# Advanced process control for continuous bioprocessing of biotherapeutic protein production

*A Thesis*

*submitted in partial fulfilment of the requirements*

*for the award of the degree of*

**Doctor of Philosophy**

*Submitted by*

**Nivedhitha S**

**(Roll No. 166152001)**

**Centre for the Environment**

**Indian Institute of Technology Guwahati**

**Guwahati -781039, Assam, India**

**MARCH 2022**

# … Dedicated to my amma and appa

**தலைசிறந்த வெற்றியை அடைவதை விட அதன்பொருட்டு எடுக்கப்படும் சிறு முயற்சிகளே மேன்மையானவை, ஏனென்றால் இலக்கை அடைவதை காட்டிலும் அதை நோக்கி நகரும் பயணமே முக்கியமானது!**

**Infinitesimal consistent progress is much better than achieving remarkable success, for all that matters is the journey and not the destination!**

**Centre for the Environment**

**Indian Institute of Technology**

**Guwahati**

**Guwahati -781039, Assam, India**

# DECLARATION

I, hereby declare that the research findings in this thesis entitled **"Advanced process control for continuous bioprocessing of biotherapeutic protein production"** are the result of original research work carried out by me under the supervision of Prof. Senthilmurugan Subbiah (supervisor), Department of Chemical Engineering and Prof. Senthilkumar Sivaprakasam (co-supervisor), Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy. The results reported herein has not been submitted elsewhere for any degree or membership of any institute or university to the best of my knowledge and belief. Also, due acknowledgements have been made wherever the research findings of other researchers have been cited in this thesis.

Date: Nivedhitha S

Place: 166152001

**Centre for the Environment**

**Indian Institute of Technology**

**Guwahati**

**Guwahati -781039, Assam, India**

# CERTIFICATE

This is to certify that the thesis entitled **"Advanced process control for continuous bioprocessing of biotherapeutic protein production"** submitted by **Nivedhitha S** (Registration No. 166152001) to the Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy is a record of bonafide research work carried out by her under our supervision and guidance. The thesis work, in our opinion, has reached the requisite standard fulfilling the requirement for the degree of Doctor of Philosophy. The results contained in this thesis have not been submitted to other university or institute for award of any degree or diploma to the best of our knowledge and belief.

Date:
Place:

| | |
|---|---|
| **(Signature of Thesis Supervisor)** | **(Signature of Thesis Supervisor)** |
| Prof. Senthilmurugan Subbiah, | Prof. Senthilkumar Sivaprakasam, |
| Professor, Department of Chemical Engineering, | Professor, Department of Biosciences and Bioengineering, |
| Indian Institute of Technology Guwahati, | Indian Institute of Technology Guwahati, |
| Guwahati - 781039, Assam, India. | Guwahati - 781039, Assam, India. |

# ACKNOWLEDGMENTS

A research journey is in itself one of the toughest paths to travel, and completing this remarkable journey won't be possible without a solid support system to rely on. I am deeply indebted to several people who travelled with me throughout this journey. I would like to take this opportunity to provide a brief acknowledgement to everyone who enabled me to see the light of this day.

First and foremost, I would like to thank my supervisor **Prof. Senthilmurugan Subbiah** for providing a wonderful opportunity to kick start my research career right after my undergraduate studies. He has been instrumental in designing my research objectives and enabling the opportunity to work in collaborative projects to mold my research career. Next, I would like to thank my co-supervisor, **Prof. Senthilkumar Sivaprakasam**, for enabling me an opportunity to work collaboratively with his research group and providing his valuable insights and suggestions towards the betterment of my thesis. I am grateful to my supervisors for their consistent support, encouragement and motivation throughout my tenure in IITG. Even during the times when I felt demotivated and clueless, they were always there to guide me and direct me towards achieving my research objectives. Their valuable insights on moral values and personality development are important lessons that I would cherish and take forward throughout my life.

I am also greatly indebted to my doctoral committee members **Prof. Kannan Pakshirajan, Dr. Soumen Kumar Maiti and Dr. Deepak Sharma** for their kind words, constant encouragement and constructive inputs during my seminars which helped me shape my thesis to its present form.

I would like to express my sincere gratitude to my external mentor from IIT Delhi **Prof. Anurag S. Rathore,** for providing an excellent opportunity to collaborate with his research team and thereby availing me a platform to perform my experimental studies. His valuable insights spearheaded my research and directed the same towards the desired direction. I also extend my gratitude to my collaborator **Dr. Priyanka** from IIT Delhi, for her meticulous mentorship and significant contributions to my research articles. I would also like to extend my thanks to the research facilities from IITD for conducting my experiments.

I owe my gratitude to the technical team from ABER instruments and Mr. Sundar (Labmate Asia) for their persistent support and assistance in setting up and troubleshooting of Futura software. I am grateful to Prof. Latha Rangan (Dept. of BSBE, IITG) for granting permission to perform flow cytometry studies in her research lab and I thank Mr. Manish Kumar Gupta for assisting me with the same. My sincere gratitude goes to Mr. Seshadri Srinivasan for helping me with troubleshooting of model development in MATLAB. I would also like to thank the Central Instruments Facility, Department of Biosciences and Bioengineering, Centre for the Environment and Department of Chemical Engineering for their state-of-the-art analytical facilities. I am also grateful to all the technical staff from all the above-mentioned departments.

I got the excellent opportunity to be a part of not just one but three research groups, and I got to interact with several people throughout the course of these five years. I would like to thank all my seniors and lab mates from each research group. From Prof. Senthilmurugan' s research group (Water and Energy nexus lab, CL), I would like to thank- Dr. Vishal Verma, Dr. Vigneshwaran, Dr. Arunkumar C, Dr. Habtom Teklu, Ms. Aanisha aktar, Mr. Senthil S, Mr. Surendhar S, Mr. Viswanth Ramba, Mr. Kranthi Munubarathi (late), Mr. Muniraja Tippa, Mr. Naveenkumar AY, Ms. Neelam Dutta, Mr. Priyamjeet Deka, Ms. Ananya Bardhan, Mr. Dinesh Kumar G, Mr. Balakumara Vignesh M, and Ms. Seema Bharati for their unconditional love, care and support.

I would like to thank the research scholars from Prof. Senthilkumar' s research group (BioPAT lab, BSBE) – Dr. Srikanth Katla, Dr. Naresh Mohan, Dr. Rengesh B, Dr. Kiran Kumar Gali, Dr. Ganesh Nehru, Mr. Subbi Rami Reddy, Mr. Satyasai Pavan, Mr. Sidharth Guhan, Ms. Payal Mukherjee, and Ms. Sandhya S. I thank all the senior scholars for their mentorship and encouragement throughout my tenure and for providing an opportunity to work in collaborative projects. Further, I would like to thank the scholars from Prof. Rathore's research group (IITD) – Dr. Nikhil Kateja, Dr. Viswanath Hebbi, Dr. Pradnya Meshram, Ms. Anamika Tiwari, Ms. Rucha Patil, Mr. Souhardya Roy and many others from the entire team for their guidance, motivation and support throughout my time in IIT Delhi.

I am glad to have made some great friends on campus and would like to thank them for their immense care and encouragement. To mention a few- Jyoti, Sanjukta, Divya, Chandrima, Manasasri, Saswati, Indumathi, Sahaya glingston, Manoj, Balakumaran,

Sundar, Krishnakumar, Muthuvel, Anilkumar, Kamalesh, Arun and many other people who I might have missed out. I would also like to give a special shout out to my fellow Xpression club mates for bringing a great diversion from the otherwise stressful research schedule. I am also glad to thank the student counselling cell, IITG, and especially Ms. Pallabita Chowdary for providing valuable support when I needed it the most.

I would also like to mention my greatest pillars of strength - Chomundeswari, Suhirja, Aishwarya and Archana, without whom I would not have come this far. Though they were miles away and stayed in different time zones, they never failed to listen to my ramblings and always stood by my side. I take this opportunity to let them know how much they mean to me. I would also thank all my friends and well-wishers who supported me throughout this journey. I am also greatly indebted to the musicians from all over the world who kept me going every day.

My final gratitude is for my family, without whom none of this would have been possible. I am immensely thankful for my parents Mr. Swaminathan K K and Ms. Shanthy Swaminathan, and my brother Arvindh for their immense love, patience and support throughout the ups and downs of this journey.

**Nivedhitha S**

# ABSTRACT

The global biotherapeutics market has been tremendously growing with a compound annual growth rate (CAGR) of 13.8%, and it is expected to reach a total sales value close to half a trillion USD within the next five years. The recent global pandemic highlighted the significance of the health sector and the inevitable necessity of therapeutic products; and the biopharmaceutical sector is expected to get greater attention in the upcoming years. The rising demand for the production of high-quality therapeutic products in affordable price range, especially in developing countries, has made the manufacturers to focus on the implementation of a continuous manufacturing process for the production of the same. Successful and efficient production of therapeutic products demands a meticulous understanding and inspection of the process. The Quality by design (QbD) principles defined by the United States Food and Drug Administration (FDA) intend to ensure the purity and efficacy of the final pharmaceutical product. The critical quality attributes (CQA) of the therapeutic products are determined by the nature of the organism, media compositions, reactor operating conditions, and the control system in place. Efficient measurement and control of the critical process parameters (CPP) governing the CQAs can be achieved by implementing the mechanisms of Process analytical technology (PAT) as imparted by the FDA.

The thesis presented herein attempts to implement the different components of PAT to venture the possibility of acquiring reliable real-time process measurements and utilizing them in a process model to develop an efficient control strategy. The expression of the therapeutic product Ranibizumab (also called Lucentis) in recombinant *Escherichia coli* has been chosen as the system for exploring the various concepts of PAT. Ranibizumab is a humanized monoclonal antibody fragment used for the treatment of age-related macular degeneration, and the strategies for implementing the goals of PAT for enhancing the expression of this therapeutic product in a fed-batch recombinant *E. coli* cultivation was investigated in this work.

The first phase of this study focused on acquiring reliable process measurements of the biomass concentration, which is a significant process parameter concerning the final product quality. The physiological changes associated with cellular population would

impact product quality, and therefore, real-time monitoring and estimation of the biomass concentration using an emerging PAT tool, namely, dielectric spectroscopy (DS), was explored. The real-time scanning capacitance data from DS were pre-processed using moving average (MA). It was then modelled through a nonlinear theoretical Cole-Cole model, and the model parameters were estimated using a global optimization technique. The parameters obtained from the GA were further applied to estimate the physiological properties like cell diameter and viable cell concentration (VCC). Further, physiological properties, namely, cell diameter and VCC obtained from the Cole-Cole model, were validated using traditional offline analytical methods like particle size analyzer and flow cytometry, respectively. The Cole-Cole model predicted the cell diameter and viable cell concentration with an error of 1.03% and 7.72%, respectively. A hybrid approach using the combination of two algorithms was proposed for real-time estimation of the physiological properties. The proposed approach can enable the operator to take real-time process decisions to achieve desired productivity and product quality.

The second phase of the study focused on developing a mechanistic model based on mass balances of the various state variables of fermentation and the application of the model to optimize the total biomass with the aid of online capacitance measurements. Model development is a crucial step, as it attempts to relate the measured CPPs to the desired CQAs. Mechanistic models offer reliable predictions and, as such, are more suitable for developing effective control strategies and overall process optimization. The developed mechanistic model was validated using experimental data sets obtained from the production of a therapeutic product, Ranibizumab, from *E. coli.* The model predicted the experimental results of the calibration set and validation set within an average error value of 12.64% and 14.97%, respectively. Developing a mechanistic model for a biotherapeutic production process would be beneficial for implementing optimization and control studies.

The final phase of the study focuses on the development of different optimization case studies for achieving enhanced productivity. The optimization of total biomass is beneficial in the case of intracellular therapeutic protein production, which is relevant in the current system of investigation. Apart from the benefits of enhanced productivity, achieving maximum biomass with a minimum reactor volume can minimize the downstream production costs. Therefore, combining the two conflicting objectives and

achieving higher volumetric productivity would be relevant in view of the growing interest in continuous processing in the biopharmaceutical industry.

The objective of the case study (1) focused on maximizing the total biomass in the reactor at a minimum broth volume. A validated mechanistic model was employed to formulate a multiobjective optimization (MOO) problem. The substrate flow rate during the fed-batch phase ($F$) was taken as the decision variable for the MOO. The Pareto front resulting from MOO revealed that for a minimum broth volume ($V$) of 1.96 L, a maximum of 58.8 g of total biomass ($XV$) could be generated. The total biomass obtained from the optimal substrate feeding profile was 20.6% higher than the experimentally achieved total biomass. Enhanced productivity was achieved by the proposed MOO formulation, which facilitates the choice of any operating point from the Pareto front based on downstream expenses of the therapeutic product. The case study (2) focuses on the development of optimization strategies for predicting an optimal fed-batch harvest time. The harvest of a batch is typically linked to the time of induction. Rather than using time as the control criterion, basing harvest on biomass concentration is likely to result in more consistent process performance. The previously developed MOO was used along with a third objective of optimizing final harvest time $t_f(t_{end})$. Simulation studies were carried out with different $t_{end}$ values to predict the optimal fed-batch harvest time. The Pareto for different $t_{end}$ values were obtained, and the objective functions were compared at different $\lambda$ values. The optimal feeding profiles and fed-batch harvest time can be chosen based on the desired volume of operation.

In a nutshell, the approach presented in the thesis integrated the real-time process measurements in a validated process model and explored the application of different optimization strategies for a therapeutic protein production process. The combination of enhanced measurement, modelling and control strategies will significantly improve the product quantity and quality, thereby paving the way for better process performance.

# NOMENCLATURE

| | |
|---|---|
| *area* | Area of the fermenter ($m^2$) |
| $air_{in}$ | Air flow rate ($L\ h^{-1}$) |
| *A, B, C* | Correlation constants for volumetric mass transfer coefficient (-) |
| $C_{biomass}$ | Raw material cost for biomass production ($ per g of biomass) |
| $C_{Downstreamprocessing}$ | Downstream processing cost for therapeutic product production ($ per L of broth volume. |
| $C_{product}$ | Cost of the therapeutic protein product ($ per g of product) |
| $C_L$ | Dissolved oxygen concentration ($g\ L^{-1}$) |
| $C_L$* | Dissolved oxygen concentration at equilibrium with gas phase ($g\ L^{-1}$) |
| $C_m$ | Membrane capacitance ($F\ m^{-2}$) |
| $D_i$ | Impeller diameter (m) |
| *DO* | Dissolved oxygen concentration (%) |
| $f_c$ | Characteristic or critical frequency for capacitance ($s^{-1}$) |
| $f$ | Objective function for multiobjective optimization |
| $f_{profit}$ | Profit function |
| $F$ | Substrate flow rate in fed-batch phase ($L\ h^{-1}$) |
| $H$ | Henry's constant ($L\ kPa\ g^{-1}$) |
| $k_La$ | Volumetric mass transfer coefficient ($h^{-1}$) |
| $k_{O_2}$ | Monod saturation constant for oxygen ($g\ L^{-1}$) |
| $k_{S_b}$ | Monod saturation constant for substrate in batch phase ($g\ L^{-1}$) |
| $k_{S_f}$ | Monod saturation constant for substrate in fed-batch phase ($g\ L^{-1}$) |
| $J_t$ | Objective function for fed-batch harvest time |

| | |
|---|---|
| $m_{O_2}$ | Maintenance coefficient of oxygen (g oxygen $g^{-1}$ biomass $h^{-1}$) |
| $m_{S_b}$ | Maintenance coefficient of substrate in batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) |
| $m_{S_f}$ | Maintenance coefficient of substrate in fed-batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) |
| $N_i$ | Stirrer speed ($h^{-1}$) |
| $N_p$ | Power number (-) |
| $N_v$ | Cell number density (cells $mL^{-1}$) |
| $O_2$ | Oxygen in gas phase (mol) |
| $O_{2,in}$ | Pure oxygen input (mol) |
| $p$ | Pressure (kPa) |
| $P$ | Product concentration (g $L^{-1}$) |
| $P_{vol}$ | Cell volume fraction (dimensionless) |
| $P_T$ | Power (watt) |
| $q_{O_{2,b}}$ | Specific oxygen consumption rate in batch phase (g oxygen $g^{-1}$ biomass $h^{-1}$) |
| $q_{O_{2,f}}$ | Specific oxygen consumption rate in fed-batch phase (g oxygen $g^{-1}$ biomass $h^{-1}$) |
| $q_P$ | Specific product production rate (g product $g^{-1}$ biomass $h^{-1}$) |
| $q_{S_b}$ | Specific substrate consumption rate for substrate in batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) |
| $q_{S_f}$ | Specific substrate consumption rate for substrate in fed-batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) |
| $r$ | Cell radius (m) |
| $R$ | Gas constant ($K^{-1}$ kPa L $mol^{-1}$) |
| $S_b$ | Substrate concentration in batch phase (g $L^{-1}$) |

| | |
|---|---|
| $S_f$ | Substrate concentration in fed-batch phase (g $L^{-1}$) |
| $S_{f,in}$ | Substrate concentration in the feed in fed - batch phase (g $L^{-1}$) |
| $T$ | Temperature (K) |
| $t_b$ | Time for batch-phase (h) |
| $t_{end}$ | Fed-batch harvest time (h) |
| $t_f$ | Final reactor time for fed-batch phase (h) |
| $t_i$ | Time of induction (h) |
| $t_{int}$ | Time interval for manipulating variable (min) |
| $u_g$ | Superficial gas velocity (m $h^{-1}$) |
| $V$ | Volume of the broth (L) |
| $V_g$ | Gas volume (L) |
| $V_T$ | Total reactor volume (L) |
| $X$ | Biomass concentration (g $L^{-1}$) |
| $X_{in}$ | Biomass concentration in the feed (g $L^{-1}$) |
| $y_{X/S_b}$ | Biomass yield coefficient from substrate in batch phase (g biomass $g^{-1}$ substrate) |
| $y_{X/S_f}$ | Biomass yield coefficient from substrate in fed-batch phase (g biomass $g^{-1}$ substrate) |
| $y_{X/O_{2,b}}$ | Biomass yield coefficient from oxygen utilized in batch phase (g oxygen $g^{-1}$ biomass) |
| $y_{X/O_{2,f}}$ | Biomass yield coefficient from oxygen utilized in fed-batch phase (g oxygen $g^{-1}$ biomass) |

***Greek***

| | |
|---|---|
| $\alpha$ | Growth associated product yield coefficient (g product $g^{-1}$ biomass) |
| $\alpha$, $\alpha_2$ | Cole-Cole parameter for capacitance and conductivity (dimensionless) |
| $\alpha$, $\beta$, $\gamma$, $\delta$ | Dispersions in the dielectric spectra |

| | |
|---|---|
| $\beta$ | Non-growth associated product yield coefficient (g product $g^{-1}$ biomass $h^{-1}$) |
| $\Delta\varepsilon$ | Dielectric increment (dimensionless) |
| $\varepsilon_0$ | Permittivity of free space (F $m^{-1}$) |
| $\varepsilon_\infty$ | High frequency permittivity (dimensionless) |
| $\lambda$ | Penalty factor for epsilon method to obtain pareto points |
| $\mu_b$ | Specific growth rate of batch phase ($h^{-1}$) |
| $\mu_f$ | Specific growth rate of fed-batch phase ($h^{-1}$) |
| $\mu_{m_b}$ | Maximum specific growth rate of batch phase ($h^{-1}$) |
| $\mu_{m_f}$ | Maximum specific growth rate of fed-batch phase ($h^{-1}$) |
| $\rho$ | Density (kg $m^{-3}$) |
| $\Delta\sigma$ | Conductivity increment (S $m^{-1}$) |
| $\sigma_L$ | Low frequency conductivity (S $m^{-1}$) |
| $\sigma_i$ | Internal conductivity (S $m^{-1}$) |
| $\sigma_e$ | Medium conductivity (S $m^{-1}$) |
| $\phi$ | Proportionality factor for biomass and therapeutic protein |
| $\omega$ | Measuring frequency (rad $s^{-1}$) |
| $\omega_c$, $\omega_{c2}$ | Characteristic angular frequency for capacitance and conductivity (rad $s^{-1}$) |

***Abbreviations***

| | |
|---|---|
| ANN | Artificial neural network |
| ATF | Alternating tangential flow filtration |
| BM | Biomass monitor |
| CAGR | Compound annual growth rate |
| CER | Carbon dioxide evolution rate |

| | |
|---|---|
| CFDA | 5-carboxyfluorescein diacetate succinimidyl ester |
| CHO | Chinese Hamster ovary |
| CPP | Critical process parameters |
| CPR | Carbon dioxide production rate |
| CQA | Critical quality attributes |
| CV | Controlled variable |
| DCW | Dry cell weight |
| DLS | Dynamic Light Scattering |
| DO | Dissolved oxygen |
| DoE | Design of experiments |
| DR | Diabetic retinopathy |
| DS | Dielectric spectroscopy |
| EIS | Electrochemical impedance spectroscopy |
| EKF | Extended Kalman filters |
| FDA | Food and drug administration |
| FESEM | Field emission scanning electron microscope |
| FTIR | Fourier transform infrared spectroscopy |
| GA | Genetic algorithm |
| HCDC | High cell density cultivations |
| HER2 | Human epidermal growth factor receptor 2 |
| HPLC | High-performance liquid chromatography |
| IB | Inclusion body |
| IPTG | isopropyl $\beta$-d-1-thiogalactopyranoside |
| LMA | Levenberg-Marquart algorithm |
| MA | Moving average |
| MIMO | Multiple inputs and multi output |
| MISO | Multiple inputs and single output |
| MOO | Multiobjective optimization |
| MPC | Model predictive control |
| MV | Manipulated variable |

| | |
|---|---|
| MVDA | Multivariate data analysis |
| NIR | Near-Infrared Spectroscopy |
| NMPC | Nonlinear model predictive control |
| NMR | Nuclear magnetic resonance |
| NVAMD | Neovascular age-related macular degeneration |
| OD | Optical density |
| ODE | Ordinary differential equations |
| OTR | Oxygen transfer rate |
| OUR | Oxygen Uptake rate |
| PAT | Process Analytical Technology |
| PCA | Principle component analysis |
| PCC | Periodic counter-current chromatography |
| PI | Propidium iodide |
| PID | Proportional Integral Derivative |
| PLS | Partial least square regression |
| PV | Process variable |
| QbD | Quality by design |
| RMSE | Root mean squared error |
| RSM | Response surface methodology |
| SCADA | Supervisory Control and Data acquisition |
| SF | Smoothness factor |
| SG | Savitzky-Golay |
| SOC | Super optimal broth with catabolite repression |
| SQP | Sequential quadratic programming |
| SS | Smoothing splines |
| SSE | Sum of squared estimate of errors |
| TFA | Trifluoroacetic acid |
| VCC | Viable cell concentration |
| VCV | Viable cell volume |
| VEGF | Vascular endothelial growth factor |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# Introduction

## 1.1 Introduction

With the advent of the recent global pandemic, the significance of vaccines as an unequivocal remedy to mitigate the crisis was evident, and their rapid development revealed the capability of next-generation bioprocessing. The approval of recombinant insulin marked the production of the first biopharmaceutical, and since then, the industry has dramatically expanded. The production of biopharmaceuticals for therapeutic application has increased tremendously worldwide with a total sales value of around USD 325 billion in 2020 and is expected to reach a revenue of USD 496 billion by 2026 (as shown in Figure 1.1), and has become one of the fastest-growing sectors of the pharmaceutical industry ("Biopharmaceuticals Market (2021 - 2026)," 2021; Walsh, 2018, 2014). A consistent rise in the demand for the development and production of life-saving drugs has led to considerable investments in the biopharmaceutical sector and rapid growth in the number of companies producing therapeutic products.

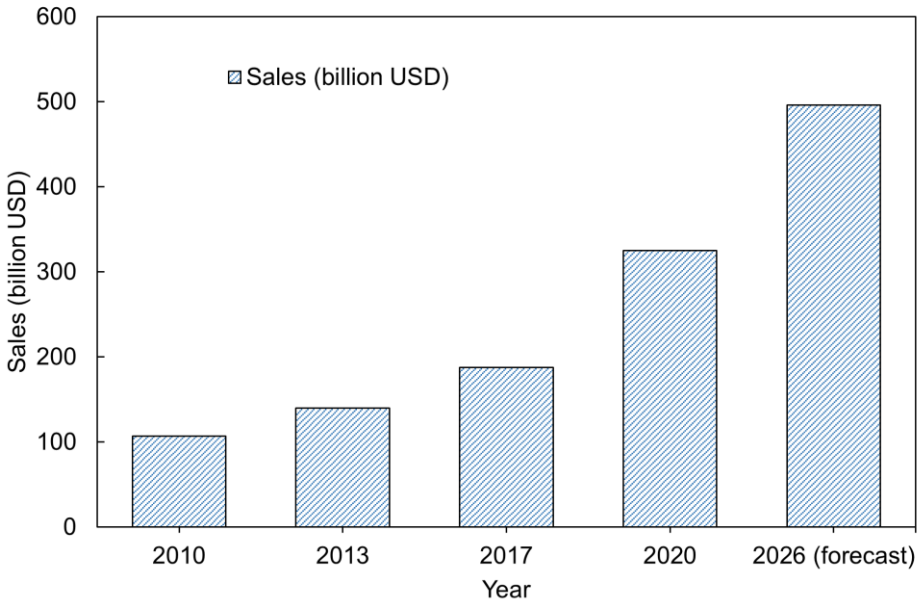



Figure 1.1. Annual revenue of biopharmaceuticals from 2010-2020 with the forecast for 2026

In accordance with the fourth industrial revolution, with the arrival of artificial intelligence and the growing interest in data science and computer-aided simulations, a tremendous revolution in the biopharmaceutical sector is anticipated (Chen et al., 2020; Gargalo et al., 2020). The main challenges concerning the implementation of Industry 4.0 to the bioprocessing sector would be, handling the delicate living organisms and

uncertainty surrounding the adaptation of new technologies. Real-time data collection using advanced sensors and implementation of suitable optimization strategies would channelize the process towards desired objectives and immensely lessen the manufacturing cost through effective resource utilization. Process optimization would especially be crucial in the therapeutic production process as millions of lives depend on the timely delivery of quality products. Autonomous bioprocessing with the aid of artificial intelligence could effectively handle the bottlenecks to ensure product quality and quicken the manufacturing process, and the biopharmaceutical sector is heading towards the same in the upcoming years. This thesis sets out to explore the application of different aspects of measurement, modelling and optimization to enhance the performance of the therapeutic production process. This chapter outlines the background behind the important concepts explored in this thesis.

### 1.1.1 Overview of Biotherapeutics

Biological therapeutics, or Biologicals, are a class of medicines in which the active substance is extracted or produced from different biological sources for therapeutic usage ("WHO: biologicals," 2018). Biotherapeutics have become an integral part of modern medicines and are indispensable to diagnose, mitigate and treat several life-threatening or chronic diseases, including cancers and autoimmune diseases. Biotherapeutics can be derived from various sources, including bacteria, yeast, plant and animal cells, and this class of medicines include recombinant proteins, vaccines, growth factors, monoclonal antibodies, and more. These therapeutic products are produced by biological means by harnessing the potential of genetically engineered living organisms through the advancements of recombinant DNA technology. The manufacturing process of biotherapeutics is complicated due to the larger size of these compounds and sophisticated control strategies that are required to ensure the quality and efficacy of the final product.

With Food and Drug Administration (FDA) approvals skyrocketing in recent years, the number of approved therapeutic products reached 112 during 2015-18, almost double the numbers until 2014. The growing interest towards the production of monoclonal antibodies (mAbs) has been reflected in the total percentage of mAb approvals reaching up to 53% out of total therapeutic product approvals in the recent period (Walsh, 2018, 2014). The production of biosimilars that are highly similar to the original biologicals is also rising in recent years, thereby contributing to an increasing number of biotherapeutic

approvals. Some examples of therapeutic products produced from different sources are presented in Table 1.1.

Table 1.1. Examples of therapeutic products produced from various sources

| **Bio-therapeutic product** | **Cumulative sales 2014-17 ($ billions)** | **Company, year first approved** | **Source** | **Application** |
|---|---|---|---|---|
| Humira (Adalimumab, anti-TNF) | 62.6 | AbbVie, Eisai, 2002 | Chinese Hamster ovary (CHO) cells | Rheumatoid arthritis |
| Rituxan (Rituximab; anti-CD20) | 29.1 | Roche, Biogen Idec, 1997 | CHO cells | Follicular lymphoma |
| Lantus (insulin glargine) | 27.4 | Sanofi, 2000 | *Escherichia coli* | Diabetes mellitus |
| Herceptin (trastuzumab, anti-HER2) | 27.1 | Roche, 1998 | Murine cell line | Metastatic breast cancer |
| Avastin (bevacizumab; anti-VEGF) | 27 | Roche, 2004 | CHO cells | Metastatic colorectal cancer |
| Neulasta (Pegfilgrastim) | 20.1 | Amgen, 2002 | *E. coli* | Chemotherapy-induced neutropenia |
| Lucentis (Ranibizumab; anti-VEGF) | 14.3 | Roche, 2006 | *E. coli* | Neovascular (wet) age-related macular degeneration |
| Opdivo (nivolumab; anti-PD-1 receptor) | 11.4 | Bristol-Myers Squibb, Ono Pharmaceutical, 2014 | CHO cells | Melanoma, renal cell carcinoma |
| Stelara (ustekinumab; anti-IL-12 & IL-23 | 12.2 | Janssen Cilag (Johnson & Johnson), 2009 | Sp2/0 cells | Moderate to severe plaque psoriasis |
| Novolog/ Novorapid (insulin aspart) | 11.7 | Novo Nordisk, 1999 | *Saccharomyces cerevisiae* | Diabetes mellitus |

TNF - tumor necrois factor; CD20 - cluster of differentiate 20; HER2 - human epidermal growth factor receptor 2; VEGF - vascular endothelial growth factor; PD-1 - Programmed cell death protein 1; IL - Interleukin.

### 1.1.2 Production methods: Continuous bioprocessing

The production of biopharmaceutical products typically begins with the upstream process, which comprises the generation of the living cells, followed by the downstream process, which includes the purification of the target product. Traditionally, the biopharmaceutical manufacturing processes were carried out using batch processes, with sufficient hold-up steps between different unit operations. Lately, the biopharmaceutical industry has been undergoing a paradigm shift towards continuous manufacturing from the traditional batch process. Previously, several industries, including steel casting, petrochemicals, and food, adapted the conversion of technology from batch to continuous, and its application in biotechnology has been gaining momentum in recent years.

In today's world, the manufacturers face a significant challenge in reducing the cost of the crucial life-saving drugs and facilitating their availability to patients in developing countries. Continuous processing has the potential to achieve this objective, and it could revolutionize the bioprocessing industry owing to the significant economic advantages it offers. Continuous bioprocessing refers to a steady-state operation and could be is defined as "A unit operation is continuous if it is capable of processing a continuous flow input for a prolonged period" (Konstantinov and Cooney, 2015). Among the several advantages of continuous processing, the significant ones include steady-state operation, smaller equipment footprint, lower cycle times, and many more (Jungbauer, 2013; Rathore et al., 2015; Zydney, 2015). Continuous bioprocessing also offers a tremendous advantage in enhancing the quality of the therapeutic product due to automation and steady-state operation.

Continuous processing for recombinant proteins has been extensively explored in the recent decade with the advent of relevant technologies that could facilitate the implementation of the same. Technologies like perfusion bioreactor, single-use bioreactors, Alternating tangential flow filtration (ATF) for the upstream processes, and high-pressure homogenizer (continuous cell lysis in case of an intracellular product), disk stack centrifuge, continuous annular chromatography for the downstream process have accelerated this conversion (Godawat et al., 2015; Holzer, 2017; Rathore et al., 2016, 2015; Rathore and Kateja, 2016).

With the advent of these rapidly growing technological advancements, the implementation of end-to-end continuous processing for bioprocesses is not a distant

dream anymore. The application of a continuous platform for the production of therapeutics was explored by Warikoo et al., (2012), wherein the perfusion bioreactor was integrated with a periodic counter-current (PCC) chromatography system for the continuous capture of the product. Despite having all the mentioned advantages, continuous processing also possesses significant challenges like sterility maintenance over an extended period and challenges related to start-up and shut-down that need to be addressed before successful industrial implementation (Zydney, 2015). Figure 1.2 illustrates the progression of biopharmaceutical manufacturing processes from a batch process to a fully integrated continuous process. A representation of a hybrid continuous manufacturing process with a batch upstream and continuous downstream process is also presented in Figure 1.2.

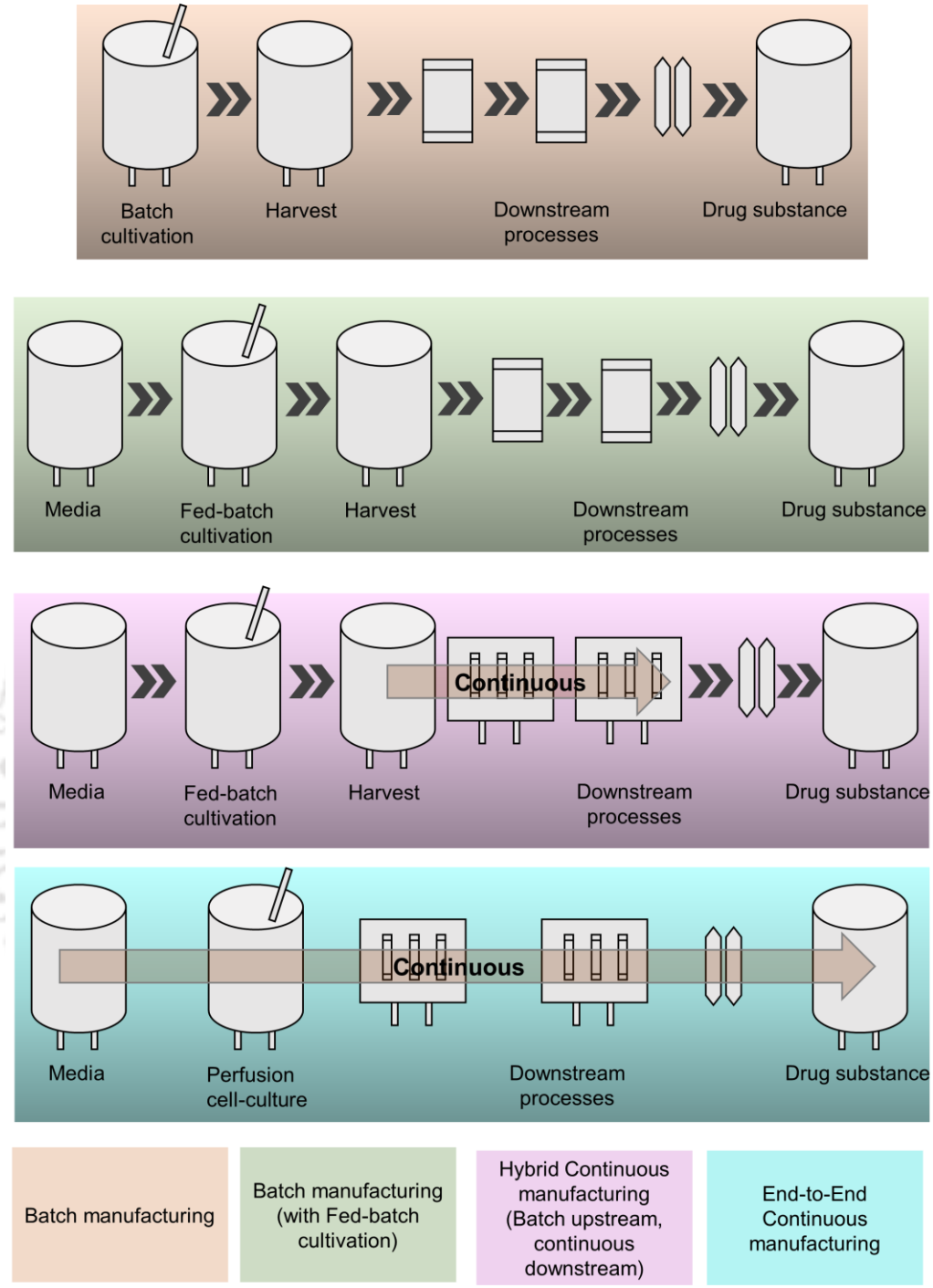



Figure 1.2. Representation of different manufacturing processes for biologicals. (adapted from (Konstantinov and Cooney, 2015)).

