Chemical Engineering, 1989. **13**(1-2): p. 49-62.

157. Hedengren, J.D., et al., *Nonlinear modeling, estimation and predictive control in APMonitor.* Computers & Chemical Engineering, 2014. **70**: p. 133-148.

158. Constantinides, A. and N. Mostoufi, *Numerical Methods for Chemical Engineers with MATLAB Applications*. 2000: Prentice Hall.

159. Byrd, R.H., J.C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming.* Mathematical Programming, 2000. **89**(1): p. 149-185.

160. Ford, J.J., et al., *X-ray computed tomography of a gas-sparged stirred-tank reactor.* Chemical Engineering Science, 2008. **63**(8): p. 2075-2085.

161. Zhao, Y., Y. Amemiya, and Y. Hung, *Efficient Gaussian Process Modeling using Experimental Design-Based Subagging*, in *Conference on Experimental Design and Analysis (CEDA)*. 2016, Institute of Statistical Science, Academia Sinica: Academia Sinica, Taiwan.

162. Chu, T.J., J. Zhu, and H.N. Wang, *Penalized Maximum Likelihood Estimation and Variable Selection in Geostatistics.* Annals of Statistics, 2011. **39**(5): p. 2607-2625.

163. Zou, H. and R.Z. Li, *One-step sparse estimates in nonconcave penalized likelihood models.* Annals of Statistics, 2008. **36**(4): p. 1509-1533.

164. Fernandes, R.L., et al., *Applying Mechanistic Models in Bioprocess Development*, in *Measurement, Monitoring, Modelling and Control of Bioprocesses*. 2013. p. 137-166.

165. Wang, G., et al., *Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses.* Engineering in Life Sciences, 2015. **15**(1): p. 20-29.

166. Jiménez-Gonzáleza, C. and J.M. Woodley, *Bioprocesses: Modeling needs for process evaluation and sustainability assessment.* Computers & Chemical Engineering, 2010. **34**(7): p. 1009 - 1017.

167. Kong, L.-n., et al., *On the measurement of gas holdup distribution near the region of impeller in a gas–liquid stirred Rushton tank by means of γ-CT.* Chemical Engineering Journal, 2012. **188**: p. 191-198.

168. Li, F., et al., *Cell culture processes for monoclonal antibody production.* Mabs, 2010. **2**(5): p. 466-479.

169. Mitchell, M.J., et al., *A model of carbon dioxide dissolution and mineral carbonation kinetics.* Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences, 2010. **466**(2117): p. 1265-1290.