### 1.1.3 Significance of Process Analytical Technology implementation

Fermentation, which is the significant segment of the upstream processing, is the crucial step in the production of biotherapeutic proteins since the main drivers of the process, the

biomass, are cultivated in this stage. Most of the process parameters that influence the final product quality are monitored and measured during the upstream process. Hence, understanding the backbone of the therapeutic protein production process by employing suitable and efficient measurement tools and process models is indispensable in achieving the desired product quality. Availability of reliable real-time process measurements will enable the application of a robust control strategy to take necessary control actions, ensuring the product quality.

In order to reduce the process variation and maintain the product quality, the FDA defined Process analytical technology (PAT) as an initiative for regulating pharmaceutical manufacturing processes. PAT can be described as a tool for designing, analyzing, and controlling manufacturing through timely measurements of critical quality and performance parameters to ensure final product quality (FDA, 2004). Accordingly, the choice of the PAT tool to be implemented is determined by the critical process parameters (CPPs), which governs the critical quality attributes (CQAs) of the biotherapeutic product. Additionally, the PAT also guided the industry to move from quality by testing to the quality by design (QbD) approach, emphasizing the development and integration of various PAT tools (Glassey et al., 2011; Rathore, 2014). The application of various PAT tools can provide us insights into the physical, chemical and biological attributes of the process. The significant steps involved in the PAT implementation include: (i) *Identify*: design of experiments to identify the CQA and the respective CPPs governing the same, (ii) *Monitor*: deploy appropriate analytical tools to monitor the process, (iii) *Analyze*: statistical analysis of the monitored CQA and understand its relationship with product efficacy, and (iv) *Control*: initiate suitable control schemes to ensure the quality of the process (Streefland et al., 2013). The major elements of PAT can be termed under the four categories of measurement, monitoring, modelling and control ($M^3C$), as represented in Figure 1.3.

PAT tools or sensors employed inside a reactor must be typically *in-situ* sterilizable, temperature and pressure stable, and should not be affected by fouling. Some of the sensor attributes include repeatability, robustness, stability, linearity and shorter response time. PAT tools can be employed in different modes, namely online, at-line, in-situ and offline. The sensors could be in direct interface with the process fluids (*in-situ*) or applied in the vicinity of the reactor where samples are analyzed outside the reactor (at-line). They could be applied for continuous measurement (online), or the analyses could be done separately

away from the reactor (offline) (Biechele et al., 2015; Gnoth et al., 2008; Vojinović et al., 2006). However, the development of PAT tools and reliable online sensors remains a challenge to be addressed in view of continuous processing for biotherapeutic products since it requires a thorough understanding of the process and product as well as real-time decision making with appropriate controls (Konstantinov and Cooney, 2015; Rathore, 2016, 2014; Streefland et al., 2013). An overview of various PAT tools is summarized in Table 1.2 (Rathore et al., 2010; Streefland et al., 2013).

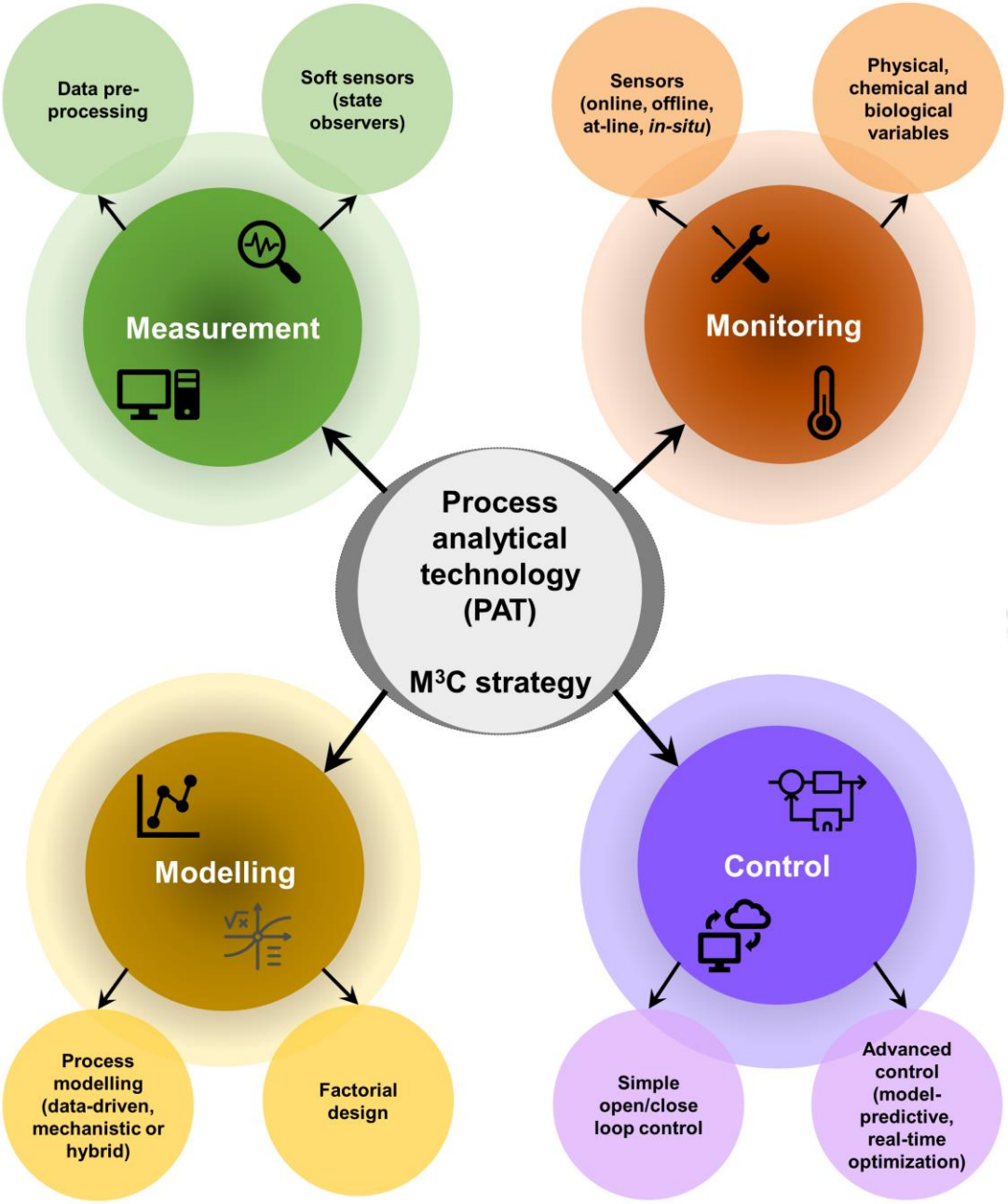



Figure 1.3. Various segments of Process Analytical Technology implementation

Table 1.2. Application of different PAT tools in bioprocesses

| Process variable | Process analyzer | PAT application | Monitoring mode | Reference |
|---|---|---|---|---|
| **Chemical** | | | | |
| pH | Potentiometric/ luminescence-based pH electrode | pH control in the fermenter | Online/*In-situ* | (Gnoth et al., 2010) |
| Oxygen/Carbon dioxide | $O_2$ and $CO_2$ sensors | Dissolved $O_2$/$CO_2$ concentration measurement for feed rate control | Online/*In-situ* | (Kuprijanov et al., 2009) |
| Substrates and metabolites | Near-infrared spectroscopy (NIR) | Monitoring of metabolite concentration like glycerol and methanol | Online/*In-situ* | (Goldfeld et al., 2014) |
| | Fourier transform infrared (FTIR) spectroscopy | The concentration of various analytes and feedback control | Online | (Gomes et al., 2015) |
| | Flow-injection analysis | Metabolite concentration monitoring for fed-batch control | Online | (Bracewell et al., 2002) |
| | High-performance liquid chromatography (HPLC) | Analysis of nutrient and metabolite concentrations during cultivation and control overflow metabolism | On/At/Offline | (Gustavsson and Mandenius, 2013) |
| | Infrared and Raman spectroscopy | Detailed chemical information of compounds | On/At-line/*In-situ* | (Abu-Absi et al., 2011) |
| Volatile gases | Mass spectroscopy | Gas analysis ($O_2$, $CO_2$, $CH_4$) and control of the metabolic rate | On/Offline | (Heinzle et al., 1990) |

| | | | | |
|---|---|---|---|---|
| Viscosity | Process viscometry | Process viscosity monitoring of dairy concentration processes. | Online | (O'Shea et al., 2019) |
| **Biological** | | | | |
| Biomass concentration | Dielectric spectroscopy (DS) | Analysis of membrane potential to assess the cell viability and cell size to develop process control strategies | Online/*In-situ* | (Ma et al., 2019) |
| | NIR spectroscopy | Monitoring of cell culture and fermentation | Online/*In-situ* | (Hakemeyer et al., 2012) |
| Metabolic activity | Biocalorimetry | Analysis of heat generated by the metabolic activity of the cells | Online | (Mohan and Sivaprakasam, 2017) |
| | 2D Nuclear magnetic resonance (NMR) spectroscopy | Control of the metabolic activity of cells | Online | (Goudar et al., 2010) |
| Cell morphology | Real-time imaging / *In-situ* microscopes | Online microscopic analysis of the cultured organism and real-time determination of cell variables | Online/*In-situ* | (Höpfner et al., 2010) |
| | Flow cytometry | Analysis of cellular morphology, investigate and monitor cell growth, viability and cell size | At/Offline | (Kuystermans et al., 2012) |
| Cell metabolism | 2D Fluorescence spectroscopy | Monitoring many biomolecules and metabolites with fluorescent properties (amino acids and coenzymes), viable and dead cells | Online/*In-situ* | (Ohadi et al., 2014) |

#### 1.1.3.1 Model development in bioprocesses

Measurement, monitoring, modelling and control are the four significant pillars of any bioprocess system (Mandenius, 2004). In view of QbD and PAT guidelines outlined by the FDA, there is an increased interest in the development of mechanistic models that can relate the desired CQAs of biologics to the measured CPPs from the sensors (FDA, 2004; Marison et al., 2013; Rathore, 2014; Rathore et al., 2010; Read et al., 2010; Streefland et al., 2013). Model development is a crucial step, as it attempts to relate the measured CPPs to the desired CQAs. Mechanistic models offer reliable predictions and are more suitable for developing effective control strategy and overall process optimization (Mears *et al.*, 2017a). The highly complex nature of biological processes and dynamic non-linearity in the responses are the major factors limiting the development of a reliable process model. In spite of its demerits, the model development for a therapeutic protein production process garners much attention owing to its prediction capability and has been attempted by many researchers (Ashoori et al., 2009; Ko and Wang, 2007; Wechselberger et al., 2013). The additional challenge includes the lack of availability of reliable sensors to measure all the process variables, and an appropriate solution would be the development of soft sensors that can measure the unmeasurable variables (Ko and Wang, 2007; Luttmann et al., 2012; Reichelt et al., 2016). In brief, the application of advanced monitoring tools and a well-designed process model could significantly enhance consistency in process and product quality through appropriate real-time control actions (Konstantinov and Cooney, 2015; Rathore, 2016, 2014).

Application of process models as a 'digital twin' that could represent the underlying physical process facilitates the implementation of optimization and control studies, which is relevant with the recent progression of the biopharmaceutical sector from batch to the continuous production process. Models can enable a virtual representation of the physical system by integrating real-time data from the analytical tools, which could foster future predictions, and eventually lead to an integrated automated smart bioprocess (Chen et al., 2020; Park et al., 2021). Additionally, soft sensors play a vital role in developing correlations to interpret the critical quality parameters that may not be directly measurable using the sensors. Performing simulations on a validated process model can minimize the use of expensive raw materials and address the bottlenecks beforehand, thereby ensuring the quality of the manufacturing process (Luttmann et al., 2012).

#### 1.1.3.2 Optimization and advanced process control

For any biotherapeutic production application, the performance of the process is determined by the product quality. Thus, in addition to maximizing the productivity of the process, quality is also significantly important. Implementing a suitable control strategy that efficiently handles the fluctuations in the raw material variability and the process conditions can result in consistent product quality. Additionally, appropriate control strategies that could maintain the process variables to the desired levels can improve the process performance. Therefore, the application of optimization and advanced control strategy is relevant to reduce the batch-to-batch variability in biotherapeutic production (Simutis and Lübbert, 2015). As mentioned previously, applying a suitable control strategy is also a primary attribute of PAT implementation for therapeutic production processes. Implementing the emerging PAT tools that could facilitate advanced control strategies can lead to holistic process control, enabling the manufacturer to improve product quality. (Glassey et al., 2011; Rathore et al., 2010).

The two essential elements before implementing any advanced process control are reliable monitoring of the critical process attributes and the availability of a validated process model that can correlate the CPPs and the CQAs. Some of the crucial CQAs are the function of the upstream unit operations concerning therapeutic protein production. Therefore, maintaining optimal reactor conditions like pH, temperature, and Dissolved Oxygen (DO) is critical to obtain a consistent product profile (Gomes et al., 2015). Additionally, there exists a high demand for real-time measurements of the process variables other than the traditional variables that can provide further information regarding the state of the process. The optimization and control strategy implementation goals can include maintaining the respective process variables at desired set points or a minimization/maximization problem. Possible control objectives could be maximizing process yield, maximizing product concentration, minimizing by-product formation, and so forth (Mears et al., 2017b).

The dynamic nonlinear nature of the biotherapeutic production processes insists on the necessity of implementing an advanced process control strategy that could handle the dynamic changes (Rathore et al., 2021). Furthermore, an advanced controller could efficiently handle the interaction between the multiple inputs and outputs of the process (MIMO). Combined with the mentioned advantages and the prediction capability of the

controller, model predictive control (MPC) has been receiving attention in recent years (Sommeregger et al., 2017). However, the application of advanced process control strategies might face certain hurdles due to the complexity of the underlying biological process, and therefore, developing reliable process models is an essential initial step towards the same. Another challenge would be the dynamic changes occurring throughout the process, and this could be addressed by implementing computer-aided strategies of real-time optimization and supervisory control strategy that could efficiently handle the variations (Junker and Wang, 2006).

## 1.2 Scope and motivation of the research

The proposed work aims to enhance the biotherapeutic protein production process by implementing the significant concepts of PAT. As discussed previously, obtaining reliable process measurements using advanced PAT tools and an appropriate process model to interpret the vital physiological variables is the initial step. The execution of optimization and advanced process control strategies are the next steps in marching towards achieving enhanced process performance. Different PAT tools could be employed to measure and monitor the critical process parameters, of which the biomass concentration is one of the significant physiological process variables. Estimation of this physiological variable is pertinent, especially in intracellular therapeutic protein production cases, wherein the productivity is correlated to the total biomass in the reactor. In this study, the implementation of the PAT tool, dielectric spectroscopy (DS), was explored to monitor and estimate biomass concentration. Interpreting reliable information from the analytical tools is a significant step to developing and implementing relevant optimization and control strategies. Therefore, this study focused on developing and validating suitable models to estimate the appropriate physiological variable from the DS measurements.

Although a multitude of literature states that scanning DS measurements can reveal cell physiological properties, the potential of utilizing the scanning frequency to take real-time process decisions in recombinant *Escherichia coli* fermentation has not been explored much to the best of our knowledge (Randek and Mandenius, 2020). Additionally, among the various methods available for developing capacitance and biomass correlations, the Cole-Cole model is considered a suitable choice to gain physiological insight into the organism. The Cole-Cole model for establishing nonlinear

correlations between scanning capacitance and the measuring frequency is generally solved using the Levenberg-Marquart algorithm (LMA). However, since LMA provides a local minimum when the model is nonlinear and nonconvex, which might not necessarily be the global solution, it might eventually estimate erroneous model parameters and physiological properties (Marquardt, 1963). A new approach using the global evolutionary algorithm, namely the genetic algorithm (GA), could be considered as an alternative to overcome this issue.

Developing a validated process model to describe the underlying therapeutic protein production process is an essential task, and it poses specific challenges, as discussed previously. In addition to that, the incorporation of the real-time process measurements in the developed process model to enable subsequent implementation of real-time optimization and control strategies is vital. Integration of the prediction capability of a mathematical model with the real-time DS measurements is likely to result in a more accurate prediction of the biomass concentration–based harvest, thereby attaining optimal protein production. Therefore, this study attempted to integrate the estimated process measurements from DS with the validated process model.

Product concentration and productivity of a biologic are significant for commercialization, and thus optimization of the total biomass in a reactor plays a significant role (Kaiser et al., 2008; Rathore, 2016). Several researchers have attempted to apply various optimization objectives, including biomass maximization and undesired product minimization, wherein different manipulated variables (MV) like substrate feed rate and air feed rate have been implemented (Atasoy et al., 2013; Geethalakshmi et al., 2011; Ko and Wang, 2006; Ochoa, 2016; Zafira and Nandong, 2019). Since the downstream operation accounts for a significant production cost for any biologicals, achieving maximum biomass with a minimum reactor volume will be beneficial concerning enhanced productivity and economics and thus could serve as an alternative for the generally developed cost-based objective functions if the total production cost is unknown (Raftery et al., 2017; Rathore, 2016). Therefore, combining the aforementioned two conflicting objectives and thereby achieving higher volumetric productivity of a biologic would be relevant in view of growing interest in continuous processing in the biopharmaceutical industry (Konstantinov and Cooney, 2015).

## 1.3 Overview of optimization and advanced process control strategy implementation

The overall process for biotherapeutic protein production from a bacterial platform, say *Escherichia coli*, can be described in a process flowsheet as shown in Figure 1.4. The production process involves a batch upstream followed by a continuous downstream process, which begins with inclusion body (IB) isolation through cell lysis followed by refolding and purification of the therapeutic product. The present study focuses on the potential of advanced PAT tools to estimate key process variables, followed by developing a reliable process model. The study then explores optimization and advanced control strategies implementation for the upstream processing of biotherapeutic protein production from *E. coli*.

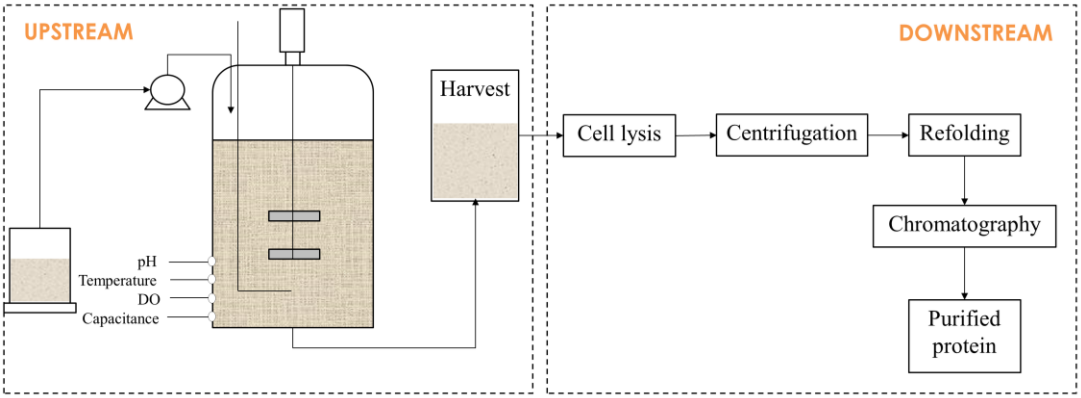



Figure 1.4. Process flowsheet for the production of therapeutic protein from *Escherichia coli*

Application of optimization and advanced process control strategies are beneficial in reducing the process variations, thereby maintaining the quality of the desired product. Since product quality is one of the CQA in any bioprocess, applying suitable control schemes to maintain the same is one of the aspects of PAT implementation, as discussed in the previous sections. The dynamic and highly nonlinear nature of the bioprocesses with changing yield coefficients emphasizes the application of a supervised control strategy to handle the dynamic changes. Advanced process control strategies such as MPC can handle the interactions that might happen between inputs and outputs. Additionally, the predictive nature of advanced process control is beneficial to take real-time actions and is relevant in reducing the batch-to-batch variations of the process.

The methodology for implementation of the advanced control strategy can be described in Figure 1.5. The first level consists of the various PAT tools like dielectric spectroscopy and exhaust gas analyzer, which monitors various process variables. The measurements from the sensors are fed into the respective regulatory controllers, and this regulatory layer maintains the process at desired set points. Online optimization and advanced process control strategies are implemented as a supervisory layer which provides the set-points for the regulatory layer (Proportional-Integral-Derivative (PID) control), which further keeps the underlying process in control by providing suitable manipulating variables (MVs).

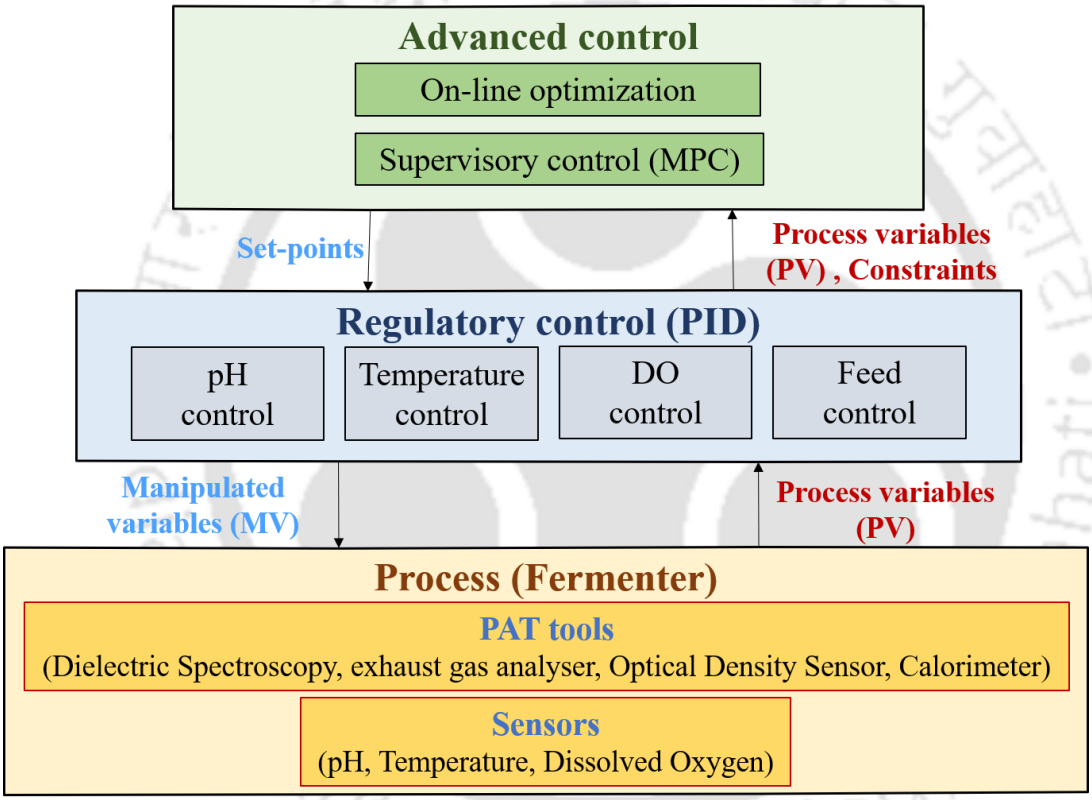



Figure 1.5. Overview of implementation of advanced process control strategies like on-line optimization and supervisory control (model predictive control (MPC)) in bioprocesses.

## 1.4 Objectives

Case study and the current system of investigation: Production of biotherapeutic protein Ranibizumab (Lucentis) as an antibody fragment in *E. coli*. The objectives of this work could be summarized as follows.

1. Application of dielectric spectroscopy for real-time monitoring of biotherapeutic protein production

This overall objective will focus on obtaining reliable process measurements from advanced PAT tools such as dielectric spectroscopy. The sub-objectives are as follows:

a. Capacitance monitoring in reactor studies of biotherapeutic protein production from *E. coli*.
b. Data pre-treatment for obtaining reliable capacitance signals.
c. Application of linear and Cole-Cole model for the development of capacitance-biomass correlations.
d. Real-time estimation of physiological properties of biomass using a robust methodology.

2. Modelling and validation of batch and fed-batch process of the fermenter

This overall objective will focus on developing a mechanistic model based on the mass balances and validation of the same. The sub-objectives are as follows:

a. Mechanistic model development based on mass balances of the fermenter.
b. Sensitivity analysis of the parameters.
c. Parameter estimation for calibration of the developed model.
d. Model validation to validate the competence of the developed model

3. Optimization studies for maximizing biomass production and predicting optimal harvest time

This overall objective will focus on the application of optimization studies using the validated model. The sub-objectives are as follows:

a. Optimization studies to be carried out for two cases:
   i) Case study (1): Optimizing the total biomass at a minimum broth volume
   ii) Case study (2): Predicting optimal fed-batch harvest time
b. Simulation studies for choosing the manipulating variable, boundary conditions and time interval
c. Development of multiobjective optimization to cater for the conflicting objectives
d. Prediction of optimal fed-batch harvest time

## 1.5 Organization of the thesis

The doctoral thesis is organized into six chapters, and the summary of these chapters are described as follows.

*Chapter 1* presents the overview and general introduction of the key concepts related to the biotherapeutic protein production process. The chapter also outlines the scope and motivation of the presented thesis, consolidates the current study's objectives, and finally presents the overall organization of the thesis.

*Chapter 2* presents the current state of the art relevant to the presented work and outlines the gaps and challenges in the existing literature.

*Chapter 3* presents the application of an established process analytical technology tool, namely, dielectric spectroscopy, in the reactor studies of the system. The chapter provides a brief theory behind the technology and elucidates the different models available for biomass estimation. This chapter presents the details about the reactor studies carried out along with the measured online and offline datasets and the analytical studies for validation purposes. The chapter further explains the different models to estimate biomass from the measured online capacitance data. Further, this chapter presents the implementation of a robust methodology using the Cole-Cole model for the real-time estimation of the physiological properties of the biomass.

*Chapter 4* presents the details regarding the development of a mechanistic model for the batch and fed-batch process of the fermenter. This chapter explains the model equations describing the process of biotherapeutic protein production from *E. coli.* This chapter then briefs the sensitivity analysis and explains the procedure for estimating the parameters present in the developed model in detail. Furthermore, the chapter explains the model validation methodology and highlights the usability of the validated model towards the execution of appropriate optimization and control studies.

*Chapter 5* presents the implementation of various optimization studies by elucidating different case studies formulated to achieve desired objectives. The case study (1) focused on developing a multiobjective optimization strategy to maximize total biomass at a minimum broth volume of the fermenter. This chapter explains the simulation studies that decided the constraints and the time interval for the chosen manipulating variable. The necessity of implementing a multiobjective optimization was inferred from the simulation studies, as explained in this chapter. Furthermore, a second case study aimed to predict the optimal fed-batch harvest time of the fermenter. This chapter thus summarizes the implementation of the different case studies for optimization, enhancing the production of the therapeutic protein considered in this study.

*Chapter 6* presents the overall summary and highlights the conclusions from each of the objectives presented in the previous chapters. This chapter concludes with some recommendations for the future and thereby serves as the overall conclusion of the thesis.

# CHAPTER 2

# Review of literature

## 2.1 Foreword

Biotherapeutics are therapeutic recombinant proteins derived from biological sources using different expression systems like microbial and mammalian culture and are applied to treat numerous life-threatening diseases, like cancer and autoimmune diseases ("WHO: biologicals," 2018). Among the available expression systems, the microbial platform is preferred due to its widely understood genomics and ease of manufacturing (Baeshen et al., 2015). *E. coli* has been the host of choice for producing a majority of recombinant proteins due to several well-established reasons, which is attributed to the organism's ability to achieve high cell density within a short time interval and well-characterized genetics and metabolism (Baeshen et al., 2015). Ranibizumab (Lucentis) is a 48 kDa non-glycosylated humanized monoclonal antibody fragment, mostly produced as inclusion bodies in the *E. coli* expression system. This therapeutic product was first approved by FDA in 2006, and it can be administered intravenously for the treatment of neovascular age-related macular degeneration (NVAMD), diabetic retinopathy (DR), and similar eye-related conditions (Kourlas and Abrams, 2007; Priyanka and Rathore, 2021; Vaidyanathan and Moshirfar, 2021). Ranibizumab was one of the 20-top selling therapeutic products and had a cumulative sales value of 14.3 billion USD from 2014-2017 (Walsh, 2018). The significance of Ranibizumab and its therapeutic application emphasizes the necessity for improving the productivity of the same. Therefore, there is adequate scope for the application of enhanced monitoring and control strategies for the fed-batch production of recombinant *E. coli* cultivation expressing this therapeutic product.

This chapter summarizes the detailed review of the current state of the art relevant to the present work. The first section of this chapter explains the different instances of the application of DS for monitoring recombinant protein production processes and various models implemented for biomass estimation. The different modeling strategies applied in bioprocesses are summarized in the next section. The review of literature focusing on the implementation of different optimization and advanced control strategies is presented in the final section. The gaps and challenges from respective sections are outlined, followed by which this chapter concludes by consolidating the overall summary of state of the art. Hence, this chapter substantiates the significance of the present work in addressing the challenges present in the current state of the art.

## 2.2 Dielectric spectroscopy as a PAT tool for real-time monitoring

The bioreactor operation is essentially the most important step in the recombinant protein production process, as the physiological changes associated with the cellular population would impact product quality in any recombinant protein production process (Flores-Cosío et al., 2020; Ma et al., 2019; Randek and Mandenius, 2020; Rathore, 2016). Harvest of a batch is typically linked to the time of induction (Priyanka et al., 2019). Rather than using time as the control criterion, basing harvest on biomass concentration is likely to result in more consistent process performance. Therefore, apart from conventional measurements such as temperature, pH, and dissolved oxygen (DO), the real-time monitoring of the significant key parameter, biomass concentration, is highly desirable, as highlighted in recent review articles (Fung Shek and Betenbaugh, 2021; Reardon, 2021).

Biomass concentration could be measured using several hard sensors, including *in-situ* microscopy (Höpfner et al., 2010; Lüder et al., 2014), flow cytometry (Kuystermans et al., 2012), near-infrared spectroscopy (Goldfeld et al., 2014; Hakemeyer et al., 2012), dielectric spectroscopy (Dabros et al., 2010; Ehgartner et al., 2015), online optical density (Kiviharju et al., 2007), fluorescence spectroscopy (Ohadi et al., 2014), and biocalorimetry (Cole et al., 2015; Mohan et al., 2019). Additionally, it could also be estimated using mechanistic models or soft sensors like off-gas analysis (Reichelt et al., 2016; Wechselberger et al., 2013). Various methods of biomass estimation and their merits and demerits were previously discussed (Kiviharju et al., 2008, 2007). Among these different sensors, DS is an emerging PAT tool for the real-time estimation of biomass. DS is robust, non-invasive, and does not require complex data analysis (Markx and Davey, 1999; Yardley et al., 2000).

DS has been proven to measure cell concentration, cell death, changes in cell size in real-time, for the application in various organisms, including bacteria (Ehgartner et al., 2015; Kaiser et al., 2008), yeast (Katla et al., 2019; Maskow et al., 2008a) and mammalian cells (Downey et al., 2014; Ducommun et al., 2001; Metze et al., 2019). DS involves the application of an electric field in the broth, where viable cells with an intact cell membrane act as capacitors and can store the charge independently. Thus, the overall capacitance of the suspension is measured to estimate total viable biomass concentration (Markx and Davey, 1999; Yardley et al., 2000). Furthermore, DS measurements could be

used to predict the induction and harvest time and maximize the productivity of the protein (Ehgartner et al., 2015; Kaiser et al., 2008; Zitzmann et al., 2018).

### 2.2.1 Dielectric spectroscopy: Principle and applications

The theory behind electrical properties of cells and impedance spectroscopy was recorded around 60 years ago by Schwan (1957). Several years later, DS was first described as a successful technique for biomass monitoring by Harris and Kell (1985), and since then, it has been used for several bioprocess applications for the past three decades (Flores-Cosío et al., 2020; Harris and Kell, 1985). DS is based on the capability of viable cells with an intact cell membrane to act as capacitors upon the polarization of ions near the membrane surface while applying an electric field across the membrane. The response of viable biological cells suspended in a conductive medium to the applied electric field can be described by its conductivity (ability to conduct electrical charge) and capacitance or permittivity (ability to store electrical charge (polarization measure)). The measured capacitance (pF $cm^{-1}$) could be correlated to the total viable biomass concentration, and this DS technique has been applied for several systems, including bacteria (Ehgartner et al., 2015; Randek and Mandenius, 2020), yeast (Katla et al., 2019; Maskow et al., 2008a), and mammalian cells (Cole et al., 2015; Isidro et al., 2021).

Additionally, DS measurements are also closely related to the cell's physiological state, and several properties like cell concentration (Zitzmann et al., 2018), changes in cell size (Randek and Mandenius, 2020), cell viability (Tibayrenc et al., 2011), and cell death can be estimated (Ma et al., 2019). However, applying this PAT tool can be challenging in the case of bacteria, especially recombinant *E. coli*, due to the smaller size of the cells (Ehgartner et al., 2015), high signal noise in the capacitance measurements (Horta et al., 2015), and sensitivity of the capacitance probe with aeration and agitation (Lira-Parada et al., 2021). Therefore, it is essential to acquire reliable capacitance signals from the fed-batch cultivation of recombinant *E. coli* producing biotherapeutic protein for its application in estimating the physiological properties of the organism.

In essence, the DS studies are based on the electrical properties of the cells. The viable cells with an intact cell membrane get polarized by applying an alternating electric field in the radio frequency range (0.1-10 MHz). The cellular cytoplasm is highly conductive due to the presence of salts and nucleic acids, while the plasma membrane formed by the

lipid bilayer is comparatively less or non-conductive. When an electric field is applied to a cell suspension, cellular polarization occurs due to the movement of intracellular ions along the applied electric field. This movement is restricted by the plasma membrane, which acts as an insulating barrier between the intracellular ions and the ions present in the culture medium. As a result, the ions from both these regions move towards the oppositely charged electrode respectively. The non-conducting plasma membrane traps the ions near the membrane surface and enables the cells to act as capacitors; and therefore, the dead cells could not be polarized due to the absence of an intact cell membrane. The presence of an intact cell membrane is therefore essential for charge separation, which contributes to measured permittivity (or capacitance) values. Therefore, dead cells or damaged cells whose plasma membrane is lysed and other solids or media particles without a plasma membrane do not contribute to the overall permittivity values. The increase in volume fraction of cells leads to the rise in the number of membranes available for polarization which further increases the capacitance of the suspension. Therefore, the measured capacitance value could reflect the total volume fraction of cells present in the suspension (Davey et al., 1993; Harris et al., 1987). DS has been successfully implemented to estimate biomass concentration in recombinant *E. coli* cultivations (Ehgartner et al., 2015; Kaiser et al., 2008; Knabben, 2011; Randek and Mandenius, 2020). However, there is still scope for nonlinear model (Cole-Cole) validation and real-time estimation of physiological properties in therapeutic protein production processes. The schematic representation of capacitance-based monitoring is presented in Figure 2.1.

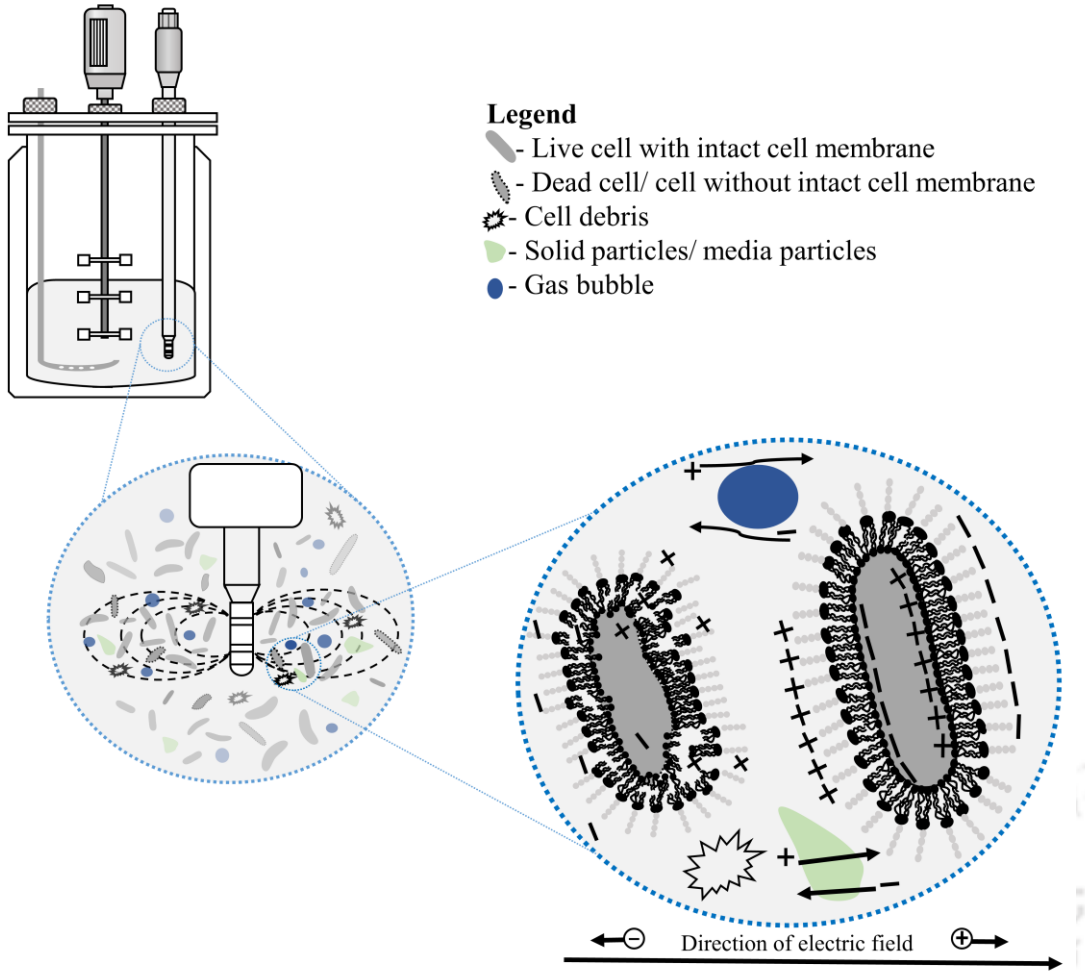



Figure 2.1. Schematic representation of capacitance-based monitoring in bioprocesses

The response of the material subjected to the applied electric field can be described by conductance ($G$) and capacitance ($C$), which are converted to conductivity [$\sigma$ (S m$^{-1}$)] and relative permittivity [$\varepsilon$ (dimensionless)], respectively, using the probe constant. With the increase in the frequency of the electric field, the permittivity of the cell suspension decreases while the conductivity increases in a step-like shift called dispersions, which occurs due to the loss of the cells' polarization abilities and increasing admittance of the plasma membrane (Davey et al., 1992). The resulting dielectric spectra consists of 4 main dispersions viz., $\alpha$, $\beta$, $\gamma$, and $\delta$ and is represented in Figure 2.2 (Schwan, 1957). The $\alpha$ dispersion occurring at low frequencies (< 10 kHz) is attributed to the diffusion of ions across cell surfaces, and the $\beta$ dispersion occurring in the radio-frequency range (0.4-10 MHz) results due to the build-up of charge at cell membranes due to the Maxwell-Wagner effect, and is of particular interest in case of biomass quantification. The $\delta$ dispersion

occurs between the $\beta$ and $\gamma$-dispersion in the high-frequency range (0.1-5 GHz) and is associated with the rotation of macromolecules, and the $\gamma$ dispersion occurs at ultra-high frequencies (>1 GHz) and is due to the dipolar rotation of water molecules. (Davey et al., 1992; Nasir and Al Ahmad, 2020).

In the case of biological cell suspensions, the $\beta$-dispersion representing the polarization process is of significance, wherein the frequency at which the electric field is reversed determines the rate of polarization. Complete polarization occurs at frequencies below 0.1 MHz since sufficient time is available before the reversal of field direction, resulting in high permittivity. With fewer ions reaching the plasma membrane at frequencies above 1 MHz, the reduced membrane polarization results in lower permittivity. The permittivity further decreases at high frequencies above 10 MHz since only a very few ions can be polarized before the electric field is reversed, and it corresponds to the residual permittivity (refer inset in Figure 2.2) (Flores-Cosío et al., 2020; Yardley et al., 2000). Analyzing the shape of the $\beta$ dispersion can provide insights into the cell's morphological state. The obtained dielectric increment is known to relate directly to the total biovolume ($P_{vol}$) as described by the equation (2.1), where $P_{vol}$ is calculated from the radius $r$ and cell number density $N_v$ as described in equation (2.2) (Schwan, 1957).

$$\Delta\varepsilon = \frac{9P_{vol}rC_m}{4\varepsilon_0} \tag{2.1}$$

$$P_{vol} = \frac{4}{3}\pi r^3 N_v \tag{2.2}$$

where $\Delta\varepsilon$ is the dimensionless dielectric increment, $P_{vol}$ is the dimensionless total volume fraction of cells, $r$ (m) is the average cell radius, $C_m$ (F m$^{-2}$) is the membrane-specific capacitance (usually assumed to be a constant), $\varepsilon_0$ (F m$^{-1}$) is the permittivity of free space and $N_v$ (cells mL$^{-1}$) is the cell number density. Therefore, one can measure the viable biomass concentration and the cell's radius by monitoring the dielectric increment $\Delta\varepsilon$ during the $\beta$ dispersion phase (Markx and Davey, 1999). Here $\Delta\varepsilon$ relates to the difference between low frequency and high frequency (residual) permittivity ($\varepsilon_\infty$) and is one of the characteristic parameters describing the $\beta$ dispersion.

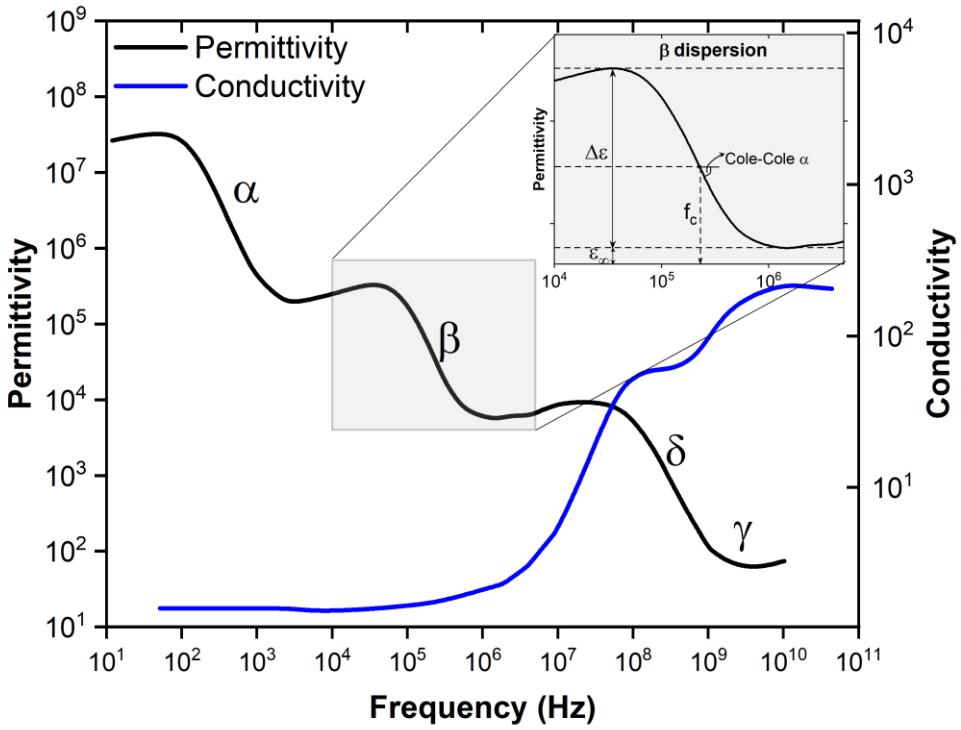



Figure 2.2. Dielectric spectra of cell suspensions representing the conductivity and permittivity with *α, β, γ,* and *δ* dispersions over the frequency spectrum. Inset: Schematic representation of the *β* dispersion with the parameters: permittivity increment (*Δε*), residual permittivity ($\varepsilon_\infty$), characteristic (or critical) frequency ($f_c$), and Cole-Cole *α*. (adapted from (Flores-Cosío et al., 2020; Nasir and Al Ahmad, 2020))

### 2.2.2 Models for estimation of physiological properties from capacitance measurements

The DS analysis can be performed either in dual-frequency mode or in frequency scanning mode, and the application of a four-annular–electrode system would reduce the electrode polarization and conductivity effects. In the dual-frequency mode, the *Δε* is measured at two different frequencies, i.e., one on the lower frequency plateau and another on the higher frequency of the *β* dispersion. Apart from the dual-frequency scanning, another capacitance measurement method is through frequency scanning, where the capacitance is measured over a range of electrical frequencies typically in the range 0.1–10 MHz, where the *β* dispersion occurs. The measured raw capacitance value

was pre-treated for noise removal, and then different models were used for estimating the physiological properties from the measured online capacitance values.

The capacitance measurements obtained from the fermentation process could be used for reliable biomass estimation by developing and deploying appropriate models. The measured capacitance data can be correlated to the viable biomass present in the reactor using a suitable calibration factor, and the correlations for biomass estimation can be developed using either linear or nonlinear models (Dabros et al., 2009b). Several reference techniques can be employed for developing this correlation (Marison et al., 2013). The linear model is the general and direct approach wherein a linear correlation is established between the measured real-time capacitance and the offline biomass values. This correlation could later estimate real-time biomass concentration from the measured capacitance values in future studies.

Even though the linear models are considered relatively simple, versatile, and applicable for different bioprocess systems, they could not provide insight into cellular physiology. Consequently, a sophisticated method capable of providing more information concerning cell physiology and reliable biomass estimation would be advantageous (Ehgartner et al., 2015; Ma et al., 2019). This additional information would be crucial in any recombinant protein production process, especially during the production of intracellular proteins in the form of inclusion bodies, where changes in cell viability and morphology occur during the course of fermentation (Ehgartner et al., 2017; Randek and Mandenius, 2020). Cole-Cole model is a theoretical nonlinear representation of the dielectric behavior of cell suspensions, which elucidates the dependency of scanning capacitance and conductivity data as a function of radiofrequency of the applied electromagnetic field (Cole and Cole, 1941). The parameters obtained from the nonlinear equations could aid in estimating additional physiological properties like cell diameter and cell viability (Dabros et al., 2009b; Díaz Pacheco et al., 2021; Opel et al., 2010; Tibayrenc et al., 2011). Reliable biomass estimation using appropriate models can eventually lead to the implementation of suitable control strategies using this critical process variable for improved process performance as demonstrated by several authors in different systems (Ehgartner et al., 2017; Knabben, 2011; Ma et al., 2019; Zitzmann et al., 2018). With the availability of real-time measurements of the organism's physiological state, there is more potential to develop robust control strategies from DS measurements (Flores-Cosío et al., 2020).

#### 2.2.2.1 Linear modeling

The linear model is widely applied for correlating the capacitance measured in the *β*–dispersion phase and the biomass concentration (Horta et al., 2015; Maskow et al., 2008a). The DS signals could be correlated with several reference techniques like cell counting, optical density values, colony-forming units, and dry cell weights (DCW) to get the linear correlation between the capacitance and the biomass concentration. For establishing a linear model, the capacitance is measured at a particular frequency that is fixed based on the type of organism, viz., 1000 kHz for bacteria and 500-600 kHz for yeast and mammalian cells. The background capacitance of the medium is measured at an elevated frequency of 15 MHz. The obtained value is then related to the biomass concentration by a calibration factor that is strain-dependent.

#### 2.2.2.2 Cole-Cole modeling

The Cole-Cole model describes the dependency of permittivity and conductivity upon the measuring frequency. The Cole-Cole equation is a variation of Debye's equation and is a theoretical way of describing the dielectric properties of cell suspensions and the shape of the *β* dispersion (Cole and Cole, 1941). The nonlinear equations describing the permittivity and conductivity with respect to the measuring frequency (equations (2.3) and (2.4)) can be solved using a nonlinear least square algorithm (Marquardt, 1963).

$$\varepsilon(\omega)=\left(\frac{\Delta\varepsilon\left(1+\left(\frac{\omega}{\omega_c}\right)^{(1-\alpha)}\sin\left(\frac{\pi}{2}\alpha\right)\right)}{1+\left(\frac{\omega}{\omega_c}\right)^{(2-2\alpha)}+2\left(\frac{\omega}{\omega_c}\right)^{(1-\alpha)}\sin\left(\frac{\pi}{2}\alpha\right)}+\varepsilon_\infty\right) \tag{2.3}$$

$$\sigma(\omega)=\left(\frac{-\Delta\sigma\left(1+\left(\frac{\omega}{\omega_{c_2}}\right)^{(1-\alpha_2)}\sin\left(\frac{\pi}{2}\alpha_2\right)\right)}{1+\left(\frac{\omega}{\omega_{c_2}}\right)^{(2-2\alpha_2)}+2\left(\frac{\omega}{\omega_{c_2}}\right)^{(1-\alpha_2)}\sin\left(\frac{\pi}{2}\alpha_2\right)}+(\sigma_L+\Delta\sigma)\right) \tag{2.4}$$

where $\varepsilon$ (dimensionless) and $\sigma$ (S $m^{-1}$) represent the permittivity and conductivity, respectively as a function of the measuring frequency $\omega$ (rad $s^{-1}$). The parameters in the Cole-Cole model can be used for describing the morphological characteristics of the cells. The parameter $\Delta\varepsilon$ (permittivity increment) can be correlated with the total volume fraction

enclosed by the cell suspension as described in the equation (2.5). The parameter $f_c$ ($\omega_c$), known as characteristic or critical frequency, represents the frequency at which measured permittivity is decreased by half and is indicative of the cell morphology as described in the equation (2.6). The empirical parameter Cole-Cole $\alpha$ (ranging from 0.1-0.2 for $\beta$-dispersion of biological cells) represents the steepness at which the permittivity decreases with increasing frequency and can provide insight into the distribution of cell sizes in suspension (Dabros et al., 2009b; Tibayrenc et al., 2011; Yardley et al., 2000). The schematic representation of the $\beta$-dispersion with the parameters described in the equation (2.3) is provided in section 2.2.1 (Figure 2.2).

$$\Delta\varepsilon = \frac{3N_v \pi r^4 C_m}{\varepsilon_0} \tag{2.5}$$

$$\omega_c = \frac{1}{rC_m\left(\frac{1}{\sigma_i} + \frac{1}{2\sigma_e}\right)} \tag{2.6}$$

$$\sigma_e = \frac{\sigma_L}{\left(1 - P_{vol}\right)^{1.5}} \tag{2.7}$$

The estimated parameters, dimensionless dielectric increment $\Delta\varepsilon$, characteristic frequency $\omega_c$ (radians $s^{-1}$) and the low–frequency conductivity $\sigma_L$ (S $m^{-1}$), along with intrinsic cellular properties like membrane capacitance $C_m$ (F $m^{-2}$) and internal conductivity $\sigma_i$ (S $m^{-1}$), are used for estimating the cell radius $r$ (m), cell number density $N_v$ (cells $mL^{-1}$), the conductivity of the suspending medium $\sigma_e$ (S $m^{-1}$) and volume fraction of cells $P_{vol}$.

The ion content of the media (which influences the suspension medium conductivity $\sigma_e$) and the internal cytoplasmic conductivity ($\sigma_i$) are considered to be influenced by the measuring frequency as described in the equation (2.6). At low frequencies, the internal conductivity value is small, and this non-conducting cell suspension is present in a conducting medium. Thus, the conductivity decrement observed relative to the suspending medium ($\sigma_L/\sigma_e$) is considered to be roughly proportional to the volume fraction of cells (equations (2.7)) (Davey et al., 1992). As described previously, the concept of $\alpha$ and $\beta$- dispersion in biological cells occurs due to the tangential flow of ions and accumulation of the ions at the membrane surface, respectively. Thus, the equations

related to the Cole-Cole model represent these phenomena related to the conductivity changes and ion movements, as explained in the recent review article by Nasir and Al Ahmad (2020). Furthermore, it could be noted that the effect of media compositions on the impedance measurements of fed-batch *E.coli* cultivation was found to be negligible in electrochemical impedance spectroscopy (EIS) as observed in a recent study (Slouka et al., 2016) (EIS and DS are both based on cellular dielectric properties).

The physiological properties, viz., cell radius and cell number density, of the respective organism are estimated using the nonlinear equations (2.5) to (2.7), where the cells are assumed to have a spherical shape (equation (2.2)). DS properties of *E. coli* suspensions were examined to consider the heterogeneity of the cell using the ellipsoidal shell model proposed in the literature, suggesting that the assumption of spherical shape has the scope of improvement and modification (Asami et al., 1980).

The various modeling methods for estimating biomass concentration from the capacitance measurements were described by Dabros et al., (2009b) and was tested in two yeast cultures, *Saccharomyces cerevisiae and Kluyveromyces marxianus.* The Cole-Cole model had more prediction errors than the linear and multivariate models in high noise conditions in their study. Further, the validation results with the model organisms highlighted the potential of DS based monitoring for optimization and control studies. The application of DS using the frequency scanning method was investigated for mammalian cells by Cannizzaro et al., (2003). The application of DS for viable cell volume (VCV) prediction in mammalian cells was explored by Downey et al., (2014). The frequency scan was performed, and the resultant permittivity values and shape of the scan was quantified using a novel area ratio method.

In contrast to the generally validated linear correlation, one research group stated a nonlinear biomass-capacitance correlation associated with the lipid accumulation in *Arxula adeninivorans*. They observed clear discrimination of the capacitance measurements between the different phases using the ratio of the $\beta$ dispersion and the characteristic frequency (Maskow et al., 2008a, 2008b). Various applications of DS for process monitoring are summarized in Table 2.1.

Table 2.1. Application of dielectric spectroscopy for process monitoring and control

| **Organism and application** | **Capacitance probe** | **Parameters measured/ estimated** | **Operating specifications** | **Cell density** | **Reference** |
|---|---|---|---|---|---|
| Mammalian cell culture (CHO) | BM 210, Aber instruments | $\Delta C$ | 0.6 MHz | 7E+6 cells $mL^{-1}$ | (Ducommun et al., 2001) |
| Mammalian cell culture (CHO) | BM 214M, Aber instruments | $\Delta C$ | Frequency scan (0.2-10 MHz); Cole-Cole, PLS | 1E+7 cells $mL^{-1}$ | (Cannizzaro et al., 2003) |
| Bacteria (*E. coli*), Recombinant protein | BM 214M, Aber instruments | $\Delta C$ | ~ | 40 g $L^{-1}$ | (Clementschitsch et al., 2005) |
| Yeast (*A. adeninivorans*), Lipid accumulation | BM 210, Aber instruments | $\Delta C$, $f_c$, Cole-Cole $\alpha$ | Frequency scan (0.1-19.49 MHz); Cole-Cole with Levenberg Marquart | 20 g $L^{-1}$ | (Maskow et al., 2008a) |
| Bacteria (*E. coli*), Recombinant protein | BM 220, Aber instruments | $\Delta C$ | Dual frequency (15.65 MHz & 1.12 MHz) | 70 g $L^{-1}$ | (Kaiser et al., 2008) |
| Yeast (*S. cerevisiae* and *K. marxianus*) | BM 210, Aber instruments | $\Delta C$, $f_c$, Cole-Cole $\alpha$ | Frequency scan (0.1-20 MHz), Dual | 12 g $L^{-1}$ | (Dabros et al., 2009b) |

| | | | | | |
|---|---|---|---|---|---|
| | | | frequency (15.56 MHz & 370 kHz) | | |
| Bacteria (*E. coli*), Recombinant protein | BM 210, Aber Instruments | $\Delta C$ | Dual frequency (15.649 MHz & 1 MHz) | 85 g $L^{-1}$ | (Knabben, 2011) |
| Mammalian cell culture (NS0 cell line), Recombinant protein | iBiomass 465 monitor, Fogale Nanotech | $\Delta\varepsilon$ | Frequency scan | 2.5E+10 $\mu m^3$ $mL^{-1}$ (VCV) | (Downey et al., 2014) |
| Bacteria (*E. coli*), Recombinant protein | BM, Aber Instruments | $\Delta C$ | Dual frequency (10 MHz & 1 MHz) | 40 g $L^{-1}$ | (Ehgartner et al., 2015) |
| Mammalian cell culture (CHO) | BM, Aber Instruments | $\Delta C$ | Dual frequency (0.58 MHz & 10MHz) | 16E+6 cells $mL^{-1}$ | (Cole et al., 2015) |
| Bacteria (*E. coli, Bacillus megaterium*), and Yeast *(Pichia pastoris, S. cerevisiae*), Recombinant protein | Biomass system, Fogale Nanotech | $\Delta\varepsilon$ | Dual frequency (10 MHz & 2.984 MHz-bacteria; 10 MHz & 2.076 MHz-yeast) | 2-7 g $L^{-1}$ | (Horta et al., 2015) |
| Fungi (*Penicillium chrysogenum*), Antibiotic | BM 220, Aber instruments | $\Delta C$ | Dual frequency (15.65 MHz & 0.1 MHz) | 60 g $L^{-1}$ | (Ehgartner et al., 2017) |

| | | | | | |
|---|---|---|---|---|---|
| Mammalian cell culture (CHO), Monoclonal antibody | FUTURA, Aber instruments | $\Delta C$, $f_c$, Cole-Cole $\alpha$ | Frequency scan (0.5 - 20 MHz.) | 10E+6 cells $mL^{-1}$ | (Ma et al., 2019) |
| Bacteria (*E. coli*), Recombinant protein | FUTURA, Aber instruments | $\Delta C$ | Frequency scan (0.1– 15 MHz) | 35 g $L^{-1}$ | (Randek and Mandenius, 2020). |

$\Delta C$ – capacitance increment; $\Delta\varepsilon$ – dielectric increment; $f_c$ - Characteristic or critical frequency; Cole-Cole $\alpha$ - Cole-Cole parameter

### 2.2.3 Control strategies using DS measurements

Suitable choice of PAT tool along with appropriate data pre-processing methods will enable efficient monitoring of the desired process variable (Marison et al., 2013). Reliable capacitance monitoring and biomass estimation using appropriate models can eventually lead to relevant control strategy implementation based on this critical process variable. Several researchers have implemented the application of this estimated biomass concentration and specific growth rate as a process variable to develop suitable feeding strategies for improved process performance.

Knabben (2011) investigated the application of impedance signal to detect high cell density fermentation of bacterial culture up to 85 g $L^{-1}$ concentrations, which was the highest ever detected biomass concentration using an impedance signal. A fed-batch mode of operation was employed using an exponential feeding determined by an open-loop control strategy. Another exponential feed-based fed-batch cultivation was carried out for four different organisms by Horta et al., (2015). The capacitance sensor monitored the cultivations, and a linear correlation between biomass concentration and permittivity values were established. Kaiser et al., (2008) applied dual-frequency mode to measure viable cell density in recombinant protein processes with a high cell density cultivation of *E. coli*. They observed that the decline in the capacitance signal corroborated with heterologous protein formation, and therefore concluded that, the productivity could be maximized by viable cell density control.

The potential application of DS in monitoring the cell culture for implementing process control in biopharmaceutical production was reviewed by Justice et al., (2011). Cole et al., (2015) explored the combined application of the PAT tools, DS, and biocalorimetry to monitor immobilized Chinese Hamster ovary (CHO) cells. Reliable calibration models for correlating the capacitance and heat flow measurements with viable cell density were obtained in this study, which demonstrated the application of complementary PAT tools for enhanced process monitoring.

The combination of DS measurements with soft sensors could yield additional physiological information regarding the respective organism. Ehgartner et al., (2015) defined a new terminology called 'bio-density' (biomass/biovolume) to describe the volumetric mass estimated by combining DS measurements and soft sensor based elemental balances. This bio-density was found to be a successful indicator of real-time

physiological changes, and additionally, a correlation was observed between this bio-density and specific product titer. Thus, it could be concluded that the application of DS with other PAT tools could enable the possibility of enhanced monitoring of bioprocess applications.

Some of the studies implementing control strategies using different PAT tools have been summarized in Table 2.2. It could be observed that, DS has been applied as a PAT tool for biomass estimation and for developing specific growth rate based control. However, there is still scope in developing advanced control strategies based on capacitance data and real-time application of the developed control strategy (Flores-Cosío et al., 2020).

Table 2.2. Application of PAT tools for control strategy implementation

| Organism, product | PAT tools | Control strategy | Observation | Reference |
|---|---|---|---|---|
| *K. marxianus, Candida utilis and P. pastoris.* | FTIR, DS and Exhaust gas analyzer | PI (modified), Exponential feeding (FF-FB) | Combination of online measurement with data reconciliation enables accurate estimation of specific growth rate and implementation of a simple robust control strategy | (Dabros et al., 2010) |
| *P. chrysogenum*, Penicillin | DS | PID, Constant, linear and exponential feeding | Predicted viable biomass concentration for two phases (growth and decline) were applied for control strategy development, which adapts to changing biomass yields | (Ehgartner et al., 2017) |
| *S. cerevisiae and E. coli* | DS, turbidity and online FTIR | PI (modified), Exponential feeding(FF-FB) | Critical specific growth rate was determined to run the process at maximum growth kinetics while preventing over flow metabolism | (Habegger et al., 2018) |
| *P. pastoris*, Human interferon α2b | DS, Exhaust gas analyzer | PID, Exponential feeding | Feeding strategy balanced the methanol consumption rate and resulted in enhanced titer and specific productivity | (Katla et al., 2019) |
| *E. coli*, Lethal Toxin-Neutralizing Factor | Flow injection analyzer | PID and DIOLC | Advanced controller scheme resulted in improvement in process performance, improved the purity and concentration of LTNF was achieved | (Priyanka et al., 2019) |
| CHO cells | DS, flowcytometry | Feed-back control | Potential of multi frequency spectra to monitor real time response of physiological changes induced by nutrient limitation was explored | (Ma et al., 2019) |
| *Streptococcus zooepidemicus*, hyaluronic acid | Calorimetery, Exhaust gas analyzer | PID (FF-FB) | Robust control of specific growth rate derived from metabolic heat rate resulted in improved molecular weight of the hyaluronic acid | (Mohan et al., 2022) |

FTIR - Fourier-transform infrared spectroscopy; DS - Dielectric spectroscopy; PI - Proportional - Integral; FF – Feed forward; FB – Feedback; PID - Proportional-Integral-Derivative; DIOLC-decoupled input-output linearizing controller;

### 2.2.4 Gaps and challenges

Availability of reliable process measurements with the help of advanced PAT tools could facilitate the real-time monitoring of the CPPs. Therefore, developing a suitable and cost-effective combination of the PAT tools for process monitoring is essential. As discussed in the previous sections, DS can be applied to achieve enhanced monitoring of biomass concentration, which is a key physiological variable. However, there are a few setbacks concerning the same, and they could be highlighted as follows.

- Implementing suitable data filtering techniques to extract valuable information from the capacitance sensor data is the first challenge. Pre-processing is significant, especially with respect to filtering the data from the frequency scans of the capacitance measurements since precise datasets devoid of signal noise are necessary to interpret the physiological changes occurring during the fermentation. Additionally, since the DS data are used to estimate biomass concentration using different correlation techniques and models, reliable data acquirement and appropriate filtering techniques are crucial.
- Published literature emphasized the possibility of estimation of physiological properties using scanning capacitance data. Since the cultivation of microorganisms for the production of recombinant proteins has significant decision steps like induction and harvest, the estimation of physiological properties can provide insight into the changes happening throughout the fermentation, and the operator can take suitable process decisions based on the same. Further advancement would be achieving this in real-time, which could further be applied for online model-based control strategies. Thus, there is a need for real-time estimation of physiological properties using measured capacitance data for enhanced monitoring and control applications.
- The real-time estimation of physiological properties can be achieved by developing a Cole-Cole model to relate the measured scanning capacitance and conductivity data to the respective properties. However, as discussed in the previous sections, there are a few setbacks regarding the implementation of the Cole-Cole model in a bacterial system for recombinant therapeutic production. Due to the small size of the cells and high signal noise, the challenges associated with Cole-Cole model

development for real-time estimation of physiological properties in a bacterial system remains to be addressed and explored.

- Real-time capacitance data can fingerprint the biomass growth profile accurately and instantaneously during the fermentation. Hence, integration of the measured capacitance data and the estimated biomass concentration in a validated process model would be beneficial for implementing online optimization and control strategies, and there is a scope to explore this notion.

## 2.3 Model development in bioprocesses

Bioprocess measurement and monitoring involve two main approaches: applying reliable and accurate real-time PAT tools and developing mathematical models that can relate the measured process variables from the sensors to the required quality attributes (Vojinović et al., 2006). The success of a PAT tool depends on the accuracy of the process model that could interpret the measured variables. An essential aspect of PAT implementation is establishing a meaningful relationship between the process conditions and product quality. A meaningful relationship can be achieved with the help of the combination of sensor data from the PAT tools employed in the reactor with a process model to obtain holistic knowledge about the various process variables. Successful process monitoring and control in the bioprocess depends on the measurement and monitoring techniques (Sonnleitner, 2012).

Model development is crucial since it is the backbone for developing and implementing optimization and control strategies for the biotherapeutic production process. Therefore, developing and validating appropriate models that could provide a reliable description of the underlying process is a significant step towards achieving desired objectives. Models can establish a mathematical relationship between the monitored CPP and CQA, which is a significant step before implementing a control strategy. A thorough understanding of the process is necessary to improve the process consistency in biopharmaceutical manufacturing, as emphasized by the recent initiative by the FDA and other regulatory bodies. Hence, mathematical model development would be the primary step (Wechselberger et al., 2010).

A mathematical model can improve the understanding of the living system through means of equations and serves as a mode of communication for knowledge sharing in both industrial and research settings and further aids in expanding the operational horizons

with the help of simulation (Mears et al., 2017a; Proß and Bachmann, 2012). A well-developed representative process model can enable performance estimation and prediction, scheduling and optimization (Montague et al., 2002). The combination of advanced sensor systems and an appropriate mathematical model could enhance the understanding of the underlying process, enable better prediction of the critical quality attribute and, subsequently, enhance the product quality by facilitating a better control strategy implementation (Sommeregger et al., 2017).

### 2.3.1 Modeling in upstream processes

Model development in the bioprocess perspective is the method by which mathematical representation of the process is developed and validated. Models can enable a convenient understanding of the relationship between different process variables even without performing actual time and resource-consuming experiments. Additionally, models could also provide insights on state variables that cannot be measured directly and provide output variables that are estimated infrequently. Another significant reason behind model development is its subsequent usage in optimization and control studies wherein the controlled variables (CV), and manipulated variables (MV) can be used to establish a suitable model-based control strategy (Luo et al., 2021).

#### 2.3.1.1 Different modeling methods

In bio-manufacturing processes, modeling techniques can be applied for different scenarios like process design, scheduling, economic analysis, process improvement, and many more (Chhatre, 2012). Among these, the process models describing fermentation fall into two broad categories: Mechanistic (white-box/ first-principles) and data-driven (black-box/ empirical). Another type of model, known as semi-parametric or hybrid modeling (grey-box), combines fundamental knowledge-based models and data-driven techniques.

Mechanistic models are mathematical models developed using the detailed description of the underlying process using first principle mechanisms. They describe the physical relationship of the input variables of the fermentation with the state and output variables in the form of an ordinary differential (ODEs) or partial differential equations. Mechanistic models can enable monitoring of critical process parameters that are not known in real-time and could further be used along with the measured experimental data to perform online parameter estimation, thereby improving the model prediction.

Mechanistic models can provide great benefit by attempting to approximate the process conditions to optimize the performance of a fermentation process (Ignova et al., 1998). The mechanistic models are more predictive and dynamic than the empirical data-driven models, as they can describe the non-linear process efficiently, and hence these models are more suitable for applications in model-based monitoring and control strategy testing (Mears et al., 2017a).

A simple example of a mechanistic process model for a fed-batch fermentation will be based on volume balance concerning feed addition ($F$). Empirical equations describing the relationship between biomass, substrate and product concentration are developed to represent the kinetics of the process. The generalized equations describing the change in biomass ($X$), substrate ($S$) and product ($P$) concentrations with respect to time of fermentation is presented in equation block (2.8). The typical kinetic model for representing growth rate ($\mu$) in microbial fermentation can be described using Monod's model, which emphasized the significance of dependence of specific growth rate on the growth-limiting substrate concentration as shown in (2.9) (Monod, 1949). The specific growth rate ($\mu$) is a significant process variable since it reflects the physiological characteristic of the organism and is related to the specific product titer. Within the substrate-dependent kinetic model, different substrate limiting and inhibiting models are available. For aerobic microbial cultivations, oxygen concentration is a substantially important variable affecting cell metabolism. Therefore, oxygen concentration could be considered an additional limiting substrate, and suitable kinetic models could be implemented (Kornaros and Lyberatos, 1997).

$$\begin{aligned}
&\frac{dV}{dt} = F \\
&\frac{dX}{dt} = \mu X - \frac{F}{V} X \\
&\frac{dP}{dt} = q_P X - \frac{F}{V} P \\
&\frac{dS}{dt} = \frac{F}{V}\left(S_0 - S\right) - \left(\frac{\mu}{Y_{XS}} + m\right) X
\end{aligned} \tag{2.8}$$

$$\mu = \frac{\mu_{\max} S}{K_s + S} \tag{2.9}$$

In the case of recombinant protein production processes, the development of a mechanistic model to depict the physiological state of the process in therapeutic protein application has been attempted previously. Ko and Wang, (2007) attempted model-based prediction of biomass and intracellular protein concentration in *E. coli* cultures using a dynamic mass balance model for the fed-batch cultivation. Ashoori et al., (2009) developed a process model with Contois kinetics and implemented a nonlinear control strategy with the validated model for penicillin fermentation. The combination of real-time biomass estimation using a soft-sensor and a mechanistic model based on elemental mass balances was demonstrated in different model systems: recombinant *P. pastoris and E. coli* by Wechselberger et al., (2013). Mechanistic model development is unique for every process, and therefore, it has to be explored and updated for implications in different applications of therapeutic protein production.

Even though mechanistic models can facilitate the process understanding and could be extended for optimization and control strategy implementation, they require more time and resources for their development. The development of mechanistic models is also difficult in cases where the complete dynamics are not understood clearly. In addition to that, the complex, dynamic and nonlinear nature of the bioprocesses adds more challenge for the same. Data-driven models, on the other hand, are entirely based on empirical observations of the actual process. These models do not require any biological understanding of the process, and they could be built with the help of plant data. Data-driven models are advantageous when the complex dynamics and fundamental mechanisms are not clearly understood. Established statistical models, including principle component analysis (PCA), partial least square regression (PLS) and artificial neural network (ANN), come under the category of data-driven models.

Hybrid modeling includes both first principle-based mechanistic equations combined with data-driven methods. Considering the merits like convenience and adaptability derived from both modeling methods, hybrid models can counteract the shortcomings of each other. Additionally, since hybrid models are centered around process understanding and provide the possibility of training them with limited data, they can serve as a digital twin in biopharma (Sokolov et al., 2021). The application of hybrid models is emerging for biopharmaceutical applications in the recent era and is expected to rise sooner. von Stosch et al., (2016a) developed a hybrid model combining the general parametric bioreactor model with a nonparametric ANN and correlated biomass and product

formation rates with the process parameters. The merits, challenges and selected examples of different modeling methods in recent literature (since 2013) are tabulated in Table 2.3. The selection of modeling methods is dependent on the particular application, data availability, ultimate goal for the model and user experience (Gnoth et al., 2008).

Several platforms are available for model development, simulation and optimization studies, like MATLAB, gPROMS, COMSOL, Aspen, to name a few (Chen et al., 2020). Modelica (Modelica Association) is an object-oriented open source modeling language used for modeling, simulation and programming of physical and technical systems and processes, for which there are multiple software tools available, both commercial as well as open source. The algebraic differential equations could be declared directly in Modelica without converting the model to ordinary differential equations (ODE) (Casella et al., 2011).

Table 2.3. Merits and challenges of different modeling approaches

| Model type | Merits | Challenges | Examples |
|---|---|---|---|
| Mechanistic | ▪ Provides a better understanding of the effect of process changes on CPP and CQA<br>▪ Adequate description of the nonlinear process behaviors<br>▪ Facilitates implementation of advanced predictive control strategies | ▪ Difficulties in the description of the entire complex biological system<br>▪ A thorough understanding of the underlying process is required<br>▪ Longer development time | (He et al., 2019; Lopes et al., 2014; Wechselberger et al., 2013) |
| Data-driven | ▪ Empirical observations of the actual process are sufficient for the model<br>▪ Data from the plant are directly used for model building<br>▪ Preferred during the design of experiments and control system design | ▪ Underlying mechanisms of the process cannot be understood completely<br>▪ Extensive calibration of parameters is required<br>▪ Availability of large sets of data is required to develop the model | (Goldfeld et al., 2014; Sokolov et al., 2018; Zürcher et al., 2020) |
| Hybrid | ▪ Benefits from both modeling methods can be combined. | ▪ Requires training data to train the data-driven | (Narayanan et al., 2019; von Stosch et al., |

| | | | |
|---|---|---|---|
| | ▪ Highly suitable for nonlinear processes and in cases where the physical process is well understood and metabolic rates are poorly understood | approach of the hybrid model | 2016a; Zhang et al., 2019) |

#### 2.3.1.2 Data pre-processing

Data pre-processing is one of the most meaningful steps for efficient interpretation of the measured process variables and establishing data homogeneity. All the real-time monitored data from the PAT tools are disturbed by instrumental or process noise and thus have to be properly filtered to acquire reliable information from the raw data. This step might be an insignificant procedure, but efficient pre-processing technique application is necessary to ensure that no valuable information is lost during the process (Marison et al., 2013). Some of the commonly used pre-processing techniques include moving average (MA) filter applied for pre-processing of DS data (Horta et al., 2012). Multivariate data analysis (MVDA) is also a part of the data-preprocessing technique, which is applied to extract the information using data-driven methods (Biechele et al., 2015).

#### 2.3.1.3 Sensitivity analysis and model validation

Sensitivity analysis is carried out to understand the influence of model parameters on the defined process variables and assess the validity of the developed process model. The simplest method for sensitivity analysis is a local sensitivity analysis ('one-factor-at-a-time'). In this method, one parameter is perturbed to a small extent by keeping other parameters at their nominal value, and the resultant changes in the process variables are observed (Qian and Mahdi, 2020; Wang et al., 2012). Thus, the sensitivity analysis provides insights into the developed process model and thereby facilitates the development of a reliable process model. Validation of the process model is necessary to ensure the successful development of a process model for its future implications in optimization and control strategy development. A recent review by Rajamanickam *et al.*, (2021) highlighted the significance of model validation in bioprocesses.

### 2.3.2 Soft sensors

Some of the challenges in the model development include difficulty in applying a nonlinear model that could exactly represent the underlying process and lack of reliable sensors to measure all the process variables (Azimzadeh et al., 1999; Feng et al., 1999). Under the circumstances when all process variables cannot be directly measured, soft-sensors serves as a pertinent solution (Ko and Wang, 2007; Luttmann et al., 2012; Randek and Mandenius, 2018). Soft sensors refer to the combination of a 'hard sensor' with software implemented models which uses the data derived from the hard sensors to estimate the variables of interest (Chéruy, 1997). The application of soft sensors could reduce the need for offline analysis and support real-time process monitoring by providing real-time information of the biological variables, thereby contributing to bioprocess automation (Sonnleitner, 2012).

Soft sensors are also employed for the estimation of parameters like oxygen uptake/transfer rate (OUR/OTR), carbon dioxide production/evolution rate (CPR/CER), which are calculated from aeration rate and exhaust gas measurements (infrared and paramagnetic sensors for $CO_2$ and $O_2$). Randek and Mandenius, (2018) extensively discussed various strategies for applying soft sensors in upstream bioprocessing. Soft sensors combine the measured process variables and measurement models (data-driven or mechanistic) to estimate the variables of interest. Soft sensors enable the possibility of real-time estimation of the process variables for online optimization and process control strategy implementation, thereby contributing to bioprocess automation (Sonnleitner, 2012).

Another important class of soft sensors is the state estimator, which uses a dynamic theoretical model to obtain the process variables (state variables). Application of state observers in a model-based estimation was carried out by Kawohl et al., (2007). The application of the Kalman filter for estimating the specific growth rate using OUR, CPR and base consumption was explored by Jenzsch et al., (2006d). Extension of a state observer with a nonlinear state model results in extended Kalman filters (EKF) based on error covariance minimization (Kalman, 1960). The Kalman-Filter can handle the time-varying characteristics and uncertainties within the dynamic model and the measurement data. Though EKF copes with the nonlinearity, some challenges like model accuracy and linearization problems exist in this method (Simutis and Lübbert, 2015).

### 2.3.3 Gaps and challenges

Developing a suitable process model has multiple applications in exploring simulation, optimization and control studies, as discussed in this section. A well-defined process model can enable enhanced interpretation of the measured process variable and pave the way for providing additional insights about the behavior of the process under investigation. Some of the challenges for developing a mechanistic model in a therapeutic production process can be summarized as follows.

- The major challenge concerning model development is the nonlinear dynamics and complex nature of the biopharmaceutical application processes. Even though employing data-driven and hybrid models could be advantageous in some aspects, a simple mechanistic model can provide insights into the physiology of the process and can be validated with minimum datasets. Hence, modifications in the mechanistic model to accommodate the process under investigation are essential to accomplish reliable model development.
- The kinetic model representing cell growth and substrate utilization is generally validated for the batch operation and then translated to the fed-batch mode. However, the volume change in fed-batch fermentation is critical in influencing the process dynamics. Developing a process model for the fed-batch fermentation of the therapeutic protein production process is challenging due to the volume changes occurring during the feeding process.
- For the development of substrate-limiting kinetic models for aerobic microbial processes, oxygen dependency of the organism is often neglected. Additionally, the changes related to input variables concerning oxygen concentration in the reactor, such as airflow rate and oxygen added in the reactor, and the corresponding changes reflected in the volumetric mass transfer coefficient ($k_La$), might be significant for achieving a reliable representation of the process dynamics. Therefore, considering the importance of oxygen-related parameters, the reactor's residual oxygen can be considered an additional limiting substrate to represent the microbial growth equation.

## 2.4 Optimization studies and control strategy development for overall process optimization

The initial stages of process development include identifying process parameters, real-time monitoring of the identified CPPs, and establishing a correlation between process parameters and quality attributes using mathematical models, as noted in the preceding sections. The next step in the sequence is developing suitable control strategies to reduce process variations and maintain product quality. One of the primary and significant requirements for establishing any control strategy is obtaining reliable measurements of the process variable (PV), based on which the control strategy is designed. The measured PV is compared with the desired set points, and the necessary control action is determined by manipulating the process inputs. Developing an integrated control strategy that can manage the process variations is a significant challenge to be surpassed for establishing continuous manufacturing for bioprocesses. The dependencies between the CPPs and CQAs have to be transformed to the integrated process model and control strategy in place for efficient adaptation of the conversion.

A descriptive model provides an understanding of the intrinsic phenomenon of the process and could predict the dynamic profiles of the state variables and determine the manipulated variables (MV) for control application, where specific product yield or maximizing the biomass concentration could be defined as an objective function (Freitas et al., 2017). In brief, the application of advanced monitoring tools and a well-designed process model could significantly enhance consistency in process and product quality through appropriate real-time control actions (Konstantinov and Cooney, 2015; Rathore, 2016, 2014). A schematic overview representing the interactions between process monitoring, modeling and advanced control strategy implementation is presented in Figure 2.3.

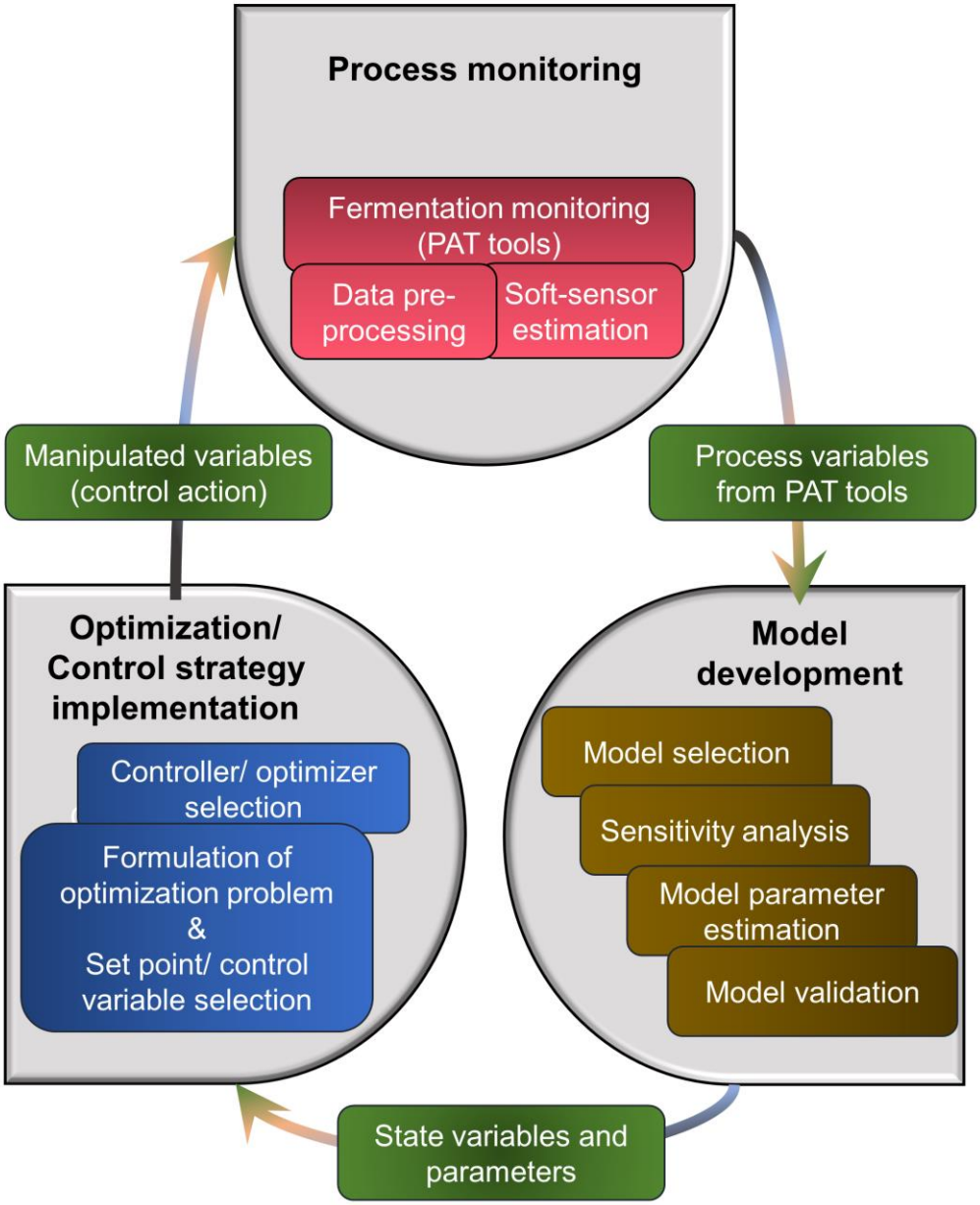



Figure 2.3. Schematic representation of the association of process monitoring, modeling and control in bioprocesses

### 2.4.1 Optimization studies for different applications

Optimization studies in the context of bioprocesses can be implemented in different stages throughout the process. Some examples include design space optimization using

traditional design of experiments (DoE), culture media optimization using response surface methodology (RSM), and model parameter estimation to optimize the process according to desired objectives (model-based optimization). Since model-based optimization studies are built up over the validated process model, the success of the method relies on the reliable representation of the system (von Stosch et al., 2016b).

The first step for applying any model-based optimization objective would be to develop a process model to estimate the parameters using direct experiments and verify the parameterization using sensitivity analysis and other additional experiments. This step is followed by optimization of the underlying process once the process model is validated (Proß and Bachmann, 2012). A general optimization workflow involves formulating an objective function J, followed by defining the constraints and bounds for the decision variables, which could be reflected as manipulating variables to achieve the desired objectives.

Among the different optimization objectives, the two major categories could be cost-based objective or productivity-based objective. In the latter category, some of the possible objectives for the dynamic optimization in fed-batch cultivation could be related to productivity, yield, end-point substrate concentration, biomass, fed-batch time, and many more. Optimizing the feed rate of limiting substrate concentration to maximize the yield or productivity of the fed-batch process using simple mathematical techniques has been in practice since the late 20$^{th}$ century (Menawat et al., 1987; Modak and Lim, 1989; Yamane et al., 1977). However, with the arrival of extensive computational power and growing technological advancements with online monitoring tools, the application of optimization with integrated models for recombinant processes has reached a whole other level (Noll and Henkel, 2020).

Product concentration and productivity of a biologic are significant for commercialization, and thus optimization of the total biomass in a reactor plays a significant role (Kaiser et al., 2008; Rathore, 2016). Several researchers have attempted to apply various optimization objectives, including biomass maximization and undesired product minimization, wherein different manipulated variables (MV) like substrate feed rate and air feed rate have been implemented (Atasoy et al., 2013; Geethalakshmi et al., 2011; Ko and Wang, 2006; Ochoa, 2016; Zafira and Nandong, 2019).

In case of multiple objectives that might also be conflicting, multiobjective optimization (MOO) is desirable to handle the same. MOO can handle the multiple objectives simultaneously, and the solution can be obtained in the form of Pareto optimal points (Sarkar and Modak, 2005). The dynamic optimization problems could be solved using various stochastic algorithms, such as differential evolution (Kapadi and Gudi, 2004), genetic algorithm (Sarkar and Modak, 2005), and many more. Patel and Padhiyar, (2017) demonstrated the implementation of MOO with different case studies and solved the optimization algorithm through a multiobjective differential evolution. Integration of process-model, with knowledge associated with measured cell-metabolism and CQAs, and undergoing a parameter estimation and optimization strategy for achieving a suggested feeding strategy have greatly enhanced the process performance in a recent study in CHO cells (Kotidis et al., 2019). Therefore, implementing an optimization framework for therapeutic protein production processes is expected to support the QbD principles for pharmaceutical manufacturing processes. A summary depicting samples of objective function formulation in optimization studies carried out for different fermentation applications is presented in Table 2.4.

Table 2.4. Summary of sample optimization studies carried out in fermentation applications

| **Objective function** | **Application** | **Optimization results** | **Reference** |
|---|---|---|---|
| $\max_{F(t)}\left[P\left(t_f\right)V\left(t_f\right)\right]$<br>Maximizing productivity using Smooth control profile parametrization (sinusoidal function) | *S. cerevisiae* Ethanol fermentation | Abrupt changes in the control profile were avoided, which could reduce the substrate shock in the organisms | (Ochoa, 2016) |
| $\max_F J = X\left(t_f\right) - \int_0^{t_f} C_E$<br>Multiobjective optimization to maximize biomass and minimize unwanted alcohol formation.<br>MV: Substrate and air feed rate | *S. cerevisiae* Ethanol fermentation | Control vector parameterization was used for solving the dynamic optimization problem, and an effective MOO algorithm for online implementation was developed. | (Atasoy et al., 2013) |

2.4 Optimization studies and control strategy development for overall process optimization

| | | | |
|---|---|---|---|
| $\max_{F(t)}\left[P(t_f)V(t_f)\right]$<br>Maximizing recombinant protein production by determining optimal substrate feed profile using GA | *E. coli* Recombinant streptokinase | The average volumetric productivity of recombinant streptokinase was 8% higher for the obtained optimal feed trajectory | (Geethalakshmi et al., 2011) |
| $\max_{F(t),X_{2F},t_f} J=\frac{P(t_f)V(t_f)}{t_f}$<br>Maximizing recombinant protein production rate by determining optimal feed rate, feed glucose concentration and fermentation time. | *E. coli* Recombinant aspartase | The two-phase approach with the optimal feed rate could improve the recombinant protein productivity by 3% compared to the experimental results | (Ko and Wang, 2006) |

### 2.4.2 Conventional control strategies for fermentation

There are three modes of reactor operation in any bioprocess: batch, fed-batch, and continuous, as discussed previously. Fed-batch cultivation sustains the growth rate at maximum specific growth rate with the addition of limiting substrate in subsequent intervals and is the most preferred operating mode in most industrial fermentations. In fed-batch mode, the feeding rate of the limiting substrate is a significant manipulating variable (MV) which is often used in control applications. The feed rate could impact the system in several ways affecting various process parameters like growth rate, substrate concentration, biomass concentration, product formation rate, and viscosity which will eventually affect OUR and OTR (Dabros et al., 2010; Mears et al., 2017b). The application of various control strategies for a fed-batch culture was reviewed by Lee et al., (1999) in the earlier decade. The control strategies could be developed using different control goals like setpoint control or trajectory control. Since the productivity of the process and the product quality are majorly growth-related, even a small deviation in the growth rate can significantly affect productivity. Therefore, majority of the conventional control strategies are applied based on the objective of specific growth rate control ($\mu$). A schematic representation of all the discussed control strategies (conventional: open and closed loop, advanced: adaptive and MPC) is presented in Figure 2.4 (Hausmann et al., 2017; Rathore et al., 2021).

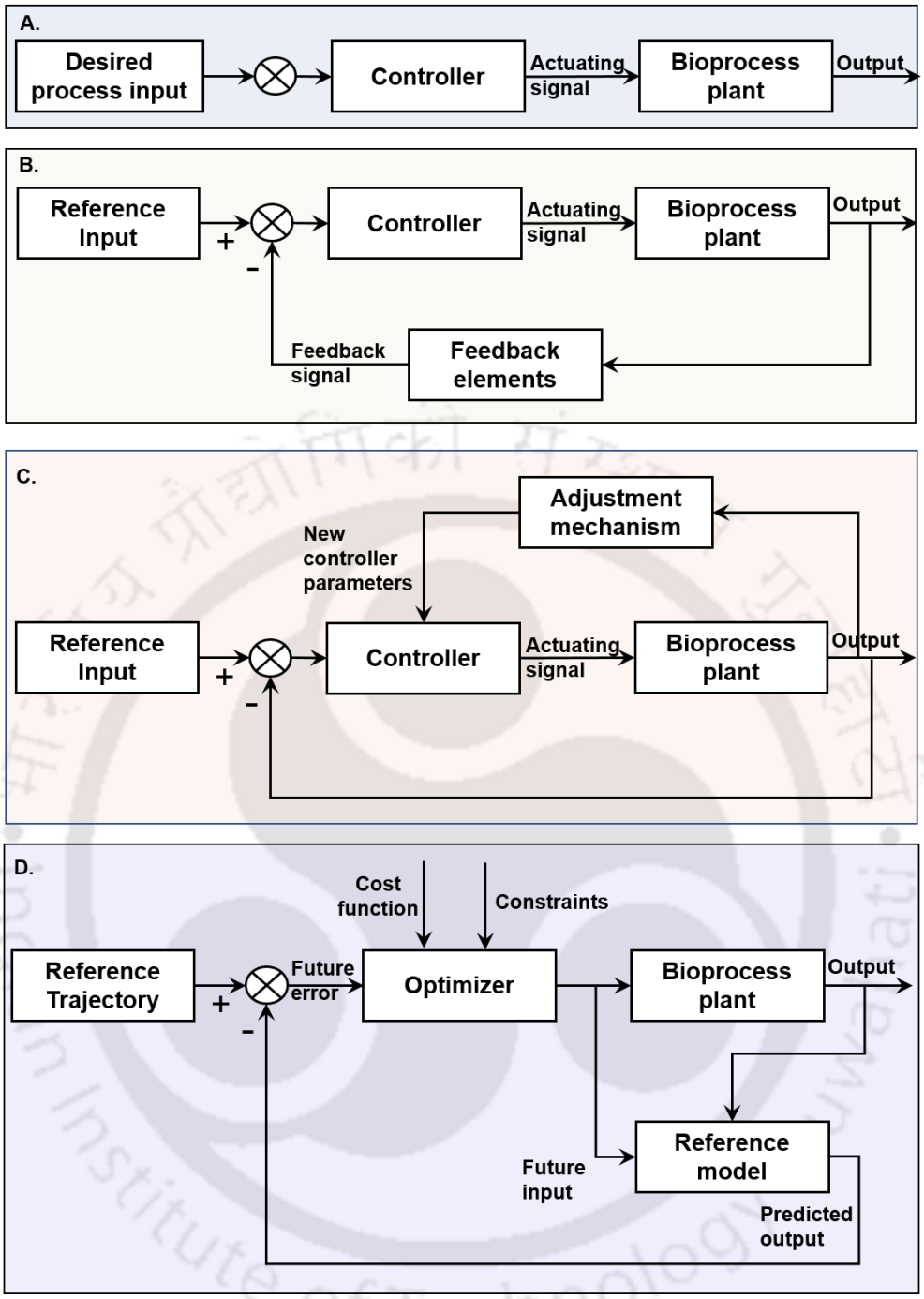



Figure 2.4. Block diagram of different control strategies discussed in this section. (A). Open-loop control, (B). Closed-loop control with feedback, (C). Adaptive control and (D). Model predictive control

### 2.4.2.1 Open-loop control

The open-loop control is the simplest method, as it does not require any online measurements. The control algorithm is set based on the desired set points, and the control law follows this predefined trajectory. The control actions are pre-computed by the

controller and are executed accordingly. This kind of open-loop could fail to predict the desired MV in case of any disturbance in the system since no reference measurement is considered by the controller (Vinet and Zhedanov, 2013). Some previous studies implemented the DS-based monitoring with an open-loop exponential feeding strategy in an *E. coli* culture (Gregory and Turner, 1993; Knabben, 2011; Turner et al., 1994). The advantages and disadvantages of open-loop control were discussed in the context of explaining the batch-to-batch variability, and the significance of reliable process monitoring was highlighted by the literature (Gnoth et al., 2007; Jenzsch et al., 2006a)

#### 2.4.2.2 Closed-loop control

In order to cope with the process disturbances and control the desired variable based on the existing process conditions, a simple closed-loop control strategy using Proportional Integral/ Proportional Integral Derivative (PI/PID) controllers could be applied. The obtained online measurements are directly used for correcting the deviations of the process variable from its desired path. Closed-loop control action (actuating variable) is the resulting action of the controller based on the applied control law. PID controllers perform the control action based on the error difference (*e(t)*) between the measured process variable and the setpoint. The PID controller parameters have to be tuned for the particular process conditions, and a simple set of rules for tuning the controller gains was described by Ziegler and Nichols, (1942). Other tuning techniques such as relay tuning and automatic tuning of PID controllers are also available (Åström and Hägglund, 2006).

For any typical closed-loop control, the measured process variable is a significant component based on which the model predicted states or the setpoints are compared to decide necessary control actions. As discussed in previous sections, DS can serve as a promising PAT tool for providing reliable biomass estimations from which the desired physiological variable $\mu$ could be obtained. The application of closed-loop control for filamentous fungi using the specific growth rate based on viable biomass estimation from DS was explored by Ehgartner et al., (2017), where the feed rate *F* was determined based on the PID output. The approach of using estimated biomass as the controlled variable (CV) instead of $\mu$ could reduce errors that could arise from the estimation procedure, as noted by several authors (Gnoth et al., 2008; Wechselberger et al., 2013). Similarly, Zitzmann et al., (2018) used the sensor signals from complementary PAT tools DS and NIR for specific growth rate prediction and coordinated the time of induction and harvest

in *Drosophila melanogaster*. Control strategies could also be developed based on the OUR and CER as measured variables to employ specific growth rate ($\mu$) control, and the manipulating variables such as stirrer speed, aeration rate, or pressure can be chosen accordingly (Levisauskas et al., 1996).

In general, the parameters like yield coefficients and consumption rates are assumed to be constant for closed-loop control. However, in the actual scenario, these parameters are dynamically changing due to the metabolic reactions in the cell, which makes it difficult to identify the controller parameters (Gnoth et al., 2007). This is considered as one of the setbacks while applying PID controllers for a biotechnological process where dynamic changes in PID gains remains a challenge to be addressed. Therefore, several alternatives for the classical PID controllers have been developed, including gain scheduling or dynamic parameter tuning (Dabros et al., 2010; Gnoth et al., 2008). Dabros et al., (2009a) proposed a relatively simple control strategy by applying an online data reconciliation methodology, based on which a PI strategy was proposed for specific growth rate ($\mu$) control. A sufficiently stable control was achieved using this strategy where the online measurements of biomass concentration verified by online balancing and data reconciliation could provide a reliable estimation of the specific growth rate leading to a robust control strategy. Due to the nonlinearities of bioprocesses and dynamic changes, the conventional control strategies could be insufficient for achieving effective controller performance.

### 2.4.3 Advanced process control strategies

In recent years, the PAT applications for biopharmaceutical production have increased significantly with emerging technologies that could enable advanced control strategy implementation (Glassey et al., 2011; Rathore et al., 2010). Model-based control strategies can tackle the shortcomings of the conventional control strategies, and therefore, an increasing drive towards the application of advanced control strategies for bioprocesses is observed (Mears et al., 2017a). Advanced control strategy aims to ensure process quality and robustness by establishing the real-time optimization of the process variables (McCready, 2017). In addition to the capability of addressing changing process dynamics, advanced control strategies can handle the interactions between various control loops in the process (Gomes et al., 2015). Some advanced control algorithms include

Adaptive control, Model predictive control (MPC), Artificial neural networks (ANN) and Fuzzy logic.

#### 2.4.3.1 Adaptive control

Adaptive control strategies are nonlinear controller algorithms where the controller parameters adapt to the changing nonlinear process dynamics and provide a better response. Gain scheduling could be considered as one of the adaptive control methodologies. (Mears et al., 2017b; Simutis and Lübbert, 2015). According to Jenzsch et al., (2006b), Artificial neural networks (ANNs) can estimate the biomass concentration ($X$) when several data records are available. A simple adaptive control law was developed where the biomass yield coefficients ($Y_{xs}$) were corrected using the change in biomass concentration ($\Delta X$) and an adaptive variable α. They observed a significant batch-to-batch reproducibility with this approach. The changes in the parameters such as specific growth rate ($\mu$), biomass yield coefficient from the substrate ($Y_{xs}$) and the maintenance coefficient ($m$) were considered to develop a control strategy by Horta et al., (2012) in high cell density cultivations (HCDC) of recombinant *E. coli*. The yield and maintenance coefficient values were re-tuned using at-line data, and the parameter estimation was carried out using the mass balance equations by minimizing an objective function. They compared four different cultivation strategies and observed that a cell density of 154 g $L^{-1}$ was achieved by implementing dynamic re-tuning ($\mu_{DYN}$, $Y_{xs}$ and $m$) compared to the 100 g $L^{-1}$ obtained with a constant $\mu_{set}$ without re-tuning. This approach could also be considered as a model-based adaptive control since the parameters are adapted based on the mass balances and the at-line data.

Jenzsch et al., (2006d) developed a generic model control where the process dynamics were described by a simple mechanistic model and a nonlinear state estimator (EKF), and it proved to be a noteworthy approach. This particular work is a Multiple input single output approach (MISO) where a nonlinear adaptive controller was designed based on OUR, CPR and base consumption to estimate the specific growth rate and to compute the correction in the feed rate $F$ accordingly. The adaptation to the process dynamics and reliable state estimation enables this generic model control to efficiently achieve the desired control goals compared with a simple PI control. Apart from $\mu$ based controls, fed-batch cultivation of an aerobic process can also be controlled using the partial pressure of oxygen ($pO_2$). One such model-based optimization and control algorithm was

developed by Galvanauskas et al., (2013) based on the cascade $pO_2$ control. An alternative approach for DO control using a nonlinear controller (decoupled geometric controller) was explored by Chopda et al., (2015), where they achieved a higher biomass concentration compared to a conventional PID control.

#### 2.4.3.2 Model predictive control

Model predictive control (MPC) has been implemented for bioprocess optimization owing to its prediction capability and facility to define desired objectives within the optimizer present inside the controller (Sommeregger et al., 2017). MPC aims to achieve minimum deviation of the selected process states from the respective setpoints by optimizing the manipulated process inputs and serves as an advanced process control strategy. Successful application of an MPC depends on the process model and online measurements of the state variables. Soft sensors can provide real-time state estimations of the control variables and correct the dynamic model predictions, which can ultimately enable the application of MPC and the prediction of future system behavior. The predictive capability of MPC can identify the problems beforehand and can also handle the interactions between multiple input and outputs (MIMO). For successful implementation of MPC within the QbD paradigm, availability of a dynamic model to predict CQA-CPP interactions and reliable state estimations by soft-sensors are some of the challenges to be met (Sommeregger et al., 2017).

The key difference between MPC and the other control strategies is that MPC has the freedom to set the objective function. It can be applied in two ways: setpoint trajectory for a specific variable or set up an objective function (minimize/maximize) (Mears et al., 2017b). Kuprijanov et al., (2013) developed an MPC strategy for controlling the biomass setpoint trajectory by using the substrate feed rate as the manipulating variable. The setpoint trajectory for the feed rate was taken from the experimentally derived 'golden batch', which was desirable. A similar setpoint trajectory-based nonlinear model predictive control (NMPC) was developed by Craven et al., (2014), where the substrate (glucose) concentration was maintained according to a setpoint trajectory. NMPC could be a promising control strategy to handle the nonlinear process dynamics if reliable process variables are available.

The alternative application of the MPC is to set up a maximization or minimization problem. The defined objective function can be used for productivity maximization

subjected to appropriate constraints when a setpoint for the controlled variable is unknown. Simulation studies of the application of NMPC for various maximization problems were carried out in previous studies (Chang et al., 2016; Santos et al., 2012). In a recent study, comparison and experimental verification of different control strategies were explored for the fed-batch process of *P. chrysogenum* by Kager et al., (2020). The nonlinear MPC resulted in efficient substrate utilization by avoiding by-product formation and resulted in a 14% higher gain in productivity. Despite the evident advantages offered by MPC, the requirement of a validated dynamic process model and expensive computational complexity remains a challenge for its implementation in biopharmaceutical applications.

Apart from MPC, other advanced controllers include Fuzzy logic and Artificial Neural Networks (ANN). Fuzzy logic uses linguistic expressions to handle uncertainties and does not need a mathematical model, while rule-based process knowledge of an expert operator is required. There are limited examples available in the literature for fuzzy control of feed rate (Mears et al., 2017b). A recent publication reviewed the history and evolution of fuzzy logic (Kahraman et al., 2016). Another advanced control strategy based on data-driven technique is Artificial Neural Network (ANN), which can describe a complex nonlinear system without explicit physical model equations. ANN models have to be trained with historical data in order to be implemented. Some of the discussed control strategies are summarized in Table 2.5.

Table 2.5. Application of various control strategies

| Type of control | Application | PAT tool applied/ Measured variables | Objective/ CV | Manipulated variable (MV) | Highlights | Reference |
|---|---|---|---|---|---|---|
| Open-loop | Bacteria (*E. coli*) | Off gas | - | *F* | $\mu$ set 0.5 $h^{-1}$; Dynamic Mass balance | (Jenzsch et al., 2006a) |
| | Bacteria (*E. coli*) | DS, off gas/ *X*, $pO_2/pCO_2$ | - | *F* | $\mu$ set 0.5 $h^{-1}$; Mass balance | (Knabben, 2011) |
| Closed-loop Feedback | Yeast (*K. marxianus*) | DS, off gas, FTIR/ *X*, $pO_2/pCO_2$, metabolite concentration | $\mu$ | *F* | Feedback with data reconciliation | (Dabros et al., 2010) |
| | Fungi (Filamentous) | DS, off gas, ion chromatography/ *X*, $pO_2/pCO_2$, metabolite concentration | $\mu$ | *F* | Viable biomass concentration prediction in growth and decline phase; $\mu$ set 0.012 $h^{-1}$ | (Ehgartner et al., 2017) |
| Adaptive | Bacteria (*E. coli*) | Off gas/ OUR, CER, pH | Maintain desired $\mu$ | *F* | EKF for state estimation (*X*, *S*) | (Jenzsch et al., 2006d) |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Bacteria (*E. coli*) | DS, HPLC/ *X*, metabolite concentration | $\mu$ ($Y_{xs}$ and $m$) | $F$ | $\mu_{set}$, $Y_{xs}$, and $m$ automatically re-tuned | (Horta et al., 2012) |
| | Bacteria (*E. coli*) | Off gas/ OUR, CER, base | $X_{tot}$ | $F$ | ANN based training with 26 batches | (Jch, Gnoth, Kleinschmidt, *et al.*, 2006) |
| | Bacteria (*E. coli*) & yeast (*S. cerevisiae*) | Off gas | Maximize $XV_{tot}$/ $pO_2$ | $F$ | Stirrer speed, oxygen enrichment and substrate feeding were used. | (Galvanauskas et al., 2013) |
| Model predictive control | Bacteria (*E. coli*) | Offline analysis/ *X*, *S* | Maintain *X*/ *X* | $F$ | Good batch-to-batch reproducibility was obtained | (Kuprijanov et al., 2013) |
| | Mammalian cells | Raman spectroscopy, off gas/ *S*, $pO_2/pCO_2$, | Maintain *S*/ *S* | $F$ | Glucose setpoint 11 mM, NMPC | (Craven et al., 2014) |
| | Bacteria (*E. coli*) | Off gas/ OTR, CER, OUR | Max glucose oxidation | $F$ | Simulation experiment | (Santos et al., 2012) |
| | Yeast (*S. cerevisiae)* | Offline analysis (*X*, *S*, *P*, *V*) | Max ethanol concentration *P* | *F, DO* setpoint | Simulation experiment | (Chang et al., 2016) |
| | Fungi (*P. chrysogenum*), | Off gas/ OUR, CER | Set-point control/ glucose uptake rate, product precursor and | F (glucose, nitrogen, precursor) | Comparison of PID, model-based and model predictive | (Kager et al., 2020) |

| | | |
|---|---|---|
| | nitrogen concentrations | control and experimental verification of the same |

*F* - feed rate (limiting substrate); $\mu$ - specific growth rate; DS-dielectric spectroscopy; *X* - Biomass concentration; $pO_2/pCO_2$ - partial pressure of oxygen/carbon dioxide in off gas; FTIR - Fourier transform infrared spectroscopy; OUR - Oxygen uptake rate; CER - carbon dioxide evolution rate; EKF - Extended Kalman filter; *S* - Substrate concentration; HPLC - High-performance liquid chromatography; $Y_{xs}$ - Yield coefficient biomass/substrate; *m* - maintenance coefficient; ANN - Artificial neural network; NMPC - Nonlinear model predictive control; OTR - Oxygen transfer rate; *P* - Product concentration; *V* - Culture volume; *DO* - Dissolved oxygen; PID - Proportional integral derivative

As elaborated in this section, advanced process control applications will face certain hurdles due to the complexity of the underlying biological process and variations in the process parameters due to the dynamic nature of the process. Applying a supervisory control layer to manage the underlying regulatory layer effectively can handle the problems/variations occurring due to the interaction of the multiple control loops. With the exploration of various control strategies, the future advancement will be applying an integrated process that integrates the individual process units and is driven by a global objective. Gomes et al., (2015) envisioned this perspective by proposing a supervisory master controller to predict process outcomes by coordinating the individual subunits and their respective PAT objectives. According to the authors, this kind of supervisory controller was proposed to be a possible advancement in the next decade to integrate the system analysis and control strategies. An overview of existing control strategies with an outlook towards the future of bioprocess automation and amalgamation of system architecture and various interfaces was provided by Rathore et al., (2021) in a recent review.

### 2.4.4 Gaps and challenges

Establishing a reliable optimization and control strategy for a biopharmaceutical application can evidently result in enhanced process performance and significantly reduce the batch-to-batch variability in the manufacturing process. Directing the process to follow the desired trajectory or operate within the specified set points will direct the manufacturing process towards achieving consistent product quality, which is beneficial in terms of biotherapeutic production. Some of the challenges in the implementation of advanced control strategies can be briefed below.

- Formulation of an optimization problem can be carried out to achieve various objectives depending on the control goals. In the case of intracellular recombinant protein production, the biomass concentration being a significant physiological variable that can influence the productivity and product quality, incorporating the same within the optimization problem can be beneficial. Therefore, developing a multiobjective optimization problem by including the biomass concentration and reactor volume for enhancing the productivity of the process under investigation would be an innovative strategy that could be explored.

- The decision regarding the harvest time for the fed-batch cultivation is often chosen intuitively considering the offline process measurements. However, a strategy for predicting the optimal harvest time would be helpful for the operator to take effective process decisions that could result in enhanced process performance.
- The application of advanced control strategies such as MPC can handle the dynamic changes and the interactions between the different input and output variables of the fermenter. However, the investigation of advanced control strategy implementation for the recombinant protein process is in the exploratory phase. Hence, the challenges that might occur for a particular application have to be addressed appropriately.
- Integration of real-time estimated process variables in a validated process model for the development of advanced control strategies would direct towards real-time optimization and bioprocess automation, and this has to be investigated.

## 2.5 Summary of the state of the art

For successful implementation of real-time optimization and advanced control strategies by obtaining reliable process measurements and a validated process model, several hurdles must be surpassed, some of which are summarized in this chapter based on the gaps present in the discussed literature. The summary of state of the art with the scope of the present work can be briefed below.

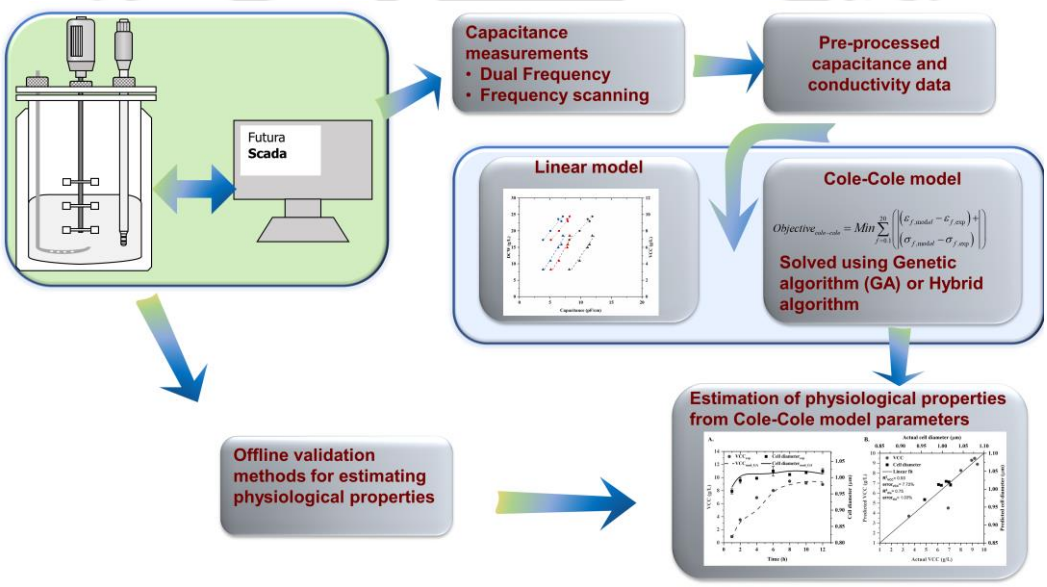
- For the past three decades, capacitance has been a successful technique for online biomass monitoring in different biotherapeutic applications. However, exploiting the tool to its complete proficiency is still in progress, especially concerning the real-time estimation of physiological properties.
- Developing a kinetic model for a particular process always has scope for improvement with respect to implementation in fed-batch cultivation, incorporating additional kinetic parameters and integrating estimated process variables from advanced PAT tools.

The application of optimization and advanced control strategies based on specific single or multiple objectives and solving them using different algorithms to provide anticipated profiles for the manipulated variables such as substrate feed rate can be explored for a biotherapeutic application.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

# CHAPTER 3

# Application of dielectric spectroscopy for real-time monitoring of biotherapeutic protein production




---

Parts of this chapter is based on the following research article(s):

**Swaminathan, N.,** Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2022. Cole-Cole modeling of real-time capacitance data for estimation of cell physiological properties in recombinant Escherichia coli cultivation. Biotechnol. Bioeng. 119, 922–935. https://doi.org/10.1002/bit.28028

**Swaminathan, N.,** Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2020. Multiobjective Optimization for Enhanced Production of Therapeutic Proteins in Escherichia coli: Application of Real-Time Dielectric Spectroscopy. Ind. Eng. Chem. Res. 59, 21841–21853. https://doi.org/10.1021/acs.iecr.0c04010

## 3.1 Foreword

Real-time estimation of physiological properties of the cell during recombinant protein production would ensure enhanced process monitoring. This study explored the application of dielectric spectroscopy to track the fed-batch phase of recombinant *Escherichia coli* cultivation for estimating the physiological properties, viz. cell diameter and viable cell concentration. The scanning capacitance data from the DS were pre-processed using moving average (MA). Later, it was modelled through a nonlinear theoretical Cole-Cole model and further solved using a global evolutionary genetic algorithm (GA). The parameters obtained from the GA were further applied for the estimation of the aforementioned physiological properties. The offline cell diameter and cell viability data were obtained from particle size analyzer and flow cytometry measurements to validate the Cole-Cole model. The offline VCC was calculated from the cell viability % from flow cytometry data and dry cell weight concentration (DCW). The Cole-Cole model predicted the cell diameter and viable cell concentration with an error of 1.03% and 7.72%, respectively, and was validated using traditional offline methods like particle size analyzer and flow cytometry for cell size and cell viability measurements, respectively. The proposed approach can enable the operator to take real-time process decisions in order to achieve desired productivity and product quality.

## 3.2 Problem statement

Real-time monitoring and estimation of the physiological variable, biomass concentration, is significant for establishing control strategies based on this variable. In the production process of intracellular biotherapeutic products, the physiological changes in the biomass concentration have been found to influence the final product titer and quality, and therefore monitoring this process variable is substantial. Therefore, establishing reliable monitoring and acquiring real-time estimation of this physiological variable is essential. Among the various methods available to monitor this variable, DS is a promising PAT tool that has been employed to monitor and estimate several physiological changes as discussed in state of the art. However, there were a few setbacks like implementing a suitable data-pre-processing technique, acquiring real-time estimation of physiological properties, implementing the Cole-Cole model in a bacterial system, and integrating the DS data into a process model, as briefed in the previous section. This chapter aims to address these challenges concerning the application of DS

for real-time monitoring and estimation of biomass concentration in the production of biotherapeutic protein Ranibizumab (antibody fragment) in recombinant *E. coli* cultivation.

In this work, the real-time capacitance signal measured using the scanning DS was used to estimate the Cole-Cole model parameters by using a global optimization technique for a recombinant *E. coli* system. The application of the Cole-Cole model to estimate the physiological properties of a recombinant *E. coli* that is producing a therapeutic protein has not been attempted so far. Additionally, the generally applied nonlinear method of the Levenberg–Marquardt algorithm (LMA) was found to be not suitable for this case study. Therefore, a Genetic algorithm (GA) was applied to obtain a global solution for the parameters. Further, physiological properties, namely, cell diameter and viable cell concentration (VCC), obtained from the Cole-Cole model were validated using the traditional offline analytical methods, like particle size analyzer and flow cytometry, respectively. In order to overcome the shortcomings of LMA, a hybrid approach using GA and LMA was also proposed for real-time implementation of the Cole-Cole model for estimating physiological properties. The proposed Cole-Cole model could be applied for real-time estimation of the physiological properties, thereby enabling the operator to take real-time process decisions. Furthermore, this approach would be useful for predicting harvest time and developing capacitance-based control strategies leading to enhanced process monitoring and improved product quality.

## 3.3 Materials and Methods

### 3.3.1 Reactor configuration and experimental setup

Fed-batch experiments were performed in a 5 L bioreactor (Sartorius BIOSTAT B Plus, Germany) with a working volume of 3.5 L. The schematic representation of the reactor system is illustrated in Figure 3.2. Gaseous phase $CO_2$ and $O_2$ mole fractions (% v/v) of the off-gas were quantified based on the infrared and paramagnetic principles, respectively, by an off-gas analyzer (BlueInOne Cell, BlueSens, Germany). pH was maintained by adding 12% (v/v) $NH_4OH$ solution and 2N $H_3PO_4$, and the temperature was maintained with the aid of a chiller unit. A cascade controller was designed to maintain DO% in the reactor by manipulating the stirrer speed and air and oxygen flow rate and also by supplying pure oxygen as and when required. The substrate flow rate was

controlled manually through a peristaltic pump. All the real-time measurements were continuously logged at a regular time interval of 1 min using the supervisory control and data acquisition (SCADA) developed using the graphical programming software LabVIEW (National Instruments, Austin, TX, USA) (Chopda et al., 2016).

### 3.3.2 Strain and reactor operating conditions

A recombinant *E. coli BL21 DE3* strain expressing the therapeutic protein Ranibizumab as the antibody fragment was used in the present study. The cloning host (*E.coli DH5α*) and expression host (*E. coli BL21 DE3)* were acquired from Promega corporation (Madison, WI, USA). The vectors pET15b and pETDuet used for Ranibizumab expression were from Addgene (Watertown, MA, USA). The cloning approach presented in this work was carried out by cloning the entire antibody fragment in 2:2 ratio (2 copies each of light and heavy chains) in a pETDuet vector to enhance the productivity. The schematic view of the 2:2 clone of ranibizumab is presented in Figure 3.1 (adapted from (Priyanka and Rathore, 2021)).

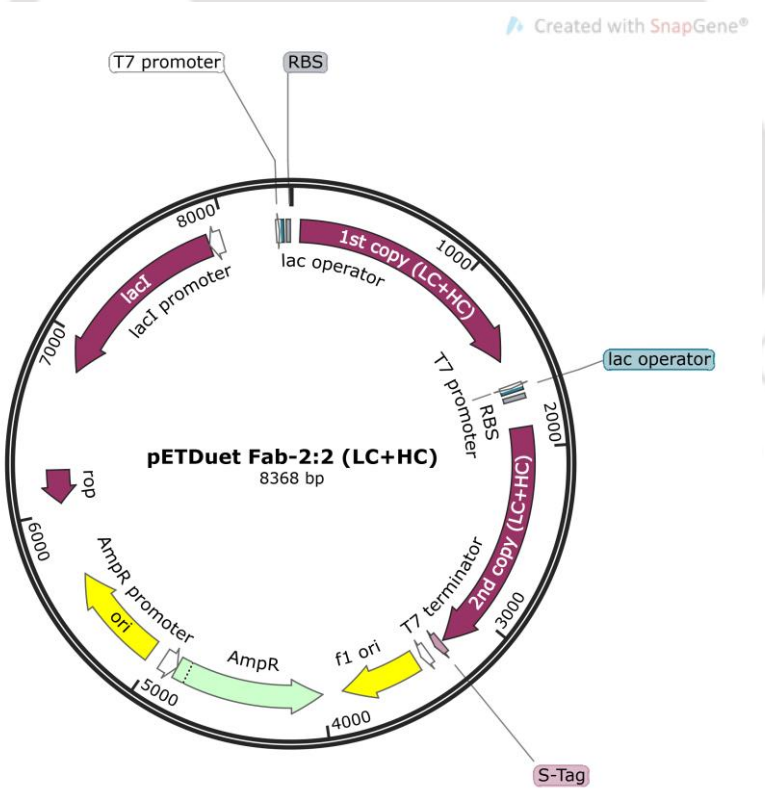



Figure 3.1. Schematic view of 2:2 ranibizumab gene cassette (adapted from (Priyanka and Rathore, 2021))

The seed culture grown overnight was prepared using the working cell bank (glycerol culture stock at -80°C) in the Luria Bertani broth for inoculating it into the bioreactor under aseptic conditions. The process was initially operated in the batch mode with the SOC medium containing 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, and 1% (w/v) $MgSO_4$ along with 2% (w/v) glucose as the sole carbon source (Atlas and Snyder, 2006). Operating conditions were maintained at pH 7 and 37°C temperature. The DO probe was polarized overnight under an initial 400 rpm and 24 L $h^{-1}$ air supply at 100% to achieve the saturation concentration of the DO. At the onset of the batch process, the DO setpoint was fixed to 10%. From the residual DO concentration, a sudden spike in DO levels closer to their saturation is a clear indication of exhaustion of the limiting substrate (glucose). Following that, a short pulse of the concentrated carbon source was fed into the reactor, following which a continuous feeding of 20% (v/v) glycerol mixed with 1% (w/v) yeast extract and 20 mM $MgSO_4$ with a constant rate of 0.054 L $h^{-1}$ was initiated. The addition of an inducer is an important step to activate the recombinant expression system in *E. coli* and initiate the production of the desired protein. In this study, induction was performed using 1 mM isopropyl $\beta$-d-1-thiogalactopyranoside (IPTG) when an optical density (OD) >40 was achieved (after 3-4 h of the fed-batch mode) in order to enable the production of intracellular recombinant proteins. The feeding was carried out until the capacitance value reached a stagnant phase, as it indicated the formation of inclusion bodies, and this occurred 3-4 hours after the induction (refer section 3.5.2). Therefore, the total feeding time in our reactor studies lasted for 7-8 hours.The batch was harvested after 12-13 h of operation.

Data from the fed-batch studies (2 sets) with frequency scanning capacitance data were consolidated for Cole-Cole model development and biomass estimation. Further, two sets of the reactor run with similar input conditions (constant feeding) were performed and used for the model calibration and validation studies, respectively.

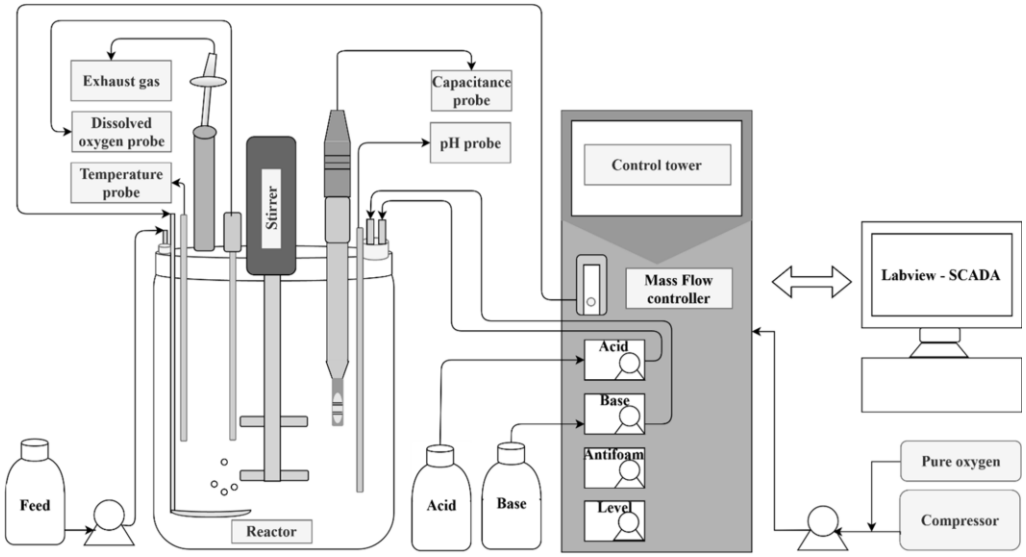



Figure 3.2. Schematic diagram for therapeutic protein production from *E. coli*

### 3.3.3 Capacitance measurements from recombinant *E. coli*

Reactor experiments were monitored using an *in–situ* annular type dielectric probe of 12 mm diameter to provide real-time capacitance measurements. The Aber Instruments Biomass Monitor (BM; Aber Instruments, Aberystwyth, UK), employing a four annular–electrode, which is non–invasive and in–situ sterilizable, was used in this study (Kaiser et al., 2008; Logan and Carvell, 2011; Priyanka et al., 2018). The dielectric probe was fitted in the reactor, as shown in Figure 3.3, and it was ensured that its four annular electrodes were well immersed inside the reactor media. The location of the capacitance probe was chosen such that it is not too close to the stirrer and aerator to minimize the disturbance of the electric field generated by the capacitance probe. The influence of agitation and aeration rates on the capacitance measurement of yeast suspension has been summarized in the literature (Maskow et al., 2008a). In our case study, a 5 L reactor was operated with a maximum working volume of 3.5 L, with an aeration rate of up to 1 vvm. Furthermore, the stirrer speed was chosen accordingly such that the hampering of the electric field was minimized.

An annular–type dielectric probe (Aber Instruments Ltd Aberystwyth, UK) equipped with the bioreactor setup offers an advantage for reliable estimation of viable cell concentration from the real-time capacitance measurements. The default bacterial dual-frequency mode was applied with the measuring and background frequencies of 1 and

15.65 MHz, respectively. The pre-processed capacitance data were used for developing the correlations and for estimating the real-time biomass concentrations in the reactor. A scanning software named FUTURA SCADA Version 2.1.0 (Aber Instruments Ltd., Aberystwyth, UK) was employed to perform frequency scanning for every 30 seconds at various frequencies, ranging from 0.1–20 MHz. Before the inoculation, a zeroing procedure was carried out for 15 min to remove the background medium capacitance from the measurements. The capacitance values were recorded at different frequencies using the inbuilt bacterial measurement mode available in the SCADA platform with a measuring frequency of 1120 kHz and a background frequency of 15.6 MHz.

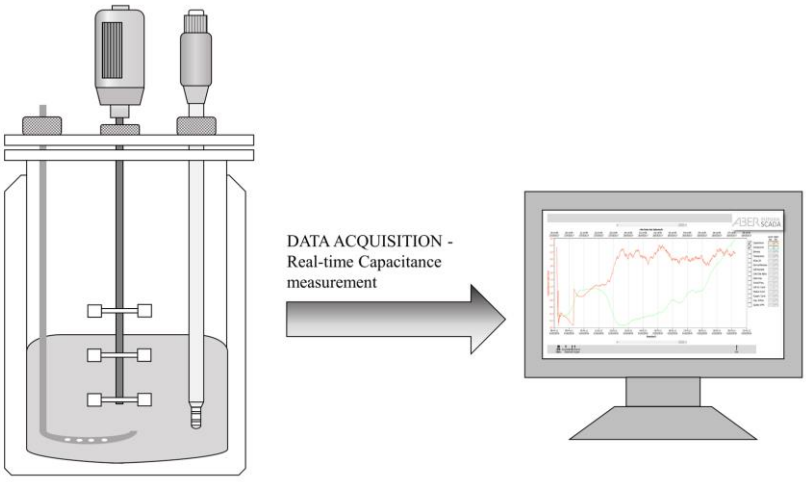



Figure 3.3. Experimental setup for measuring real-time capacitance using frequency scanning

### 3.3.4 Offline biomass measurements

The dry cell weight (DCW) was determined using the gravimetric method (centrifuging the broth samples at 10,000 rpm), followed by drying (the pellet was dried at 50°C overnight) in pre-weighed centrifuge tubes. The OD of the collected samples was measured from absorbance values at 600 nm. The deduced correlation between capacitance data and DCW was extended for different batch runs, employing the SOC medium and same operating conditions. Additionally, the following equation 3.1 was applied to introduce the correction factor for calculating viable cell concentration (VCC) from the measured DCW and the cell viability percentage (%) obtained from the flow cytometry analysis (Ehgartner et al., 2017) (refer to section 3.3.7).

$$VCC = DCW * Viability \quad (3.1)$$

### 3.3.5 Analytical methods

#### 3.3.5.1 Substrate concentration measurement (glucose and glycerol)

The residual substrate concentration (glucose and glycerol) in the supernatant of the collected samples was estimated using the analytical high–performance liquid chromatography (HPLC) (Agilent Technologies, USA) using an isocratic flow of 5 mM $H_2SO_4$ (mobile phase) at the rate of 0.6 mL $min^{-1}$ at 50°C through a Hi–plex H column (300 × 7.7 mm, Agilent technology, USA).

#### 3.3.5.2 Therapeutic protein measurement

About 25 mL of the culture broth was collected every hour after the induction, followed by centrifugation at 7000 rpm and 4°C. The pellet was washed with 1X phosphate-buffered saline (PBS) (8 g $L^{-1}$ NaCl, 0.2 g $L^{-1}$ KCl, 1.42 g $L^{-1}$ $Na_2HPO_4$ and 0.24 g $L^{-1}$ $KH_2PO_4$), followed by which 1 g cell pellet was lysed in 10 mL lysis buffer I (100 mM Tris-HCl buffer (pH 9), 1 mM ethylenediaminetetraacetic acid (EDTA) and 100 mM NaCl). The cell pellet was subjected to ultrasonication at 1 amplitude with 30 sec ON/OFF cycles. The lysate was then centrifuged and the pellet was resuspended in 1g/ 10 mL of lysis buffer II (100 mM Tris-HCl buffer (pH 9), 1 mM EDTA, 100 mM NaCl and 1% Triton X1000). Further, the pellet was washed with 1X PBS thrice and finally the washed pellet (1g/10 mL) was disolved in solubilization buffer (50 mM Tris-HCl buffer (pH 9), 8M urea) to obtain solubilized inclusion body (IB).

The solubilized IBs obtained from the harvest were subjected to refolding. Refolding is the crucial rate limiting step where the inactive and non-native protein conformation from IBs is refolded to its active and native conformation. Refolding was done for around 48 h with the addition of refolding buffer (50 mM Tris buffer (pH 10.1), 5% w/v sorbitol and 0.7 M arginine). The refolding step was followed by multiple chromatographic steps (Hydrophobic interaction chromatography (HIC) and Cation exchange chromatography (CEX)) for capture and purification of the recombinant protein, wherein the process and product related impurities were cleared and > 98% pure therapeutic product (ranibizumab) was obtained.

Quantitative estimation of the recombinant protein (Ranibizumab) was performed using the reversed-phase gradient HPLC (RP-HPLC) with a Zorbax 300 SB C8 column. The

mobile phase combination was A [water + 0.1% trifluoroacetic acid (TFA)] and B (acetonitrile + 0.1% TFA), with elution by gradient flow at 1.2 mL $min^{-1}$ and at 80 °C using an Agilent 1200 HPLC (Agilent Technologies, USA). The eluted proteins were detected at 235 nm. The total concentration of the recombinant therapeutic protein was calculated from the RP-HPLC of the standard and the recombinant therapeutic protein obtained from reactor studies according to equation 3.2, where $A_{sam}$ and $A_{std}$ represents area of sample and standard respectively, $V_{sam}$ and $V_{std}$ are the volume of sample and standard injected (µL), $C_{std}$ is the concentration of standard from UV absorbarance and p% is the percent purity of standard (Priyanka and Rathore, 2021).

$$\text{Protein concentration } (\mu g/\mu L) = \frac{A_{sam} * V_{std} * C_{std} * p\%}{A_{std} * V_{sam}} \quad (3.2)$$

### 3.3.6 Data pre-processing and noise removal

Data pre-processing is a method to remove outliers from raw data and reduce the signal-to-noise ratio without losing actual process data (Marison et al., 2013). The noisy raw capacitance data from the frequency scanning must be pre-treated using a suitable pre-processing technique to reduce the signal to noise ratio without losing valuable process data (Horta et al., 2012). The pre-processing methods applied in previous literature for filtering raw capacitance data from the frequency scan include moving average (MA) (Downey et al., 2014), and Savitzky-Golay (SG) (Dabros et al., 2009b; Maskow et al., 2008b, 2008a), to name a few.

The recorded raw data were smoothened by subjecting them to the second-order preprocessing algorithms of DYMOLA 2019 (version 2019 FD01, Dassault Systèmes, Vélizy–Villacoublay, France). DYMOLA has the options of various analog filters (Critically damped, Bessel, Butterworth and Chebyshev) that can filter the signals according to the specified cut-off frequency obtained from the frequency analysis. The reactor data from the SCADA were pre-processed in DYMOLA using one of the inbuilt filters (second-order algorithms) to reduce the signal noise.

The sampling interval for dual frequency and frequency scanning capacitance data collection was 60 and 30 seconds, respectively. Moving average filtering with a moving window of 10 min was applied for the raw capacitance data (Horta et al., 2015, 2012). Subsequently, MATLAB R2019a was used to apply various pre-processing techniques to the raw capacitance data. Different inbuilt pre-processing algorithms like moving average

(MA), Savitzky–Golay filtering (SG) for rejecting noise using least-squares fitting, smoothing splines (SS) as a curve fitting technique were used as a pre-processing algorithm, and their statistical significance were compared. Different methods were compared using a smoothness factor (SF), which was calculated from the sum of squared differences of the normalized values of the first derivatives of the respective signal, sum of squared estimate of errors (SSE), root mean squared error (RMSE) and correlation coefficient ($R^2$) as shown in equations 3.3-3.5. Among these parameters, more weight was given to the SSE and $R^2$ compared to the SF to emphasize our aim of reducing the signal noise without the loss of actual process data.

For the validation of the preprocessing methods, RMSE was applied as one of the statistical measures and was considered to be adequate for the comparison of the methods. Normalized RMSE (NRMSE) (RMSE/std deviation), is generally applied to normalize the error values when the models/algorithms are in different scales, which is not the scenario in this case, since the objective here is to compare different preprocessing algorithms (Rajamanickam et al., 2021).

$$\hat{y} = \left( \frac{y_i' - \bar{y}'}{\sigma_{y'}} \right)_{i=1}^{n} \tag{3.3}$$

$$SF = \sum \left( \frac{\left( \hat{y}_i - \hat{y}_{i+1} \right)^2}{4} \right) \tag{3.4}$$

$$SSE = \sum_{i=1}^{n} \left( y_{i,raw} - y_{i,pre-processed} \right)^2 \tag{3.5}$$

where $y$ represents the respective signal, including raw data and the pre-processed data obtained from various algorithms, $\bar{y}'$and $\sigma_{y'}$ represent mean and standard deviation of first derivative ($y'$) and $\hat{y}$ represents normalized first derivatives of the respective signal.

### 3.3.7 Cell size and viability studies

#### 3.3.7.1 Cell size studies: Particle size analysis (quantitative) and Field Emission Scanning Electron Microscope (FESEM) (qualitative)

Cell size and viability studies were performed to validate the physiological properties estimated by the Cole-Cole model in this study. The bacterial cell size was quantitatively

measured using the particle size analyzer, Zetasizer Nano ZS (Malvern Instruments, UK). The collected samples were centrifuged, and the cells were washed in 0.9% (w/v) saline solution before measuring the particle size.

The *E. coli* cells were qualitatively visualized using Field Emission Scanning Electron Microscope (FESEM) (Zeiss, Sigma300). Before visualizing the *E. coli* cells using Field Emission Scanning Electron Microscope (FESEM) (Zeiss, Sigma300), the washed cell pellets from the samples collected during induction and harvest time were re-suspended in 100 $\mu$L of 0.9% saline. 20 $\mu$L of this suspension was placed in 1×1 mm aluminium foil mounted on a glass petri plate. The samples were fixed using 2% (v/v) glutaraldehyde for 15 min. The samples were dehydrated using ascending grades of ethanol with a concentration from 30% to 100%. Each step lasted for about 15 min, and finally, the samples were incubated overnight for drying until the analysis. The samples were mounted on a SEM sample holder using coal tape and were sputter-coated with a thin gold layer for conduction. The images were taken in inlens mode with an acceleration voltage of 2 kV at different magnifications.

#### 3.3.7.2 Cell viability studies: Flow cytometry (quantitative) and Fluorescence microscopy (qualitative)

Fluorescence imaging of the bacterial cells was carried out using Nikon inverted microscope eclipse Ti–S (EINST Technology, Singapore) for qualitative visualization of the cell viability. The cells were diluted to maintain $10^7$ cells/mL (~OD of 1) in all the samples to visualize the cell viability using Fluorescence imaging. To 1 mL of the diluted sample, 1.33 $\mu$L (1.5 mM stock conc.) of propidium iodide (PI) (Sigma) was added for staining. The stained samples were incubated for 15 min at 37°C and then washed with 0.9% saline before the imaging. About 5–10 $\mu$L of the cells were re-suspended in saline and were mounted in the slide and visualized using 527–553 nm excitation filter under the microscope.

In addition to the microscopy, quantitative cell viability analysis was also performed using flow cytometry. The samples were subjected to dual staining using 5-carboxyfluorescein diacetate succinimidyl ester (CFDA) and PI (Sigma) for the live and dead cell staining, respectively. The $0^{th}$ h sample just after inoculation was considered as the positive control (100% viability). The samples from different time intervals of the fermentation (a- batch, b- fed-batch, c- induction and d- harvest) were centrifuged at

10,000 rpm and resuspended in saline to maintain $10^6$–$10^7$ cells/mL (~OD >0.6). About 2 $\mu$L (0.5 mg/mL working conc.) of the respective dyes were added to 0.5 mL of the samples for single staining. The CFDA stained samples were incubated in the water bath at 50°C for about 1 h, and the PI stained samples were incubated at room temperature (37°C) for 10 min in the dark. For dual staining, CFDA staining was carried out initially, followed by PI staining after the incubation. The stained samples were centrifuged and washed with saline after the incubation before the analysis. The modified protocol for sample preparation and dual staining was adapted from the previous literature (Amor et al., 2002; Mahato et al., 2016).

The samples were analyzed using a CytoFLEX flow cytometer (Beckman coulter, US) using CytoFLEX solution as the sheath fluid. The CFDA stained samples were detected by FL1 channel with 530/30nm bandpass filter, and PI stained samples were detected by FL3 channel with 670 nm long-pass filter. The resultant flow cytometry data collected from the CytExpert software were analyzed using FlowJO software to visualize and compare the control (without dye) and dual stained variants of the samples. The quadrant gating method was employed to obtain the cell viability percentage, which was further used to calculate viable cell concentration (VCC) as described in the equation 3.1 (section 3.3.4).

## 3.4 Model development for real-time biomass estimation

### 3.4.1 Linear modeling

Pre–processed capacitance data was used for developing correlations for estimating the real-time biomass concentrations in the reactor. In linear modeling, a linear calibration was developed using OriginPro2020 between the pre-processed capacitance data obtained from the dual-frequency scanning and the offline biomass concentrations (DCW and VCC).

### 3.4.2 Cole-Cole modeling

The Cole-Cole model was developed from the capacitance and conductivity values measured at various frequencies. The LMA and GA methods, which are deterministic and stochastic, respectively, were applied to solve the nonlinear least-square fitting of the pre-processed permittivity and conductivity data in the Cole-Cole equation. The objective

function for the nonlinear least-square minimization of capacitance and conductivity for a single time interval is defined in the equation 3.6, subjected to the bounds for the parameters, which were fixed based on previous literatures (especially $\omega_c$ and $\alpha$). The initial guesses for the parameters ($\Delta\varepsilon$, $\alpha$, $\omega_c$, $\varepsilon_\infty$, $\Delta\sigma$, $\alpha_2$, $\omega_{c2}$, $\sigma_L$) were chosen to be within the bounds, and it was updated after every iteration. The parameters obtained from the nonlinear least-square fitting were used in solving the nonlinear equations to estimate the physiological properties such as cell diameter and viable biomass concentration using equations (2.2),(2.5)-(2.7) as explained in sections 2.2.1 and 2.2.2.2. MATLAB R2019a was used for nonlinear least-square fitting and also for solving the nonlinear equations. The overall process of Cole-Cole modeling is illustrated in the flowchart in Figure 3.4.

$$Objective_{Cole-Cole} = Min \sum_{f=0.1}^{20} \left( \left| \left( \varepsilon_{f,model} - \varepsilon_{f,\exp} \right) + \left( \sigma_{f,model} - \sigma_{f,\exp} \right) \right| \right) \qquad (3.6)$$

where $f$ represents the scanning frequency at which capacitance and conductivity are measured (0.1–20 MHz).

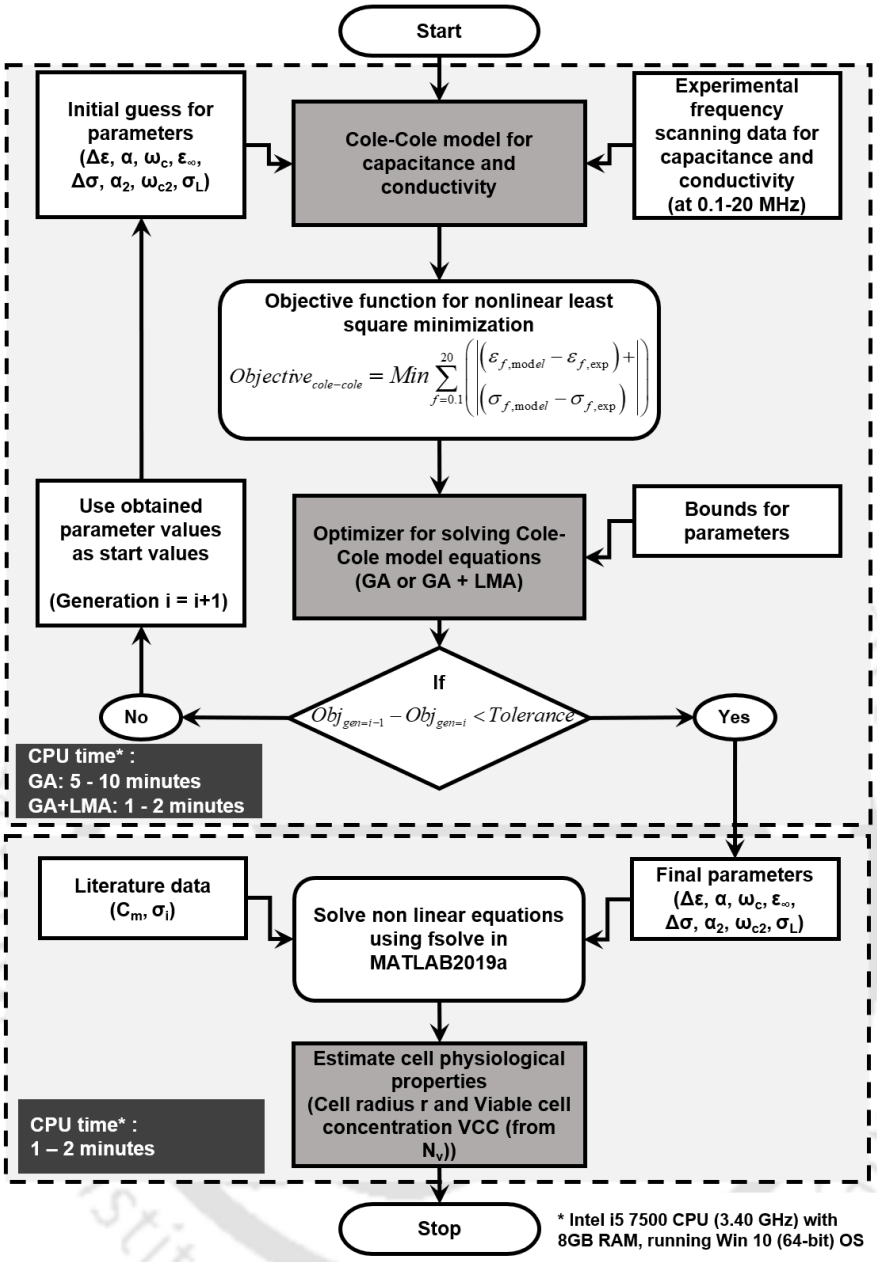



Figure 3.4. Flowchart for estimating the cell physiological properties using Cole-Cole model

## 3.5 Results and discussion

### 3.5.1 Data pre-processing and noise removal

The noise in the raw capacitance data was reduced by applying various pre-processing algorithms. Applications of different filters to the raw capacitance data are presented in Figure 3.5. As mentioned previously, three pre-processing methods, namely, MA, SG and

SS were applied, and the comparison of the aforementioned methods is presented in Table 3.1. Additionally, the smoothness factor SF of the three methods were compared. It is observed that the MA not only reduced the signal noise by smoothening the raw data up to 46%, as indicated by the SF factor, but also comparatively has a smaller SSE value and higher correlation coefficient. Therefore, it could be inferred that among the three best–suited methods, MA filtering eliminates the signal noise without much loss of the actual data. As reported in several literature, SG filtering could also be applied for scanning frequency capacitance signals (Cannizzaro et al., 2003; Ehgartner et al., 2017; Maskow et al., 2008a). In this study, MA filtering with a moving window of 20 points (10 min) was applied for the capacitance signals obtained from the frequency scanning data, considering that there would only be a minimal change in biomass concentration in 10 min duration. Figure 3.5 shows the various pre-processing techniques used for the raw capacitance data at three different representative frequencies (384 kHz, 1120 kHz, 4472 kHz). The three distinct frequencies were chosen for representation purposes, arbitrarily choosing them from three different points throughout the frequency range (0.1-20 MHz) to represent the $\beta$ dispersion curve at different frequencies (Randek and Mandenius, 2020). However, the pre-processing procedure was implemented for the data points from all the different frequencies throughout the scanning region, which were further used for the Cole-Cole model development and validation. The capacitance data at the three different frequencies are represented with decreasing thickness and colour transparencies in Figure 3.5.

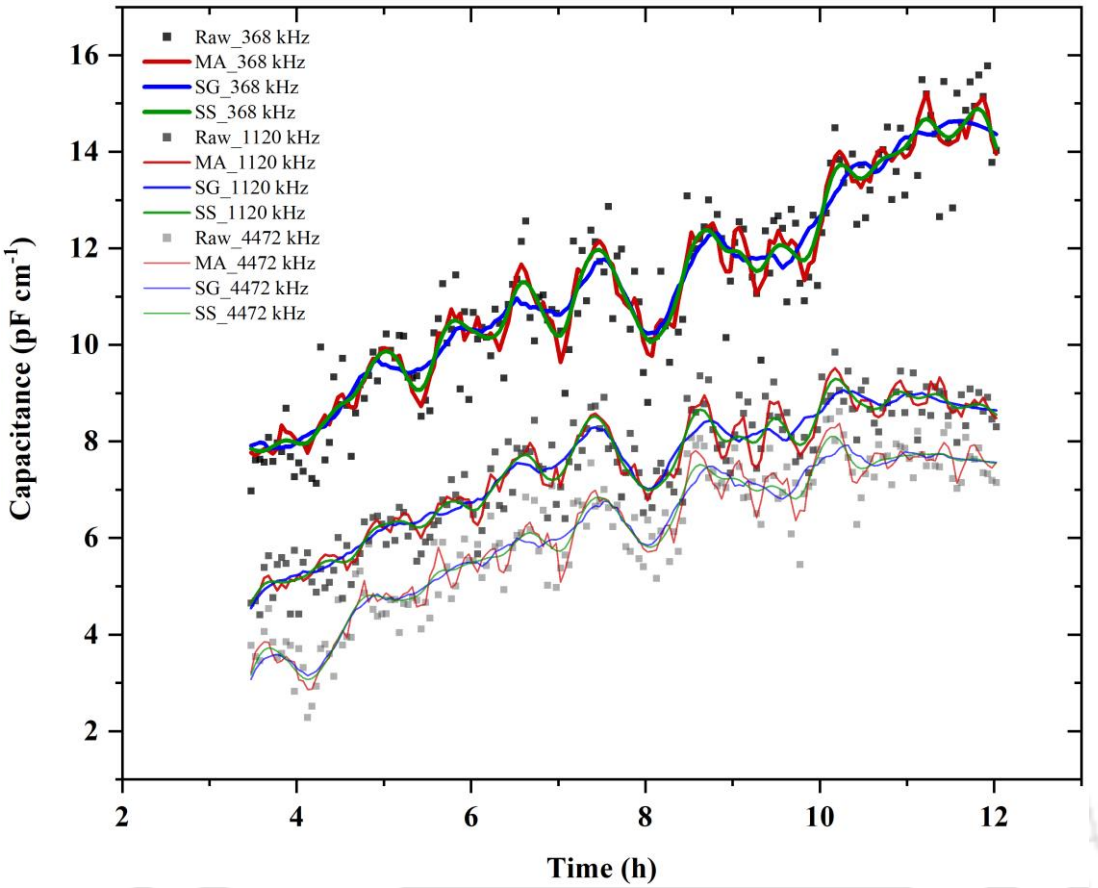



Figure 3.5. Data pre-processing for raw capacitance data using moving average (MA), Savitzky Golay (SG) and Smoothing spline (SS) represented at three different frequencies, 384 kHz, 1120 kHz and 4472 kHz (arbitrarily chosen for representation)

Table 3.1. Statistical comparison of various pre-processing methods for the raw capacitance data

| **Pre-processing Method** | **Raw data** | **Moving average (MA)** | **Savitzky - Golay (SG)** | **Smoothing spline (SS)** |
|---|---|---|---|---|
| Smoothness factor (SF) | 1030.25 | 557.45 | 315.90 | 5.14 |
| Sum of squared estimate of errors (SSE) | NA | 18874.11 | 23286.01 | 20113.75 |
| Correlation coefficient ($R^2$) | NA | 0.922 | 0.916 | 0.920 |
| Root mean squared error (RMSE) | NA | 0.870 | 0.880 | 0.886 |

### 3.5.2 Production of therapeutic recombinant protein from *E. coli*

Reactor studies were performed for 12 h, and the data acquired in SCADA were pre-processed, as explained in section 3.3.6. Figure 3.6 shows the pre-processed DO and capacitance values with respect to batch time for the production of Ranibizumab from *E. coli*. It was observed that around the fifth hour, the DO concentration increased up to 60%, indicating completion of the batch phase. When the fed-batch phase was started, it was observed that the DO spike subsided and reached its setpoint value of 10%. The pre-processed capacitance profile shown in Figure 3.6 indicates physiological changes between the various phases, including batch, fed-batch, and induction (Ehgartner et al., 2015). The capacitance values have been proven to be influenced by cell morphological changes and internal protein composition (Kaiser et al., 2008). As observed in Figure 3.6, the capacitance value reaches a stagnant value during the production phase (after 8-9 h). This stagnant value could be associated with the formation of IBs, where the flux gets diverted from biomass growth toward the production of the intracellular therapeutic recombinant protein. The decrease/ stagnation in the capacitance values toward the end of the batch is an indicator of cell morphological changes and a possible decrease in cell viability during the production phase.

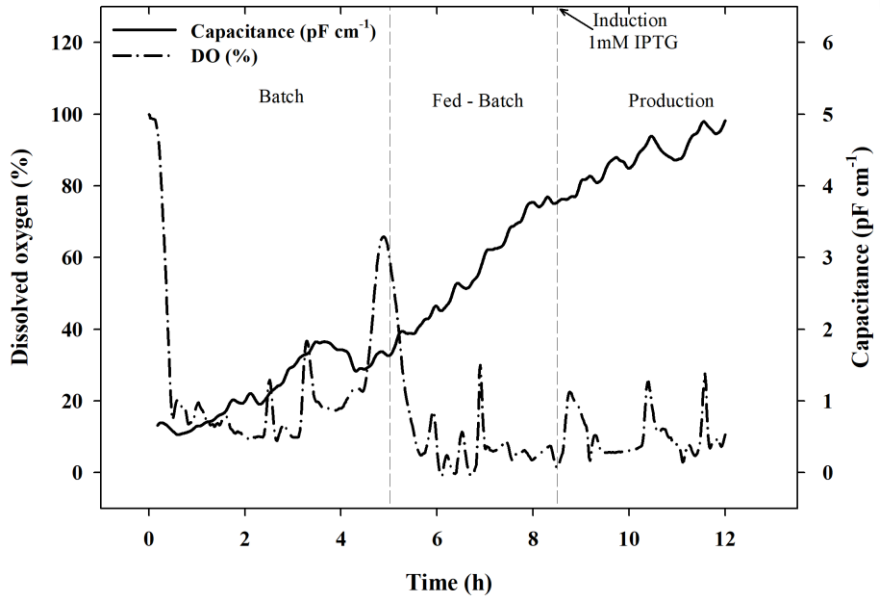



Figure 3.6. Experimental profiles for Dissolved oxygen and capacitance throughout the different phases of fermentation (batch, fed-batch and production)

### 3.5.3 Capacitance measurements from recombinant *E. coli*

#### 3.5.3.1 Real-time biomass estimation from dual-frequency capacitance measurements

Real-time biomass concentration was estimated from the recorded capacitance values, as explained in section 3.3.3. Figure 3.7 (A) represents the comparison of raw capacitance data and pre-processed capacitance data with the offline DCW values. It is observed that the capacitance values rose as high as 5 pF $cm^{-1}$. The capacitance plot also indicated that the rate of change in DCW varies for the three phases in the order of batch > fed-batch > production. The correlations for biomass estimation from the measured capacitance could be developed using linear or nonlinear models (Dabros et al., 2009b). As shown in Figure 3.7 (B), the linear relationship between the pre-processed capacitance and DCW is observed. The measured capacitance signals corroborated well with the offline DCW values. The corresponding linear equation for estimating the biomass concentration from the measured capacitance data is shown in Figure 3.7 (B), with an $R^2$ value of 0.9663. It could be noted that the dielectric probe records the non-zero values for the background capacitance of the medium without cells. These recorded non-zero capacitance values contribute to the negative intercept, and since the value of the intercept is very small, it could be ignored (Ehgartner et al., 2015; Maskow et al., 2008a; Metze et al., 2019).

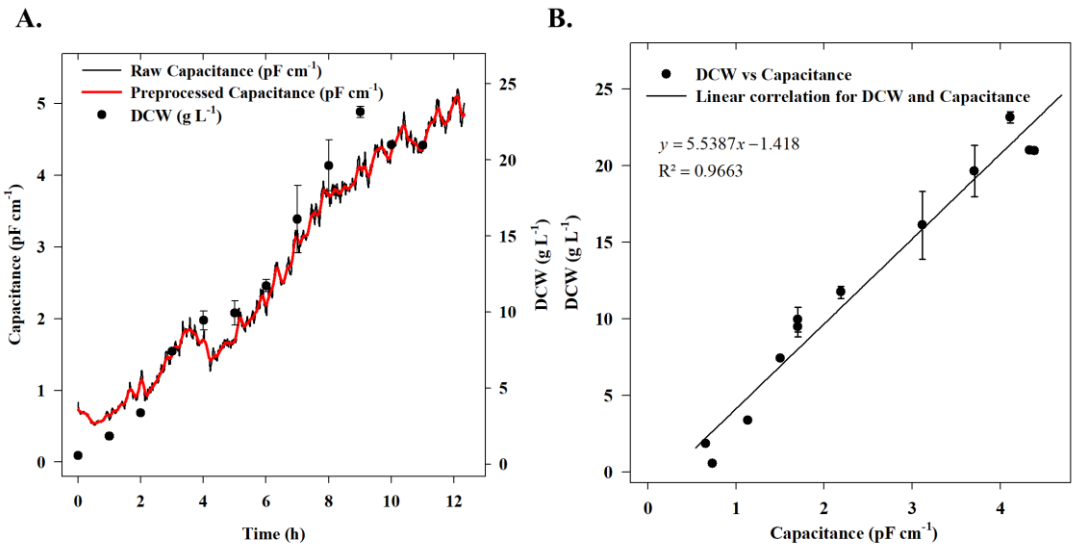



Figure 3.7. (A). Comparison of raw capacitance, pre-processed capacitance and offline DCW values. (B). Linear correlation for DCW and capacitance

### 3.5.3.2 Capacitance measurements using frequency scanning

Figure 3.8 represents the pre-processed capacitance profile plot at different frequencies for various time intervals of the fermentation. Data from four different time intervals corresponding to the different phases of fermentation are depicted, which would henceforth be listed as follows: a: Batch ($2^{nd}$ h); b: Fed-batch ($5^{th}$ h); c: Induction ($8^{th}$ h); and d: Harvest ($12^{th}$ h). It is evident from Figure 3.8 that the measured capacitance values increase with time during the fermentation, which is in accordance with the increasing concentration of biomass. The characteristic $\beta$ dispersion curve from the frequency scanning indicates that the capacitance values decrease with increasing frequency (Figure 3.8 (A)). At higher frequencies, the cells do not have sufficient time to get polarized and hence the capacitance values are closer to zero. The observed characteristic $\beta$ dispersion curve is in accordance with the observation in previous studies for various organisms (Downey et al., 2014; Randek and Mandenius, 2020; Tibayrenc et al., 2011).

Figure 3.8 (B) shows a comparison of the pre-processed capacitance data at default bacterial measuring frequency 1120 kHz and the offline measured DCW. Figure 3.8 (B) further illustrates the viable cell concentration (VCC) calculated from the measured DCW and cell viability obtained from flow cytometry analysis according to the equation 3.1. It is observed that during the induction phase (after 8 h), the capacitance attains a stagnant value, which reveals a possible change in cell morphology and/or decrease in cell viability (Figure 3.8 (B)). The observation also corresponds to the VCC profile because, after 10 h, the cell viability decreased and reached a value of 42% around the harvest period (section 3.5.5). However, this was not observed in the case of DCW, which continued to increase until the end of the fermentation, thereby deviating from the measured capacitance values and the corrected offline VCC values. Therefore, it can be inferred that the capacitance measurement corresponds only to the live cells in the reactor, as shown by the online capacitance profiles and corrected VCC in Figure 3.8 (B). Furthermore, as described in the subsequent sections, the estimation of real-time biomass concentration from the reported capacitance values using different models was performed.

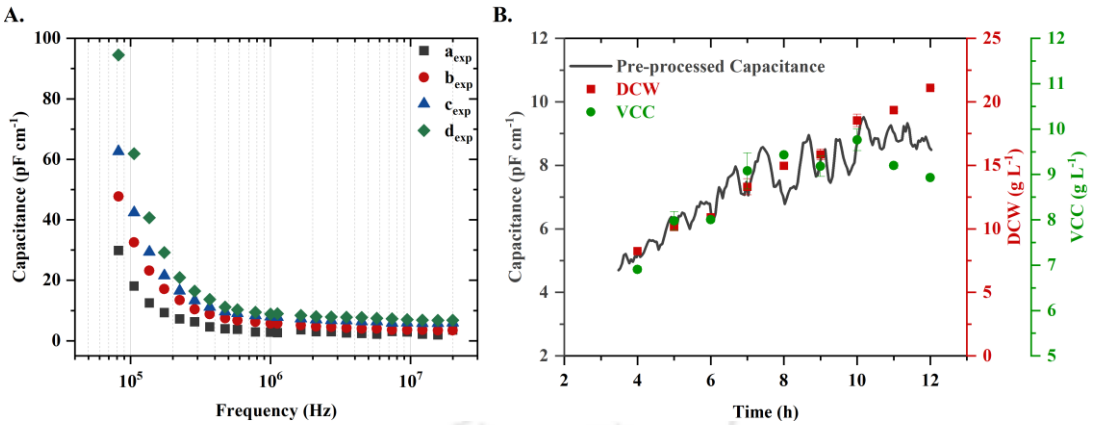



Figure 3.8. (A). Frequency scanning capacitance profile represented for four different time intervals of fermentation, namely a (batch), b (fed-batch), c (induction) and d (harvest). (B). Pre–processed capacitance (at 1120 kHz), offline dry cell weight (DCW, red filled square), and viable cell concentration (VCC, green filled circle) throughout the fermentation.

### 3.5.4 Offline analysis

The offline analysis for the biomass (DCW), substrate (batch phase substrate - $S_b$, fed-batch phase substrate- $S_f$) and product concentration ($P$) is represented in Figure 3.9. This data was used for the model calibration and validation studies. The values are plotted as an average of duplicate reactor runs, and the error bars are plotted accordingly.

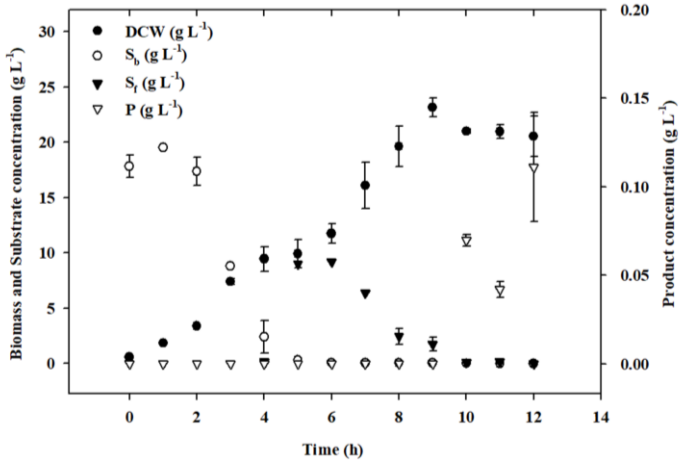



Figure 3.9. Offline analysis for biomass (DCW, filled circle), substrate ($S_b$, empty circle, $S_f$, filled downward facing triangle) and product ($P$, empty downward facing triangle) concentration.

### 3.5.5 Cell size and viability studies

Offline cell size and viability analysis were performed to validate the application of capacitance as a PAT tool to estimate the physiological properties of recombinant *E. coli* considered in this study. As the recombinant therapeutic protein considered in this study is obtained in the form of inclusion bodies, changes in cell size or cell shape could be expected to occur (Randek and Mandenius, 2020). The viability analyses were performed using flow cytometry and fluorescence microscopy techniques. The percentage of viability was calculated using the flow cytometry dual staining analysis.

#### 3.5.5.1 Cell size analysis

The cell size analysis was carried out using a particle size analyzer, and FESEM is shown in Figure 3.10. The cell diameter was measured using particle size analysis, which measures the average size of the particles in the suspension using the principle of dynamic light scattering (DLS) (Loske et al., 2014). As observed from the results of particle size analysis shown in Figure 3.10 (A), the cell diameter increased up to 6 h and then is observed to be maintained around 1 $\mu$m throughout the fermentation. The particle size analyzer measures the average size of the particles suspended in the sample, and it could be noted that the measured average value of 1 $\mu$m can be correlated to the average diameter of the typical rod-shaped *E. coli* cells present in the suspension.

Additionally, the difference between the cell size at the time of induction and harvest (samples c and d) was also observed from the FESEM analysis as presented in Figure 3.10 (B). The FESEM analysis also confirmed the rod-shaped appearance of the *E. coli* cells, and it was observed that there was no significant change in cell shape between the induction and harvest samples and that there is a minor change with respect to cell size. The slight decrease in cell size from induction to harvest could be attributed to the stress-induced due to the formation of inclusion bodies. The observations obtained from the FESEM can be corroborated with the measured capacitance values to reveal more insights related to the physiological changes occurring throughout the fermentation (Randek and Mandenius, 2020).

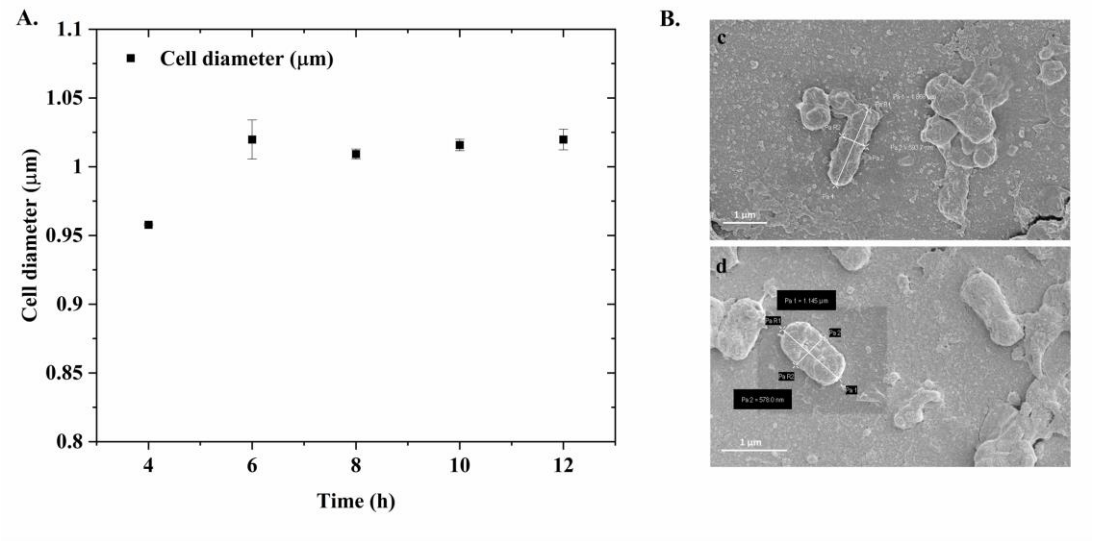



Figure 3.10. Bacterial cell size analysis. (A). Cell diameter measurement using particle size analyzer. (B). Scanning electron microscopy images of bacterial cells at the time of induction (c) and harvest (d)

### 3.5.5.2 Cell viability analysis

The results from the cell viability studies are summarized in Figure 3.11. Figure 3.11 (A) represents the percentage of viable and dead cells analyzed through quadrant gating of flow cytometry. Figure 3.11 (B) shows the quadrant gating for the control and dual stained samples for samples a, b, c, and d, respectively. The quadrants were segregated to represent the different subpopulations of the samples, with the bottom left quadrant indicating the unstained cells. In the contour plot of dual staining studies, the top left quadrant corresponded to the percentage of live cells in the samples and the bottom right quadrant corresponded to the percentage of dead cells in the respective samples. It could be observed that viability decreased up to 42% towards the end of the fermentation, with a proportional increase in the percentage of dead cells by 24%. These cell viability values were used for calculating the corrected value of viable biomass concentration (VCC) from the measured DCW using the equation 3.1. The cell viability decreased significantly after the induction, as reflected in the viability values of samples c and d in Figure 3.11 (A, B). This observation further indicates that the measured capacitance values can correlate well with the viable cell concentration present in the bioreactor. The changes in cell viability were visualized qualitatively from the fluorescence microscopy studies. The decrease in the viability was also observed in the fluorescence microscopy with PI staining from Figure 3.11 (C).

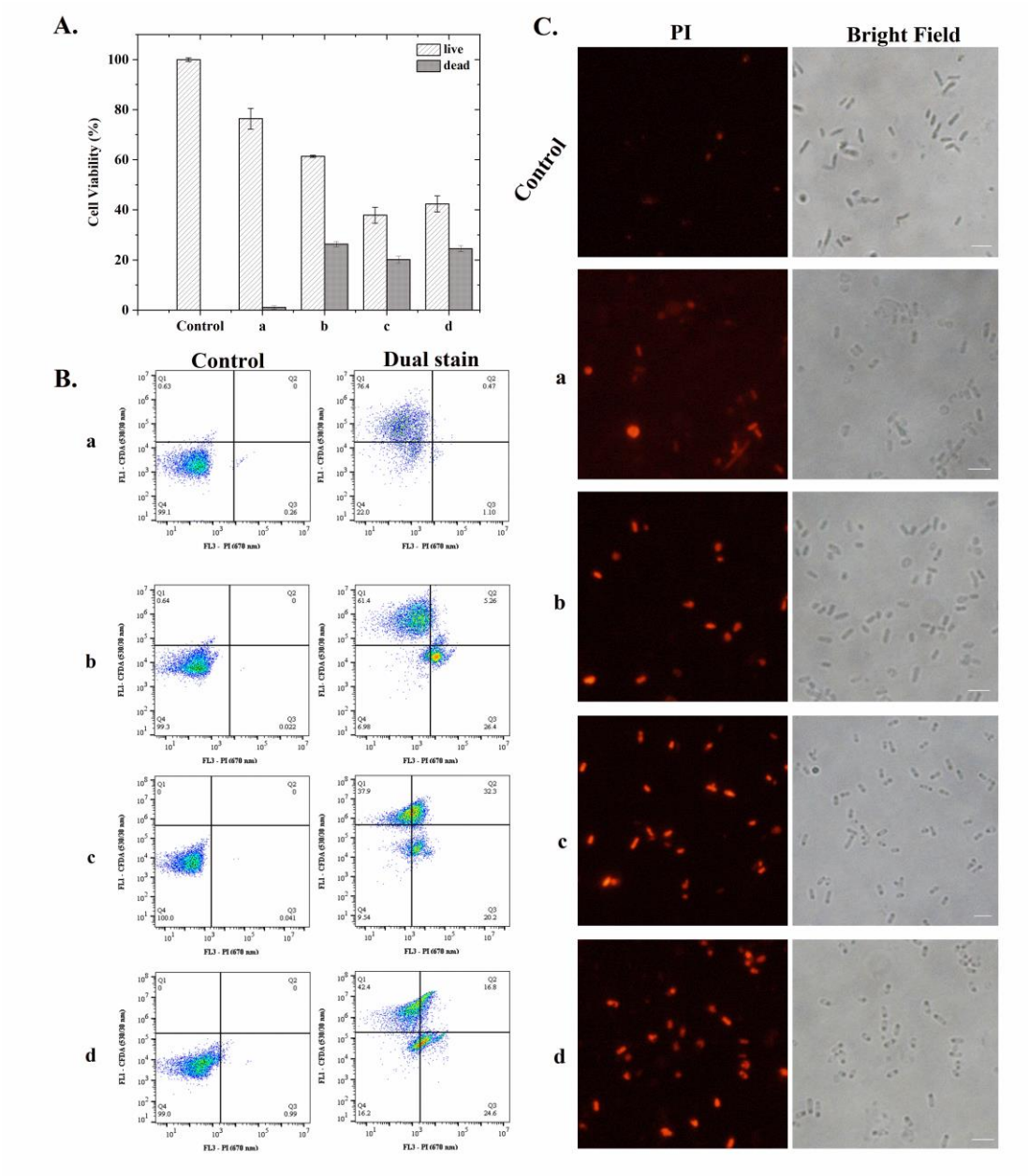



Figure 3.11. Bacterial cell viability analysis. (A). Cell viability percentage obtained from flow cytometry. (B). Quadrant gating for control and dual stain obtained from flow cytometry for four different time intervals of fermentation, namely, a (batch), b (fed-batch), c (induction) and d (harvest). (C). Propidium iodide staining studies using fluorescence microscopy for the samples a, b, c and d.

## 3.6 Real-time biomass estimation from capacitance measurements

### 3.6.1 Linear modelling

Application of different models for the correlation of measured capacitance and biomass concentration is a significant step in developing a capacitance-based control strategy (Knabben, 2011). Linear modeling is the most commonly used approach for establishing a direct correlation between biomass concentration and capacitance data. Figure 3.12 represents the linear correlation of pre-processed capacitance data with the offline DCW and corrected VCC values. Three representative frequencies, 384 kHz, 1000 kHz, and 4472 kHz, were chosen to represent the linear correlation model for estimating biomass concentration. The slope and $R^2$ values of the represented linear correlations are presented in Table 3.2. It is observed that the average $R^2$ values of the linear correlation established using VCC values were higher than that of those developed using DCW values. The $R^2$ values indicate that the measured capacitance values have a better correlation with the VCC values, which was also observed from steeper slopes with a three-fold increase than the slope values from the DCW correlation.

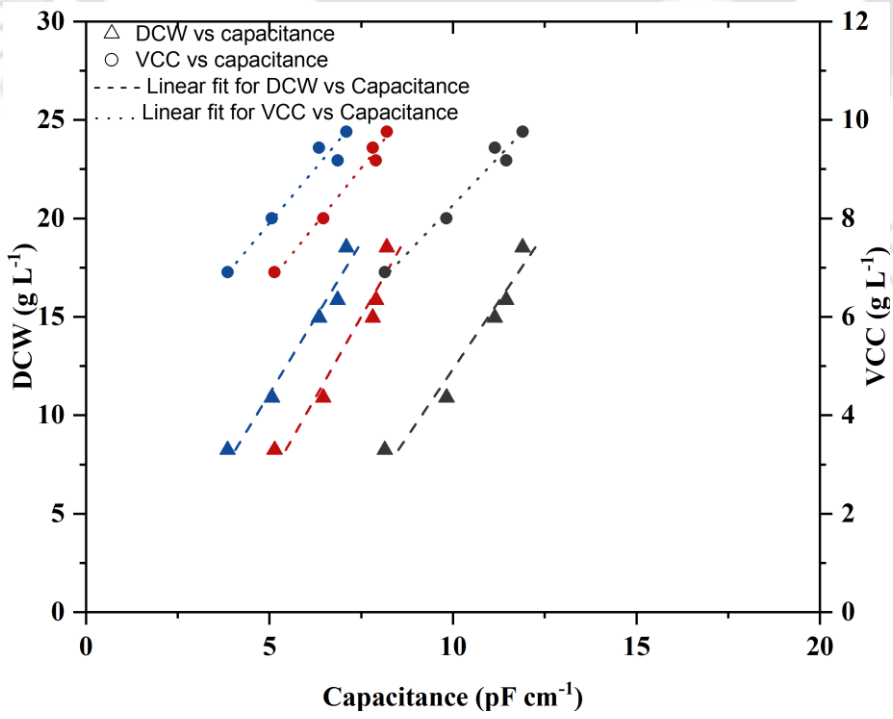



Figure 3.12. Linear correlation for DCW vs capacitance represented in triangles (▲) and correlation for VCC vs capacitance represented in circles (●). Three representative frequencies, 368 kHz (grey), 1120 kHz (red), 4472 kHz (blue), are represented.

Table 3.2 Linear correlation model for capacitance and offline biomass concentration (DCW and VCC)

| **Offline analysis** | **DCW (g $L^{-1}$)** | | | **VCC (g $L^{-1}$)** | | |
|---|---|---|---|---|---|---|
| Freq (kHz) | 368 | 1120 | 4472 | 368 | 1120 | 4472 |
| color | grey | red | blue | grey | red | blue |
| Slope | 0.363 | 0.324 | 0.306 | 1.272 | 1.181 | 1.126 |
| $R^2$ | 0.952 | 0.939 | 0.923 | 0.974 | 0.986 | 0.957 |

### 3.6.2 Cole-Cole modelling

The Cole-Cole model is a theoretical nonlinear model to estimate the cell radius and cell number (volume fraction) from the scanning capacitance and conductivity data. Initially, the fitting of permittivity and conductivity data of the Cole-Cole equation (equation (2.3) , (2.4)) was carried out using Levenberg–Marquardt algorithm (Dabros et al., 2009b; Maskow et al., 2008a; Tibayrenc et al., 2011). Although the algorithm seems to give a satisfactory agreement between the experimental and model-predicted values of permittivity and conductivity at various time intervals, the obtained parameters were erroneous and unsatisfactory. Moreover, the different start values yielded different values for the parameters indicated that the function has multiple local minima, and LMA yielded different local minima for different start points. As a result, applying these parameters for solving the nonlinear equations could not yield reliable results for the cell's physiological properties, and it was concluded that in our current case study, a stochastic global algorithm is needed. In view of this observation, a global evolutionary Genetic Algorithm (GA) was implemented to solve the Cole-Cole model. More details with respect to the comparison of both methods are presented below.

#### 3.6.2.1 Comparison of the deterministic (LMA) and stochastic (GA) method for solving the non–linear equations of the Cole-Cole model

The nonlinear least square fitting for the capacitance and conductivity equations (equation (2.3), (2.4)) of the Cole-Cole model was carried out using both deterministic and stochastic methods. It was observed that, while the LMA method was employed, different start values resulted in different optimal values for the parameters. This observation

indicated that the objective function has multiple local optima. Therefore, it was concluded that, for the current case study, the stochastic method GA yielded better and consistent values for the Cole-Cole model parameters, irrespective of the start values.

Obtaining reliable parameter values from the nonlinear least-square fitting of the Cole-Cole model is crucial for the estimation of the cell's physiological properties. Hence, the significance of choosing an appropriate method for solving the nonlinear equations of the Cole-Cole model were observed in this study. The comparison of the parameters estimated by the two methods LMA and GA, with different start values, are presented in Figure 3.13. It is observed that for different start values of the parameters, LMA resulted in different optimal values, while optimal values obtained from GA was similar to each other despite choosing different start values. Figure 3.13 represents the optimal values obtained for three parameters $\Delta\varepsilon$, $\omega_c$ and $\sigma_L$ (1, 3 and 8) which are essential for estimating the cell diameter and viable cell concentration using equations (2.2),(2.5)-(2.7).

Moreover, it was also observed that while choosing a start value nearer to the global optimum, both methods could yield almost similar results. This further emphasizes the significance of the start values and also reveals the constraints of the respective algorithms. Therefore, it could be emphasized that a hybrid methodology using both GA and LMA could also be considered, where the LMA algorithm is applied using the start values obtained from GA. This hybrid approach could be carried out to obtain quicker results for the parameter estimation and could later be useful for real-time implementation of the proposed strategy. Hence, it was concluded from our study that a GA algorithm or a hybrid algorithm was better suited to obtain consistent global values for the parameters, and thereby a reliable estimation of the physiological properties of recombinant *E. coli* cultivation could be achieved.

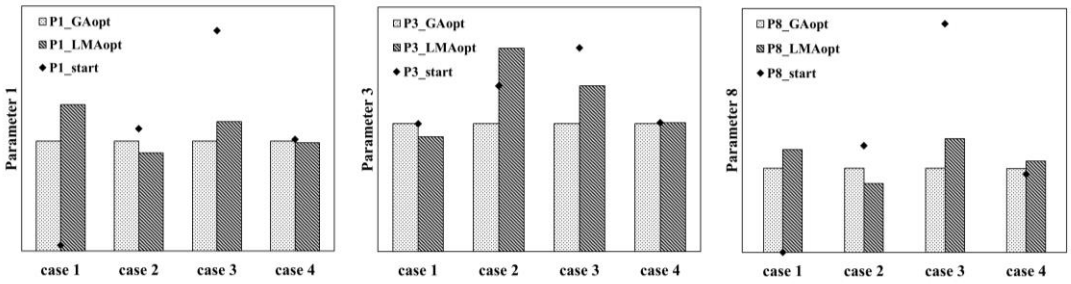



Figure 3.13. Comparison of parameter values ($\Delta\varepsilon$, $\omega_c$ and $\sigma_L$ (1, 3 and 8)) obtained from LMA and GA for different start values.

### 3.6.2.2 Cole–Cole model predictions

The comparison of experimental and Cole-Cole model predicted capacitance profile for the samples a, b, c and d through different stages of the fermentation is represented in Figure 3.14. The model predicted capacitance profile was obtained using nonlinear curve fitting of the equation (2.3) using GA. It could be observed from Figure 3.14 that there is considerable agreement between the experimental and Cole-Cole model predicted capacitance profile, with an average error value of 15.94%.

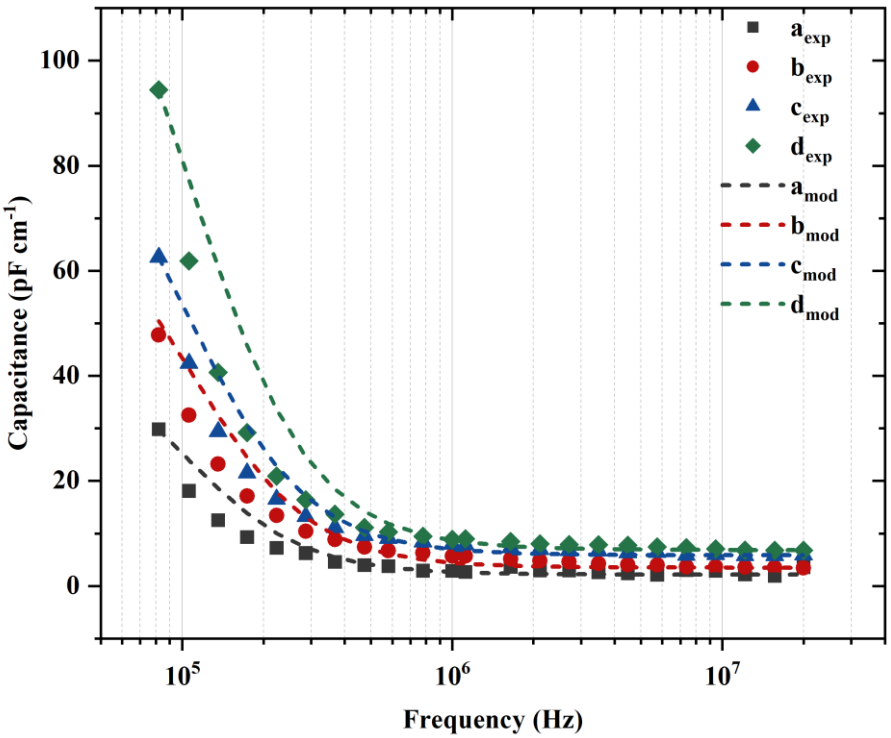



Figure 3.14. Experimental and Cole-Cole model predicted capacitance profile for four different time intervals of fermentation, namely, a (batch), b (fed-batch), c (induction) and d (harvest).

The estimated parameters obtained from the nonlinear fitting of the capacitance and conductivity data through the GA approach were later applied to solve the nonlinear equations, as mentioned previously. The comparison of experimental and model-predicted values of the physiological properties of the recombinant *E. coli*, namely the cell diameter and viable biomass concentration VCC, is presented in Figure 3.15 (A). The error for the model predictions were 1.03% and 7.72% for the cell diameter and VCC, respectively. $R^2$ for the Cole-Cole prediction of biomass concentration VCC was 0.93, as shown in Figure 3.15 (B).

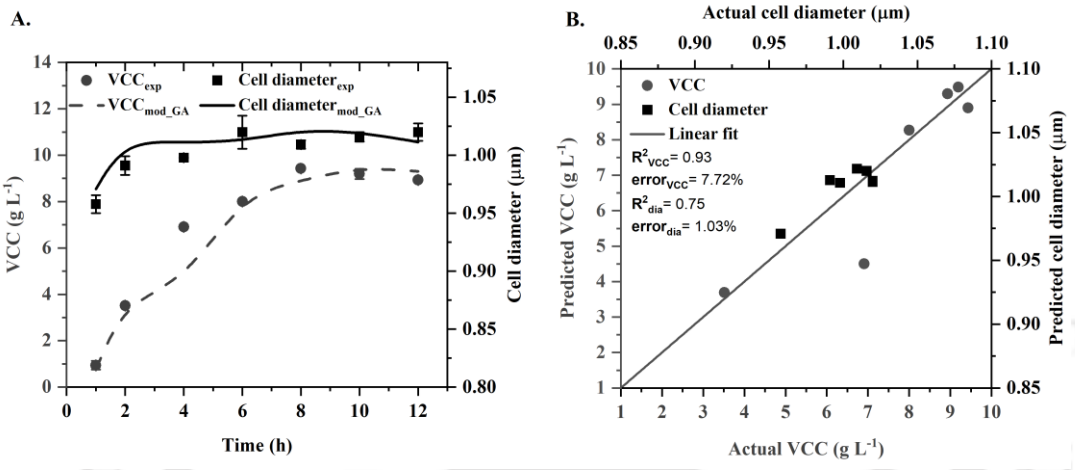



Figure 3.15. (A). comparison of experimental and Cole-Cole model predicted cell physiological properties viz., cell diameter and viable cell concentration (VCC). (B). Cole-Cole model predictions for VCC (filled circle) and cell diameter (filled square).

### 3.6.2.3 Real-time estimation of physiological properties using Cole-Cole model

The overall procedure for real-time estimation of physiological properties using the Cole-Cole model is summarized in Figure 3.16. The parameters of the Cole-Cole model can be estimated by two strategies. The first strategy is to employ GA to estimate the parameters from the nonlinear fitting of equations (2.3) and (2.4) using the pre-processed capacitance and conductivity data, which would take about 5–10 min for every dataset. In case if there is a need to estimate the parameters for shorter durations, the second strategy with hybrid methodology could be applied. In this second strategy, the parameters are initially estimated using GA (for the first iteration at time t =10, considering a 10–min moving average for pre-processing) and then LMA is applied further (for 2$^{nd}$ iteration onwards from time t >10), where the results from GA are used as start values for LMA. This hybrid strategy could further reduce the time to solve the Cole-Cole model as indicated by the

processor time. Further, the estimated Cole-Cole parameters from both strategies are used for estimating the cell's physiological properties.

The comparison of predicted cell physiological properties using both strategies is presented in Figure 3.16. It was observed that the hybrid strategy of both GA and LMA also yielded comparable predictions. The hybrid strategy (GA+LMA) estimated the physiological properties very similar to the GA strategy with an average improvement of 0.18% for both cell diameter and VCC, as observed in Figure 3.16. However, the major advantage in hybrid strategy is the overall reduction in the time taken for solving Cole-Cole model. In that aspect, the hybrid strategy leads to 80% reduction (per dataset) in time taken to estimate the physiological properties while also providing comparable results similar to the GA strategy. Therefore, this hybrid strategy could be considered for real-time application of the Cole-Cole model to predict the physiological properties, considering the shorter processing time required by the same. Furthermore, these estimated physiological properties could be employed to take real-time process decisions and develop capacitance-based control strategies.

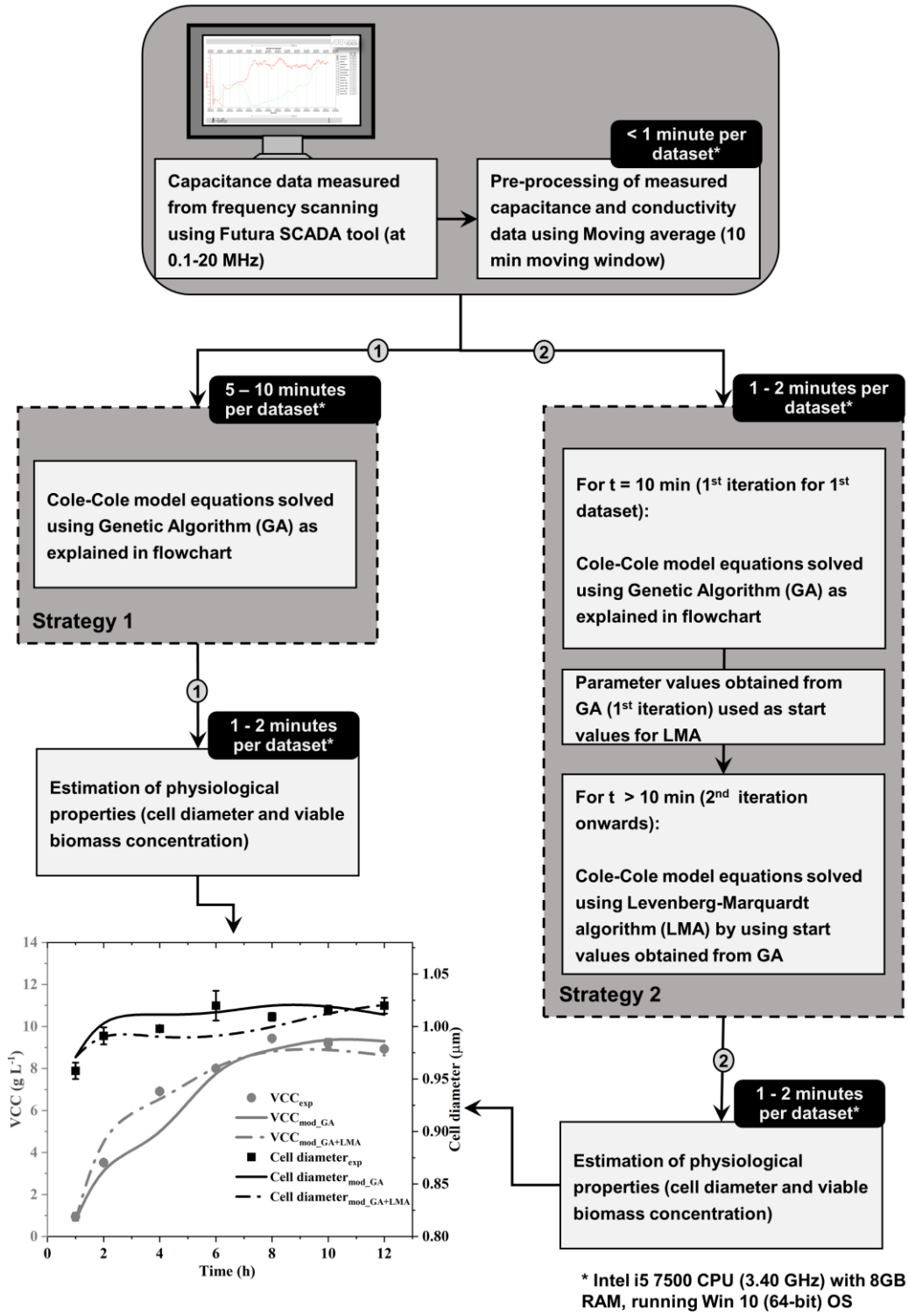



Figure 3.16. Overall procedure for real-time estimation of physiological properties using Cole-Cole model

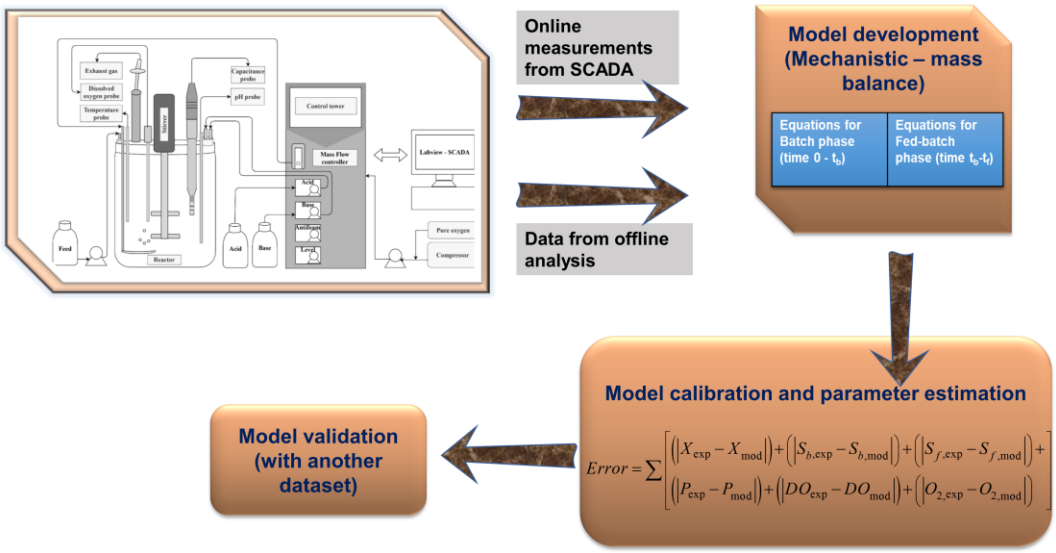
## 3.7 Summary

Reactor studies were performed for the production of recombinant therapeutic protein Ranibizumab from *E. coli*. Real-time biomass was monitored by the PAT tool, namely, dielectric spectroscopy. The measured capacitance signals were pre-processed, and a linear model was developed for obtaining the correlations between the online capacitance and biomass concentration.

In the current case study, estimation of physiological properties of recombinant *E. coli* was attempted by solving the Cole-Cole equation using the GA approach. DS was used for monitoring real-time biomass in the production of recombinant therapeutic protein Ranibizumab from *E. coli*. Capacitance signals were measured using frequency scanning methodology, and the obtained raw data was pre-processed using different methods, of which the moving average was found to be best suited for our system. Models for estimating the correlations between the online capacitance and biomass concentration were developed. The Cole-Cole model was solved using the genetic algorithm, and the physiological properties, namely cell radius and viable biomass concentration were estimated using the nonlinear fitting parameters. Among the two methods, LMA and GA, for solving the Cole-Cole model, it was observed that global optima values for the respective parameters were obtained from GA irrespective of the start values. The significance of obtaining global optima for the parameters for reliable estimation of cell's physiological properties was thereby examined in this study. The Cole-Cole model predicted the cell diameter and viable cell concentration with an error of 1.03% and 7.72%, respectively. The estimated physiological properties were validated using the offline studies of cell viability and cell size. Furthermore, a hybrid approach using GA and LMA was proposed to reduce the time taken to solve the Cole-Cole model and enable the real-time implementation of the proposed strategy. Therefore, the application of dielectric spectroscopy for the real-time monitoring of *E. coli* cultivation and estimation of cell's physiological properties in a biotherapeutic production process was successfully demonstrated in this chapter.

# CHAPTER 4

# Modelling and validation of batch and fed-batch process of the fermenter




---

Parts of this chapter is based on the following research article(s):

**Swaminathan, N.,** Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2020. Multiobjective Optimization for Enhanced Production of Therapeutic Proteins in Escherichia coli: Application of Real-Time Dielectric Spectroscopy. Ind. Eng. Chem. Res. 59, 21841–21853. https://doi.org/10.1021/acs.iecr.0c04010.

## 4.1 Foreword

The present chapter focuses on developing a mechanistic model based on mass balances of the various state variables of fermentation and the application of the model to optimize the total biomass with the aid of online capacitance measurements. The developed mechanistic model has been validated using experimental data sets obtained from the production of a therapeutic product, Ranibizumab, from *E. coli*. DS was deployed to monitor the production process and estimate real-time biomass. The model predicted the experimental results of the calibration set and validation set within an average error value of 12.64% and 14.97%, respectively. To the best of our knowledge, this is the first study to use a mechanistic model for exploring the potential application of capacitance-based biomass estimation. Developing a mechanistic model for a biotherapeutic production process would be beneficial for implementing optimization and control studies.

## 4.2 Problem background

A mathematical model representing the upstream process of a biotherapeutic production is valuable for several purposes. It can provide a detailed understanding of the underlying process and enable the possibility of conducting simulation studies to understand the relationship between various CPPs and CQAs. Furthermore, a validated process model serves as the first step in developing control and optimization studies for deploying appropriate control actions to keep the process under control by maintaining various setpoints for different state variables of the process. However, the mechanistic model must be developed and modified with appropriate kinetic models for the particular case study. Followed by which, the parameters from the respective model equations have to be estimated to achieve a reliable representation of the underlying process.

In this chapter, a mechanistic model for representing the fed-batch fermentation process of a biotherapeutic protein production from recombinant *E. coli* was developed. The production of Ranibizumab as the antibody fragment was considered as the case study, and the data obtained from the reactor experiments were used for calibrating and validating the developed model. For representing the bacterial growth, a dual-substrate Monod-type model was implemented for this case study with residual oxygen concentration as a limiting substrate along with the carbon source. Furthermore, the biomass concentration estimated from the capacitance measurements was integrated with the developed model for the model calibration and validation. The validated model from

this study was further considered for the development of optimization and control studies for achieving enhanced production of the therapeutic product.

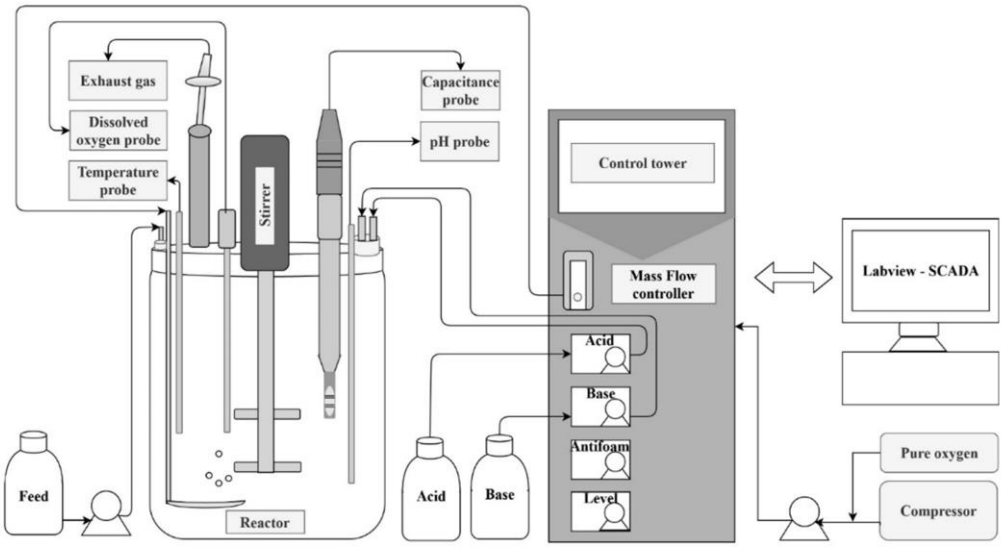
## 4.3 Methodology

### 4.3.1 Experimental studies

The fed-batch experiments for the production of the therapeutic protein Ranibizumab were carried out as explained in chapter 3 (section 3.3.1 and 3.3.2). The reactor studies were carried out for 12 h approximately, and the reactor operation was designated into two phases: batch and fed-batch. The recombinant *E. coli* culture was inoculated in SOC medium with 2% (w/v) glucose as the carbon source during the batch phase, followed by a constant feeding of 20% (v/v) glycerol in the fed-batch phase, as explained in detail in section 3.3.2. The reactor was equipped with PAT tools like pH and temperature probe, capacitance probe, DO probe and exhaust gas analyzer to measure the respective state variables. The real-time measurements for the state variables were logged using SCADA, as explained in section 3.3.1.

The reactor studies were conducted using four input variables, namely, air flow rate [$air_{in}$ (L $h^{-1}$)], pure oxygen added [$O_{2,in}$ (%)], stirrer speed [$N_i$ ($h^{-1}$)], and substrate flow rate during the fed-batch phase [$F$ (L $h^{-1}$)]. The measured capacitance data were used to estimate the online biomass concentration [$X$ (g $L^{-1}$)]. The online DO concentration [$DO$ (%)], and oxygen in the gas phase [$O_2$ (%)] were measured using the DO probe and exhaust gas analyzer, respectively. Offline analysis was performed to estimate the substrate concentrations in batch [$S_b$ (g $L^{-1}$)] and fed-batch phase [$S_f$ (g $L^{-1}$)], product concentration [$P$ (g $L^{-1}$)], and biomass concentration [$X$ (g $L^{-1}$)], which were summarized in sections 3.3.4 and 3.3.5 of chapter 3. Different sets of reactor experiments with similar input conditions and a constant feeding during the fed-batch phase were performed, and the datasets were further used for the calibration and validation steps of the model development, as presented in this chapter. The experimental setup for conducting the reactor studies and a summary of the measured state variables are presented in Figure 4.1.

The reactor studies presented in chapter 4 are the same as presented in chapter 3. In chapter 3, the capacitance data (frequency scanning) from reactor studies was utilized for developing biomass estimation models. In chapter 4, the SCADA data and offline data from reactor studies were used for fermentation model development.




**Duration**:
- Batch phase (from time 0 to batch time $t_b$)
- Fed-batch phase (from batch time $t_b$ to final harvest time $t_f$)
- Production phase (from induction time $t_i$ to final harvest time $t_f$)

| Input variables | Measured state variables | |
|---|---|---|
| | Online | Offline |
| • Air flow rate, $air_{in}$ (L $h^{-1}$)<br>• Pure oxygen added, $O_{2,in}$ (%)<br>• Stirrer speed, $N_i$ ($h^{-1}$)<br>• Substrate flow rate during fed-batch phase, F (L $h^{-1}$) | • Biomass concentration from Capacitance measurements, X (g $L^{-1}$)<br>• Dissolved oxygen concentration, DO (%)<br>• Oxygen in gas phase from exhaust gas measurements, $O_2$ (%) | • Substrate concentrations, $S_b$ and $S_f$ (g $L^{-1}$) for batch and fed-batch phase respectively<br>• Product concentration, P (g $L^{-1}$)<br>• Biomass concentration from Dry cell weight (DCW) analysis, X (g $L^{-1}$) |

Figure 4.1. Experimental setup for reactor studies with a summary of reactor duration, input variables and measured state variables

### 4.3.2 Model development for biotherapeutic production from *E. coli*

The mechanistic model for a fed-batch process was developed by formulating unsteady-state material balances for the state variables across the reactor system (Shuler and Kargi, 2002). The duration of reactor operation was separated into batch phase (0 - $t_b$) and fed-batch phase ($t_b$ - $t_f$) according to the time interval of operation, where $t_b$ and $t_f$ represent batch time and final harvest time, respectively. Accordingly, mass balance equations were developed, parameters were defined separately for the two different phases, and the mass balance for the protein production was applied after the time of induction, $t_i$. The air flow

rate [$air_{in}$ (L $h^{-1}$)], pure oxygen added [$O_{2,in}$ (%)], stirrer speed [$N_i$ ($h^{-1}$)], and substrate flow rate during the fed-batch phase [$F$ (L $h^{-1}$)], were the input variables defined in the model equations. The state variables defined were biomass concentration [$X$ (g $L^{-1}$)], substrate concentrations, for the batch [$S_b$ (g $L^{-1}$)] and fed-batch phase [$S_f$ (g $L^{-1}$)], DO concentration [$DO$ (%)], oxygen in the gas phase [$O_2$ (%)] and product concentration [$P$ (g $L^{-1}$)]. The substrate mass balance for $S_b$ was active during the batch phase (0 - $t_b$), and the mass balance for $S_f$ was active during the fed-batch phase ($t_b$ - $t_f$), where $S_b$ and $S_f$ represent the concentration of the two substrates, glucose and glycerol, in the batch and fed-batch phase, respectively (Ko and Wang, 2007). Equations 4.1–4.7 represent the model equations developed based on the mass balances.

Biomass mass balance:

$$\frac{dX}{dt} = \mu_b X + \mu_f X + \frac{F}{V}\left(X_{in} - X\right) \tag{4.1}$$

Substrate mass balance ($S_b$) (0–$t_b$):

$$\frac{dS_b}{dt} = -q_{S_b} X \tag{4.2}$$

Substrate mass balance ($S_f$) ($t_b$–$t_f$):

$$\frac{dS_f}{dt} = \frac{F}{V}\left(S_{f,in} - S_f\right) - q_{S_f} X \tag{4.3}$$

DO mass balance:

$$\frac{dC_L}{dt} = k_L a\left(C_L^* - C_L\right) - q_{O_{2,b}} X - q_{O_{2,f}} X - \frac{F}{V} C_L \tag{4.4}$$

$O_2$ mass balance:

$$\frac{dO_2}{dt} = \frac{air_{in}}{V_g}\left(O_{2,in} - O_2\right) - \left(\frac{\frac{RT}{p}\left(V k_L a\left(C_L^* - C_L\right)\right)}{32 V_g}\right) - \frac{F}{V_g} O_2 \tag{4.5}$$

$$\frac{dV}{dt} = F \tag{4.6}$$

Product mass balance ($t_i$–$t_f$):

$$\frac{dP}{dt} = q_P X - \frac{F}{V} P \quad (4.7)$$

$$V_g = V_T - V \quad (4.8)$$

where $\mu_b$, $\mu_f$, $q_{S_b}$, $q_{S_f}$, $q_{O_{2,b}}$, $q_{O_{2,f}}$ and $q_P$ represent the specific rates of biomass growth ($h^{-1}$), substrate consumption (g $g^{-1}$ $h^{-1}$), oxygen consumption (g $g^{-1}$ $h^{-1}$), and product formation (g $g^{-1}$ $h^{-1}$), respectively, and the subscripts *b* and *f* refer to the batch phase and fed-batch phase, respectively. $X_{in}$ and $S_{f,in}$ refer to the biomass and substrate concentration in the feed (g $L^{-1}$), respectively. According to the defined equation 4.3 and as observed in our experiments, the batch–phase substrate was completely exhausted at the end of the batch, and thus $S_b$ was expected to be zero in the fed-batch phase, thereby ignoring any residual substrate term for $S_b$ in the mass balance equation (equation 4.3). Subsequently, in the fed-batch phase, only the substrate balance for $S_f$ was assumed to be active according to the model. The mass balance for biomass was defined based on the assumption that the substrate was used only for biomass growth and maintenance. The product mass balance was assumed to be zero until $t_i$ and assumed to be following the equation 4.7 from $t_i$ to $t_f$.

It was assumed that oxygen was well mixed in the reactor, and the oxygen bubbles in the liquid phase were of negligible volume. The total volume of the reactor is $V_T$ (L), which comprises broth volume [$V$ (L)] and volume of the gas above the broth [$V_g$ (L)] as represented by the equation 4.8. $C_L$ is the concentration of *DO* (g $L^{-1}$), $C_L$* is the saturation concentration of *DO* (8 mg $L^{-1}$ at standard temperature and pressure), and $k_La$ ($h^{-1}$) is the volumetric mass transfer coefficient. Other terms related to the gas law were ideal gas constant [$R$ ($K^{-1}$ kPa L $mol^{-1}$)], pressure [$p$ (kPa)], and temperature [$T$ (K)]. Equations related to *DO* conversion and $k_La$ estimation for both phases are explained below.

In order to convert the *DO %* to $C_L$ (g $L^{-1}$), $C_L$* was calculated based on Henry's law, and the following conversion equations (4.9, 4.10) were applied.

$$C_L^* = \frac{O_{2,in} p}{H} \quad (4.9)$$

$$DO = \frac{C_L}{C_L^*} 100 \quad (4.10)$$

Where, $C_L$* is the saturation concentration of *DO* (8 mg $L^{-1}$ at standard temperature and pressure) and *H* is Henry's constant (L kPa $g^{-1}$) (Sander, 2015). Additionally, the following equations (4.11–4.13) were applied to estimate the volumetric mass transfer coefficient $k_La$ ($h^{-1}$) in case of changing stirrer speed and inlet air flow rate. Equation 4.11 represents the most successful correlation for determining the mass transfer coefficient in the reactor. However, since oxygen transfer is strongly influenced by additives in the media, the mass transfer situation can become complex. Additionally, apart from the empirical equation such as equation 4.11, experimental methods could also be used for $k_La$ estimation (Doran, 2012).

$$k_L a = A\left(\frac{P_T}{V}\right)^B \left(u_g\right)^C \tag{4.11}$$

$$P_T = N_p \rho N_i^3 D_i^5 \tag{4.12}$$

$$u_g = \frac{air_{in}}{area} \tag{4.13}$$

where *A*, *B* and *C* are the correlation coefficients for $k_La$ estimation (Doran, 2012). The total power consumption [$P_T$ (watt)] and superficial gas velocity [$u_g$ (m $h^{-1}$)] in the $k_La$ correlation (equation 4.11) were further defined using the fermenter and fluid characteristics, including the cross-sectional area [*area* ($m^2$)], impeller diameter [$D_i$ (m)], stirrer speed [$N_i$ ($h^{-1}$)], power number [$N_p$ (-)] and density [$\rho$ (kg $m^{-3}$)] in equations 4.12 and 4.13.

The specific growth rate $(\mu_b, \mu_f)$ for the biomass was assumed to follow the Monod-type dual substrate kinetics, considering limiting substrates, viz., $S_b$ or $S_f$ and $C_L$ (Kornaros and Lyberatos, 1997). This dual substrate kinetic model was adapted after observing the effect of oxygen-related inputs on the biomass concentration. The specific product formation rate for the intracellular protein was described using the Luedeking–Piret equation (Jenzsch et al., 2006c; Luedeking and Piret, 2000).

The equations for the specific growth rates are given in equations 4.14 and 4.15. The specific rates, $q_S$, $q_{O_2}$ and $q_P$ for the two phases are represented in equations 4.16-4.20

$$\mu_b = \mu_{m_b}\left(\frac{S_b}{k_{S_b} + S_b}\right)\left(\frac{C_L}{k_{O_2} + C_L}\right) \tag{4.14}$$

$$\mu_f = \mu_{m_f}\left(\frac{S_f}{k_{S_f}+S_f}\right)\left(\frac{C_L}{k_{O_2}+C_L}\right) \tag{4.15}$$

$$q_{S_b} = \frac{\mu_b}{y_{X/S_b}} + m_{S_b} \tag{4.16}$$

$$q_{O_{2,b}} = \frac{\mu_b}{y_{X/O_{2,b}}} + m_{O_2} \tag{4.17}$$

$$q_{S_f} = \frac{\mu_f}{y_{X/S_f}} + m_{S_f} \tag{4.18}$$

$$q_{O_{2,f}} = \frac{\mu_f}{y_{X/O_{2,f}}} + m_{O_2} \tag{4.19}$$

$$q_P = \alpha\mu + \beta \tag{4.20}$$

where, $\mu_m$, $k_S$ and $k_{O_2}$ are maximum specific growth rate ($h^{-1}$), substrate saturation constant (g $L^{-1}$) and oxygen saturation constant (g $L^{-1}$), respectively. The specific substrate and oxygen consumption rates were given as a function of yield coefficient, $y_{X/S}$, $y_{X/O_2}$ (g $g^{-1}$) and maintenance coefficient $m_S$, $m_{O_2}$ (g $g^{-1}$ $h^{-1}$). The specific product formation rate for the intracellular protein was described using the Luedeking–Piret equation where α and $\beta$ are the growth (g $g^{-1}$) and non-growth associated product yield coefficients (g $g^{-1}$ $h^{-1}$), respectively.

All the model equations (as summarized in Table 4.1) were numerically solved by implementing them in the Modelica language and DassL solver available in the DYMOLA®2019 tool. The initial conditions for the state variables were given at time t = 0. The developed model has 7 state variables and 17 parameters, as summarized in Table 4.2. These parameters are expected to vary with respect to the reactor dimension, input conditions, and the kinetics of the system under study.

Table 4.1. Summary of model equations used in this study

| **Equations** | **Batch phase (0–$t_b$)** | **Fed-batch phase ($t_b$–$t_f$)** |
|---|---|---|
| Volume balance | $\frac{dV}{dt} = F \quad V_g = V_T - V$ | |
| Biomass mass balance (X) | $\frac{dX}{dt} = \mu_b X + \frac{F}{V}\left(X_{in} - X\right)$ | $\frac{dX}{dt} = \mu_f X + \frac{F}{V}\left(X_{in} - X\right)$ |
| Substrate mass balance ($S_b$) | $\frac{dS_b}{dt} = -q_{S_b} X$ | NA |
| Substrate mass balance ($S_f$) | NA | $\frac{dS_f}{dt} = \frac{F}{V}\left(S_{f,in} - S_f\right) - q_{S_f} X$ |
| DO mass balance ($C_L$) | $\frac{dC_L}{dt} = k_L a\left(C_L^* - C_L\right) - q_{O_{2,b}} X - \frac{F}{V} C_L$ | $\frac{dC_L}{dt} = k_L a\left(C_L^* - C_L\right) - q_{O_{2,f}} X - \frac{F}{V} C_L$ |
| $O_2$ mass balance ($O_2$) | $\frac{dO_2}{dt} = \frac{Air_{in}}{V_g}\left(O_{2,in} - O_2\right) - \left(\frac{\frac{RT}{p}\left(Vk_L a\left(C_L^* - C_L\right)\right)}{32V_g}\right) - \frac{F}{V_g} O_2$ | |
| Product mass balance ($t_i$–$t_f$) ($P$) | NA | $\frac{dP}{dt} = q_P X - \frac{F}{V} P$ |

| | |
|---|---|
| Equations for specific rates estimation | $\mu_b = \mu_{m_b}\left(\frac{S_b}{k_{S_b}+S_b}\right)\left(\frac{C_L}{k_{O_2}+C_L}\right)$ $\mu_f = \mu_{m_f}\left(\frac{S_f}{k_{S_f}+S_f}\right)\left(\frac{C_L}{k_{O_2}+C_L}\right)$ $q_{S_b} = \frac{\mu_b}{y_{X/S_b}} + m_{S_b}$ $q_{S_f} = \frac{\mu_b}{y_{X/S_f}} + m_{S_f}$ $q_{O_{2,b}} = \frac{\mu_b}{y_{X/O_{2,b}}} + m_{O_2}$ $q_{O_{2,f}} = \frac{\mu_f}{y_{X/O_{2,f}}} + m_{O_2}$ $q_P = \alpha\mu + \beta$ |
| Equations for DO conversion and $k_La$ estimation | $C_L^* = \frac{O_{2,in}p}{H}$ $DO = \frac{C_L}{C_L^*}100$ $k_La = A\left(\frac{P_T}{V}\right)^B\left(u_g\right)^C$ $P_T = N_p\rho N_i^3 D_i^5$ $u_g = \frac{air_{in}}{area}$ |

Table 4.2. Summary of state variables, input variables and parameters of the developed model and start values

| State variables | | | Input variables | Parameters | |
|---|---|---|---|---|---|
| **Batch** | **Fed batch** | **Start values (*t*=0)** | | | |
| $V$ | $V$ | 1.5 L | $air_{in}$ | $k_{O_2}$ | $\mu_{m_f}$ |
| $X$ | $X$ | 2.07 g $L^{-1}$ | $O_{2,in}$ | $\mu_{m_b}$ | $k_{S_f}$ |
| $S_1$ | - | 20 g $L^{-1}$ | $F$ | $k_{S_b}$ | $y_{X/S_f}$ |
| - | $S_2$ | 0 | $N_i$ | $y_{X/S_b}$ | $y_{X/O_{2,f}}$ |
| $C_L$ | $C_L$ | 0.008 g $L^{-1}$ | | $y_{X/O_{2,b}}$ | $m_{S_f}$ |
| $O_2$ | $O_2$ | 19.6 % | | $m_{S_b}$ | $\alpha$ |
| | $P$ | 0 | | $m_{O_2}$ | $\beta$ |
| | | | | $A, B, C$ | |

### 4.3.3 Model calibration and validation

#### 4.3.3.1 Parameter sensitivity analysis

Model calibration was performed by estimating various model parameters, as listed in Table 4.2. The optimization algorithm used for parameter estimation required bounds for the parameters that have to be tuned, and thus, a model sensitivity analysis was performed. The sensitivity analysis of the model output was studied by perturbing each parameter by ± 10% from the literature reported values, while other parameters were maintained at their nominal values (Jenzsch et al., 2006a; Lopes et al., 2014). The corresponding changes in the concentrations of the state variables were observed, and thereby, the sensitivity of the parameters was identified.

#### 4.3.3.2 Parameter estimation and model validation

Parameter estimation was done by using the DYMOLA's in-built design library known as the model calibration tool. This design library enables us to formulate an error function that must be minimized to match the experimental and model-predicted results by

manipulating the model parameters, as shown by the equation 4.21. The data flow between the model calibration tool and model equations and the overall procedure for parameter estimation are presented in Figure 4.2. The dynamic data of the input variables $air_{in}$, $O_{2,in}$, $N_i$ and $F$ were integrated with the calibration tool using the data file extracted from SCADA. Other model constants such as reactor dimensions and basic fluid and gas properties (*area*, $N_p$, $D_i$, $\rho$, $p$, $R$, and $T$) were coded within the model. The calibration tool used the genetic algorithm (GA) as the optimization technique to minimize the error with a tolerance of 0.001. The bounds for the parameters were used from the sensitivity analysis, as summarized in Table 4.3.

$$Error = \sum \begin{bmatrix} \left(\left|X_{\exp} - X_{\mathrm{mod}}\right|\right) + \left(\left|S_{b,\exp} - S_{b,\mathrm{mod}}\right|\right) + \left(\left|S_{f,\exp} - S_{f,\mathrm{mod}}\right|\right) + \\ \left(\left|P_{\exp} - P_{\mathrm{mod}}\right|\right) + \left(\left|DO_{\exp} - DO_{\mathrm{mod}}\right|\right) + \left(\left|O_{2,\exp} - O_{2,\mathrm{mod}}\right|\right) \end{bmatrix} \quad (4.21)$$

where $X$, $S_b$, $S_f$, $P$, $DO$, and $O_2$ represent the state variables, and the subscripts *exp* and *mod* represent the experimental and model-predicted values of the respective state variables. In equation 4.21, the different state variables were not scaled to be of equal size and the error terms were defined to have equal weights in the objective function.

The calibrated model was then subjected to validation using another set of experimental data with similar input conditions. Model accuracy was analyzed by estimating the error between the experimental and model-predicted values. The calibrated model is applicable for various operating conditions, provided that the input variables are within their respective constraints and physical boundaries. Varying the input variables within their respective limits results in various combinations of operating conditions within which the calibrated model is applicable.

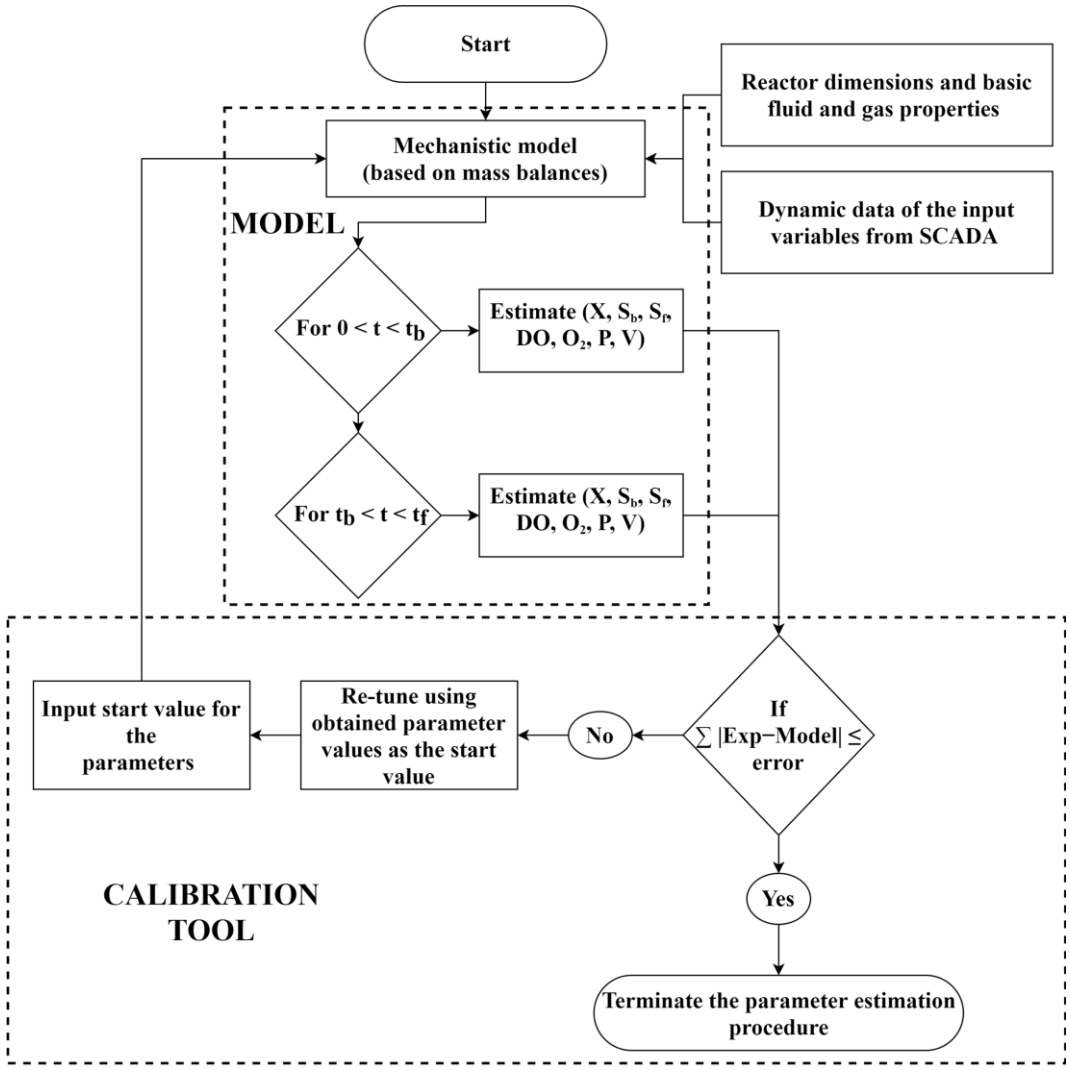



Figure 4.2. Flowchart for parameter estimation

Table 4.3. Boundary conditions for the parameters considered for model calibration

| Parameter | Unit | Lower bound | Upper bound |
|---|---|---|---|
| $k_{O_2}$ | g $L^{-1}$ | 0 | 0.003 |
| $\mu_{m_b}, \mu_{m_f}$ | $h^{-1}$ | 0 | 0.8 |
| $k_{S_b}$ | g $L^{-1}$ | 0 | 0.1 |
| $k_{S_f}$ | g $L^{-1}$ | 0 | 0.01 |
| $y_{X/S_b}, y_{X/S_f}, y_{X/O_{2,b}}, y_{X/O_{2,f}}$ | g $g^{-1}$ | 0 | 1 |
| $\alpha$ | g $g^{-1}$ | 0 | 0.002 |

| | | | |
|---|---|---|---|
| $\beta$ | g $g^{-1}$ $h^{-1}$ | 0 | 0.002 |
| $m_{S_b}$, $m_{S_f}$,$m_{O_2}$ | g $g^{-1}$ $h^{-1}$ | 0 | 1 |
| *A, B, C* | (-) | 0 | 1 |

## 4.4 Results and discussion

### 4.4.1 Model calibration and validation

#### 4.4.1.1 Parameter sensitivity analysis

Sensitivities of different parameters with respect to all the state variables are tabulated in Table 4.4. The sensitivity analysis was performed to understand the effect of different parameters on the state variables and estimate the range of the parameters for model tuning, that is, lower bound and upper bound (Shiue et al., 1995; Zhong and Tang, 2004). The result of positive perturbation in all the parameters concluded that the oxygen yield coefficient in the batch phase was the most sensitive parameter for biomass concentration compared to all other parameters. Similarly, the negative perturbation study indicated that the maintenance coefficient of oxygen is found to be the most sensitive parameter for biomass concentration. This observation, in turn, implies that the oxygen-related parameters are sensitive towards biomass concentration. In the case of substrate concentration, biomass yield coefficients are found to be more sensitive. A similar analysis for other state variables can be established, as shown in Table 4.4.

Table 4.4. Sensitivity ranking of the parameters for the different state variables of the developed model

| **State variable** | **Rank of parameters in the order of sensitivity** | |
|---|---|---|
| | **+10% perturbation in parameters** | **-10% perturbation in parameters** |
| *X* | $y_{X/O_{2,b}} > y_{X/O_{2,f}} > \mu_{m_b} > \mu_{m_f} > y_{X/S_f} > y_{X/S_b} > m_{S_b} > \beta > \alpha > k_{S_f} > k_{S_b} > m_{S_f} > k_{O_2} >$ *(A, B, C)* $> m_{O_2}$ | $m_{O_2} > k_{O_2} >$ *(A, B, C)* $> m_{S_f} > k_{S_b} > m_{S_b} > k_{S_f} > \beta > \alpha > y_{X/S_b} > \mu_{m_f} > y_{X/S_f} > \mu_{m_b} > y_{X/O_{2,f}} > y_{X/O_{2,b}}$ |
| $S_b$ | $y_{X/S_b} >$ *(A, B, C)* $> k_{O_2} > m_{O_2} > k_{S_b} > m_{S_f} > \beta > \alpha > y_{X/O_{2,f}} > y_{X/S_f} > k_{S_f} > \mu_{m_f} > \mu_{m_b} > m_{S_b} > y_{X/O_{2,b}}$ | $y_{X/O_{2,b}} > \mu_{m_b} > m_{S_b} > m_{S_f} > \beta > \alpha > y_{X/O_{2,f}} > y_{X/S_f} > k_{S_f} > \mu_{m_f} > k_{S_b} > m_{O_2} > k_{O_2} >$ *(A, B, C)* $> y_{X/S_b}$ |

| | | |
|---|---|---|
| $S_f$ | $y_{X/S_f} > m_{O_2} > (A, B, C) > k_{O_2} > y_{X/S_b} > k_{S_b} > k_{S_f} > \beta > \alpha > m_{S_b} > \mu_{m_f} > \mu_{m_b} > y_{X/O_{2,b}} > m_{S_f} > y_{X/O_{2,f}}$ | $y_{X/O_{2,f}} > m_{S_f} > y_{X/O_{2,b}} > \mu_{m_b} > \mu_{m_f} > m_{S_b} > k_{S_b} > \beta > \alpha > k_{S_f} > y_{X/S_b} > (A, B, C) > k_{O_2} > m_{O_2} > y_{X/S_f}$ |
| $P$ | $\beta > \alpha > y_{X/S_f} > \mu_{m_b} > \mu_{m_f} > y_{X/O_{2,b}} > y_{X/O_{2,f}} > y_{X/S_b} > k_{S_f} > k_{S_b} > m_{S_b} > k_{O_2} > m_{O_2} > (A, B, C) > m_{S_f}$ | $m_{S_f} > m_{O_2} > k_{O_2} > k_{S_f} > k_{S_b} > m_{S_b} > y_{X/S_b} > \mu_{m_f} > \mu_{m_b} > y_{X/O_{2,b}} > \alpha > (A, B, C) > y_{X/S_f} > \beta > y_{X/O_{2,f}}$ |
| $DO$ | $y_{X/O_{2,f}} > m_{S_f} > y_{X/O_{2,b}} > (A, B, C) > \mu_{m_b} > \mu_{m_f} > m_{S_b} > y_{X/S_b} > k_{S_b} > k_{S_f} > \beta > \alpha > k_{O_2} > y_{X/S_f} > m_{O_2}$ | $y_{X/S_f} > m_{O_2} > (A, B, C) > k_{O_2} > y_{X/S_b} > k_{S_b} > m_{S_b} > \beta > \alpha > k_{S_f} > \mu_{m_b} > \mu_{m_f} > m_{S_f} > y_{X/O_{2,b}} > y_{X/O_{2,f}}$ |
| $O_2$ | $y_{X/O_{2,f}} > y_{X/O_{2,b}} > m_{S_f} > (A, B, C) > \mu_{m_f} > k_{O_2} > m_{S_b} > y_{X/S_b} > k_{S_f} > \beta > \alpha > k_{S_b} > \mu_{m_b} > y_{X/S_f} > m_{O_2}$ | $y_{X/S_f} > m_{O_2} > \mu_{m_b} > (A, B, C) > y_{X/S_b} > k_{S_b} > m_{S_b} > k_{O_2} > k_{S_f} > \beta > \alpha > \mu_{m_f} > m_{S_f} > y_{X/O_{2,f}} > y_{X/O_{2,b}}$ |

#### 4.4.1.2 Parameter estimation and model validation

Model calibration was performed by tuning the parameters within the defined bounds, as mentioned in section 4.3.3.2. The comparison of experimental and model-predicted values is presented in Figure 4.3. It is observed that the model is able to predict the respective experimental values with an average error of 12.64%. Figure 4.3 (A) shows the variation in the input variables with respect to time, and the same was used as an input for the model. Also, it could be noticed that a constant substrate flow rate was adapted for this experimental study. The residual substrate concentration ($S_f$) of the process was maintained near zero, as observed in Figure 4.3 (D). The offline analysis of biomass, substrate, and product concentration is represented as dotted points, as shown in Figure 4.3 (C) and (D). These data were used for the model calibration and validation studies. The values were plotted as an average of two reactor runs.

The comparison of experimental and model-predicted values for *DO* and $O_2$ is shown in Figure 4.3 B. The model predicted the experimental *DO* and $O_2$ within an error value of 28.21% and 3.41%, respectively. The comparison of experimental and model-predicted values for biomass, substrate, and product concentration is shown in Figure 4.3 (C) and (D). The model predicted the experimental results within error values of 11.97%, 7.23%, 15.81%, and 9.20% for *X*, $S_b$, $S_f$ and *P*, respectively. It can be concluded that the parameter

estimation procedure has to ensure that all the mass balances are satisfied and the parameters are within their bounds in order to obtain a reliable calibrated model that could be applied further for optimization strategy implementation.

The estimated values for the parameters are summarized in Table 4.5. The parameters were constrained to be positive and also fall within the defined bounds. Following the model calibration, validation was carried out, and the results are plotted in Figure 4.4. The model predicted the experimental values of the validation dataset with an average error value of 14.97%.

The reactor studies carried out for model development and validation were open-loop experiments with constant feed rates throughout the fed-batch phase. The input signals can be subjected to disturbance/excitation (5 or 10%) to verify the model predictions. The initial parameter values for the search were chosen randomly to fall within the bounds presented in Table 4.3. Start values for some parameters like yield coefficients were chosen to be roughly near the values calculated from offline analysis.

Compared to the model predictions using initial parameter values, the calibrated parameters resulted in improvements greater than 50%. The error value for validation set was ~14%, which is close enough to the error value of model calibration (~12%). This shows that the model is reliable and valid within the given range of input parameters. Further, this error could be reduced by one of the following means: a. Choosing different kinetic models for biomass estimation; b. Smoothening/ filtering of experimental DO data, since the error value of DO was highest among the different state variables, owing to high signal noise; c. Acquiring more experimental data points (especially offline analysis) and implementing online estimation methods.

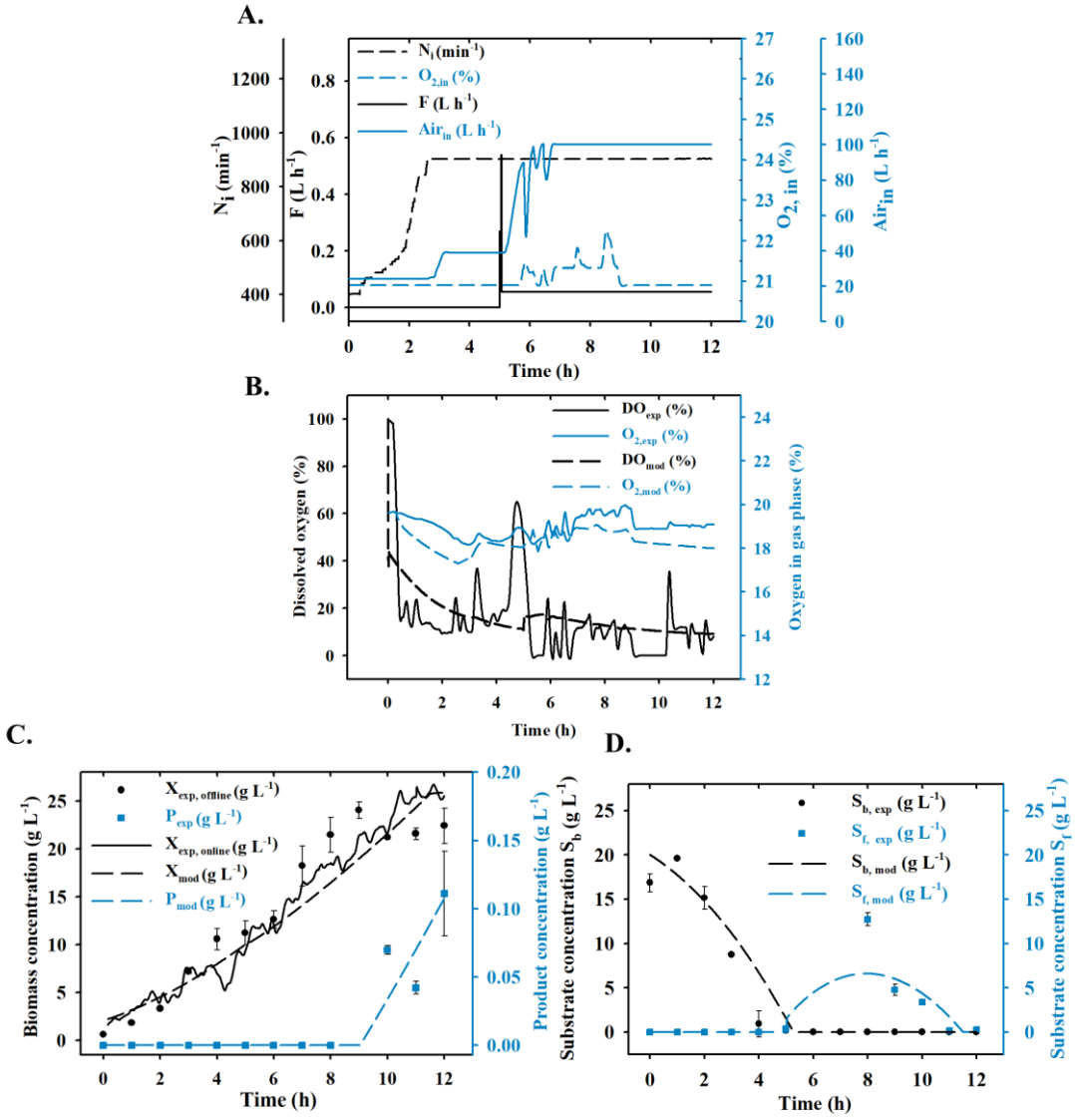



Figure 4.3. Model calibration and parameter estimation: (A). Input process variables; (B). Experimental (continuous lines) and model-predicted (dashed lines) comparison for dissolved oxygen and gas phase oxygen; (C). Experimental and model-predicted (dashed lines) comparison for biomass concentration (online (continuous line) & offline (black filled circle)) and product concentration (blue filled circle); (D). Experimental and model-predicted (dashed lines) comparison for substrate concentration in batch (black filled circle) and fed-batch phase (blue filled circle).

Table 4.5. Estimated values for the parameters after model calibration

| Parameter symbol | Parameter name | Value |
|---|---|---|
| $k_{O_2}$ | Monod saturation constant for oxygen (g $L^{-1}$) | 0.002 |
| $\mu_{m_b}$ | Maximum specific growth rate of batch phase ($h^{-1}$) | 0.750 |
| $\mu_{m_f}$ | Maximum specific growth rate of fed-batch phase ($h^{-1}$) | 0.555 |
| $k_{S_b}$ | Monod saturation constant for substrate in batch phase (g $L^{-1}$) | 0.075 |
| $k_{S_f}$ | Monod saturation constant for substrate in fed-batch phase (g $L^{-1}$) | 0.008 |
| $y_{X/S_b}$ | Biomass yield coefficient from substrate in batch phase (g biomass $g^{-1}$ substrate $S_b$) | 0.666 |
| $y_{X/S_f}$ | Biomass yield coefficient from substrate in fed-batch phase (g biomass $g^{-1}$ substrate $S_f$) | 0.699 |
| $y_{X/O_{2,b}}$ | Biomass yield coefficient from oxygen utilized in batch phase (g oxygen $g^{-1}$ biomass) | 0.931 |
| $y_{X/O_{2,f}}$ | Biomass yield coefficient from oxygen utilized in fed-batch phase (g oxygen $g^{-1}$ biomass) | 0.976 |
| $\alpha$ | Growth associated product yield coefficient (g product $g^{-1}$ biomass) | 0.010 |
| $\beta$ | Non-growth associated product yield coefficient (g product $g^{-1}$ biomass $h^{-1}$) | 0.0002 |
| $m_{S_b}$ | Maintenance coefficient of substrate in batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) | 0.235 |
| $m_{S_f}$ | Maintenance coefficient of substrate in fed-batch phase (g substrate $g^{-1}$ biomass $h^{-1}$) | 0.227 |
| $m_{O_2}$ | Maintenance coefficient of oxygen (g oxygen $g^{-1}$ biomass $h^{-1}$) | 0.069 |
| $A$<br>$B$<br>$C$ | Correlation constants for volumetric mass transfer coefficient (-) | 0.451<br>0.013<br>0.195 |

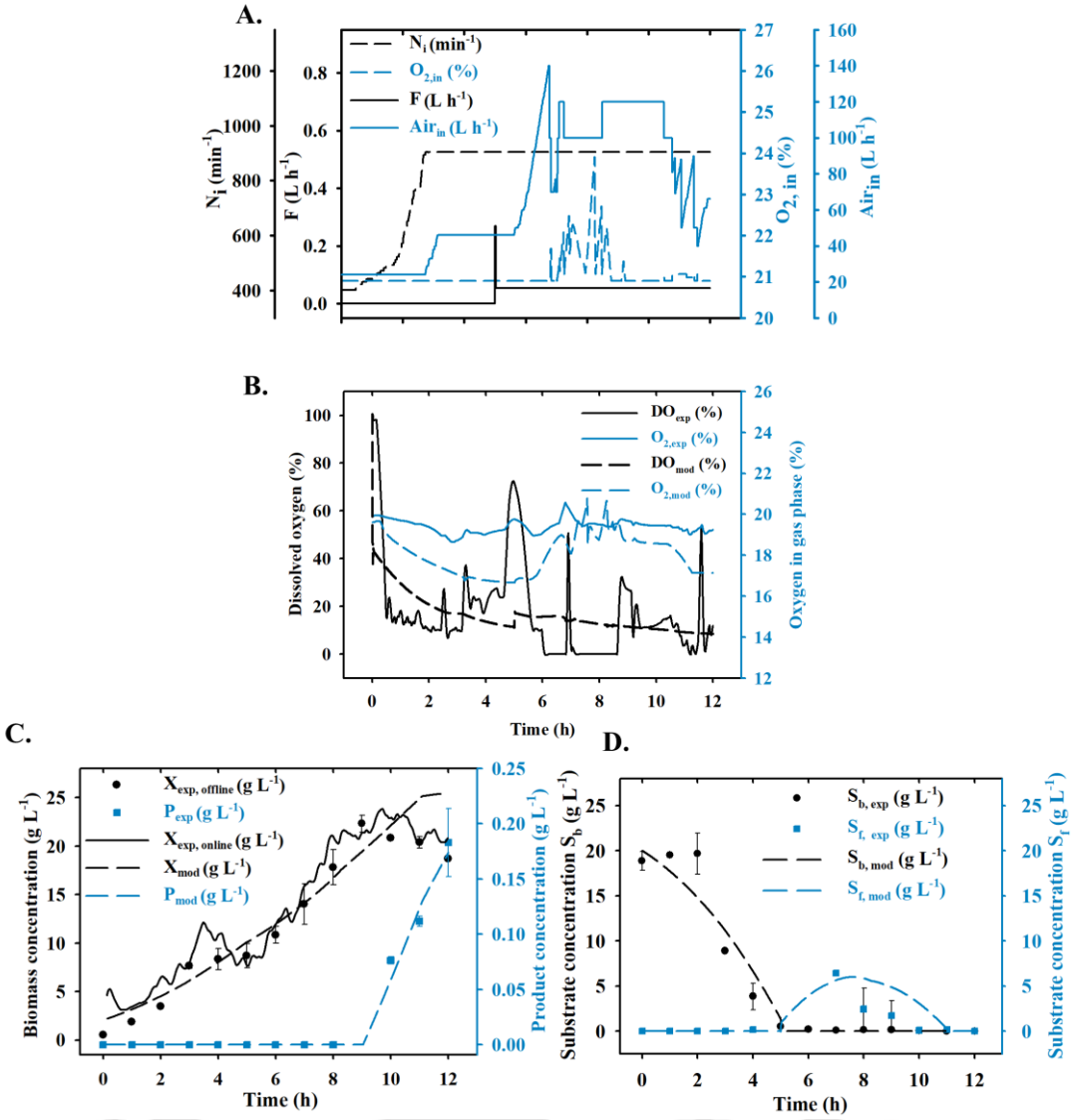



Figure 4.4. Model validation: (A). Input process variables; (B). Experimental (continuous lines) and model-predicted (dashed lines) comparison for dissolved oxygen and gas phase oxygen; (C). Experimental and model-predicted (dashed lines) comparison for biomass concentration (online (continuous line) & offline (black filled circle)) and product concentration (blue filled circle); (D). Experimental and model-predicted (dashed lines) comparison for substrate concentration in batch (black filled circle) and fed-batch phase (blue filled circle).

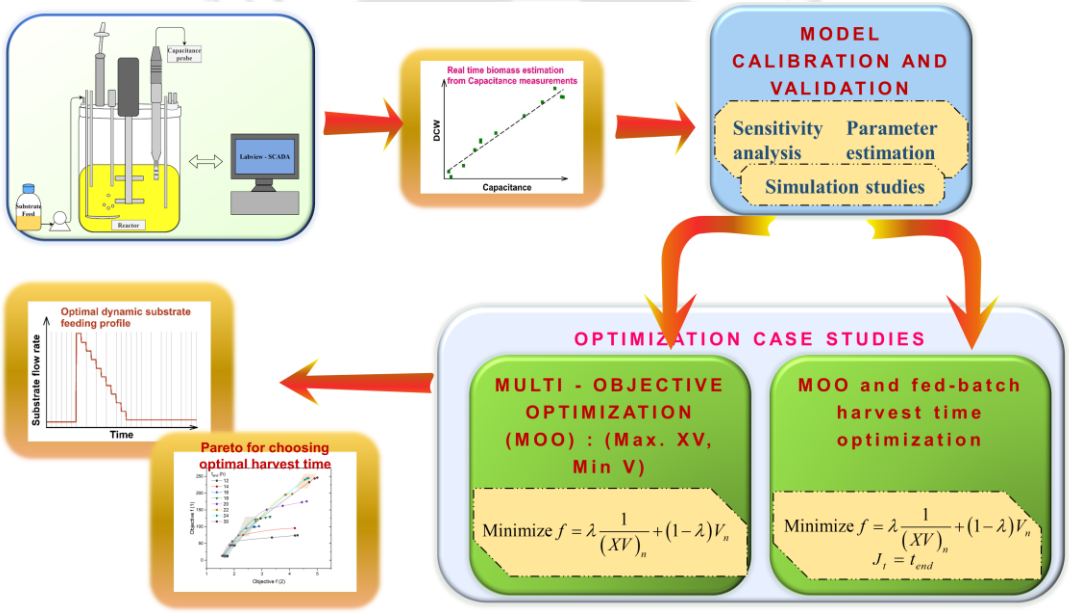
## 4.5 Summary

A mathematical model was developed, and the model was validated for similar experimental conditions. The model predicted the experimental results of the calibration set and validation set within an average error value of 12.64% and 14.97%, respectively.

Sensitivity analysis was carried out on the defined parameters to understand and investigate their influence on the respective state variables. The significant effect of oxygen-related parameters and the aeration-related inputs on biomass growth rate was observed from the simulation studies and sensitivity analysis. Consequently, a dual-substrate Monod-type kinetic model including residual oxygen concentration as an additional limiting substrate was implemented to represent the specific growth rate of the organism. Further, this study presented the mathematical model for batch, and fed-batch phases of the fermentation and the respective parameters from both phases were estimated. The validated mathematical model obtained from this study represents the fed-batch cultivation of recombinant *E. coli* expressing the biotherapeutic protein, Ranibizumab, and can further be extended for implementing optimization and control studies.

# CHAPTER 5

# Optimization studies for maximizing biomass production and predicting harvest time




---

Parts of this chapter is based on the following research article(s):

**Swaminathan, N.,** Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2020. Multiobjective Optimization for Enhanced Production of Therapeutic Proteins in Escherichia coli: Application of Real-Time Dielectric Spectroscopy. Ind. Eng. Chem. Res. 59, 21841–21853. https://doi.org/10.1021/acs.iecr.0c04010.

## 5.1 Foreword

Investigating the usage of optimization studies to acquire optimal inputs for fermentation could enhance the productivity of the product of interest. In this study, two case studies were proposed to target relevant objectives that can improve the therapeutic production process. Case study (1) formulates a multiobjective optimization (MOO) with respect to the conflicting objectives of maximizing the total biomass ($XV$) and minimizing the broth volume ($V$). A validated mechanistic model was employed to formulate a multiobjective optimization (MOO) problem. This study focuses on the optimization of substrate feeding profile for enhanced production of Ranibizumab, a recombinant therapeutic protein, in *Escherichia coli*. The substrate flow rate during the fed-batch phase ($F$) was taken as the decision variable for the MOO. The Pareto front resulting from the MOO revealed that for a minimum broth volume ($V$) of 1.96 L, a maximum of 58.8 g of total biomass ($XV$) could be generated. The total biomass obtained from the optimal substrate feeding profile was 20.6% higher than the experimentally achieved total biomass. Different simulation studies have been performed to demonstrate the ability of the developed optimization strategy in effectively dealing with the fault in the actuators observed during fermentation. The optimization results obtained from this case study (1) was extended by including the cost factors, and a sample procedure for selecting optimal point for given cost factors was demonstrated.

Case study (2) focuses on predicting the optimal harvest time of the fed-batch operation. The previously developed MOO was used along with a third objective of optimizing final harvest time $t_f$ ($t_{end}$). Simulation studies were carried out with different $t_{end}$ values to predict the optimal fed-batch harvest time. The Pareto for different $t_{end}$ values were obtained, and the objective functions were compared at different $\lambda$ values. The optimal feeding profiles and fed-batch harvest time can be chosen based on the desired volume of operation.

## 5.2 Problem statement

Implementation of optimization studies targeting the desired objectives can aid in enhancing productivity and thereby enhance the performance of the process. Optimization studies serve as an alternative for rigorous and time-extensive experimentation in order to achieve desired targets. The outcomes of optimization studies can result in enhanced biomass growth and efficient metabolism by directing the fed-batch process to operate

with optimal feeding rate without under or overfeeding the substrate (Ochoa, 2016). The application of various optimization studies is especially relevant in the case of expensive biotherapeutic products since performing the fed-batch cultivation under optimal conditions can enhance the yield of the process, thereby reducing the cost of these life-saving therapeutic products. Additionally, the application of various optimization studies provides a better understanding of the system under investigation and provides greater flexibility to the operator to take real-time process decisions like induction and harvest time for the cultivation. Operating the reactor under optimal conditions enhances the efficiency of the process and can significantly reduce the number of experiments to be conducted.

This chapter demonstrates the implementation of various optimization studies by elucidating different case studies. Two case studies were formulated to focus on different objectives to improve the process performance by focusing on particular targets like biomass maximization and predicting optimal harvest time. Case study (1) focused on developing a multiobjective optimization (MOO) strategy to maximize total biomass at a minimum broth volume of the fermenter. The MOO used in this study could be of significant interest to various applications, compared to the existing cost-based objective functions, which differ from application to application.

Further, the simulation studies that were initially carried out to decide the constraints and the time interval for changing the chosen manipulating variable is explained in this chapter. Also, the necessity of implementing a multiobjective optimization objective was inferred from the simulation studies. Enhanced productivity was achieved by the proposed MOO formulation, which facilitates the choice of any operating point from the Pareto front based on downstream expenses of the therapeutic product. The dynamic parameter estimation and the MOO implemented in this study would be beneficial for upscaling the therapeutic protein production at an industrial scale. Furthermore, case study (2) was carried out to predict the optimal fed-batch harvest time of the fermenter. This chapter thus summarizes the implementation of the different case studies for optimization, which can enhance the production of the therapeutic protein (Ranibizumab) considered in this study.

## 5.3 Methodology

### 5.3.1 Process model and experimental data

Reactor studies were conducted in a 5 L bioreactor for the fed-batch cultivation of recombinant *E. coli BL21 DE3* producing Ranibizumab as an antibody fragment, as explained in chapter 3 (section 3.3.1). The reactor was equipped with various PAT tools, including dielectric spectroscopy, and the real-time process measurements were logged using SCADA developed using LabView as briefed in section 3.3.1. Fed-batch experiments were carried out approximately for 12 h with glucose and glycerol as the carbon source in the batch and fed-batch phase, respectively. More details regarding the strain, media compositions and operating conditions are provided in sections 3.3.1 and 3.3.2 of chapter 3. The different datasets obtained from the experimental studies consisting of online and offline measurements were utilized to validate the mechanistic model developed in chapter 4, as explained in section 4.3.1.

The optimization studies presented in this chapter were carried out using the validated process model presented previously in chapter 4. The model equations were developed for the batch (0 - $t_b$) and fed-batch phases ($t_b$ - $t_f$) separately, as explained in section 4.3.2. The validated process model with the optimized parameter values was used for the simulation studies and for investigating the implementation of the desired optimization objectives. The data for the input process variables, namely, $air_{in}$ (L $h^{-1}$), $O_{2,in}$ (%), $N_i$ ($h^{-1}$), and $F$ (L $h^{-1}$), for the simulation and optimization studies were taken from the fed-batch experiments which were presented in section 3.3.1. The optimization studies were thus demonstrated for the validated process model obtained from the experimental studies of fed-batch cultivation of recombinant *E. coli* expressing Ranibizumab as the therapeutic product.

### 5.3.2 Optimization studies

Optimization is a tool for obtaining optimal operating conditions that could channelize the process in order to satisfy the desired objectives. The objectives for optimization studies could be chosen for the particular process of interest according to the specific target of the operator. In this study, two optimization case studies were formulated focusing on different objectives to cater to the needs relevant to the particular system under investigation. Since the therapeutic product (Ranibizumab) considered in this study

is intracellular, it was deduced that maximizing the total biomass in the reactor would be beneficial to enhance the productivity of the therapeutic product. Additionally, this could be combined with a minimization of broth volume, which would be useful in downstream processing of the product, which was investigated in the case study (1). Another valuable optimization goal apart from these two would be determining optimal harvest time for the fed-batch process, which was explored in the second case study.

### 5.3.3 Case study (1): Optimizing the total biomass at a minimum broth volume

In order to achieve maximum biomass with a minimum broth volume, the validated mechanistic model was subjected to optimization studies. Achieving maximum biomass within the reactor with a minimum broth volume could, in turn, aid in achieving higher productivity and also be beneficial in the downstream steps owing to the less broth volume.

#### 5.3.3.1 Simulation studies

Simulation studies were carried out to understand and analyze the system before formulating an objective function for the implementation of the optimization strategy. For case study (1), the following simulation studies were carried out:

(i) Choosing the MV among the different input variables
(ii) Fix the boundary condition for the chosen MV
(iii) Testing the single objective dynamic optimization
(iv) Choosing optimum time interval for the MV

Simulation studies were carried out to study the effect of different input variables and choose the MV accordingly. Different profiles for the input variables were tested in the validated model, and finally, $F$ was chosen to be the MV for this study. In order to fix the boundary conditions for $F$, the simulation studies were carried out, and the respective values of objective functions $f(1)$ and $f(2)$ were analyzed.

The standard dynamic optimization formulations with a single objective were tested to understand the nature of the two objectives defined in this study. The simulation studies were also conducted to choose the optimum time interval for the MV. The tested cases were as follows:

(i) E1–E6: equal time intervals of 5, 10, 15, 20, 25, and 30, respectively, throughout the batch

(ii) V1–V6: varied time intervals (smaller time interval of 1 or 2 min for the first 1-3 h followed by 10 min intervals for the remaining hours of feeding).

#### 5.3.3.2 Formulation of the objective function

The four input variables, namely, $air_{in}$ (L $h^{-1}$), $O_{2,in}$ (%), $N_i$ ($h^{-1}$), and $F$ (L $h^{-1}$), could also be considered as the MVs. However, based on the simulation studies, the substrate flow rate during the fed-batch phase ($F$) was found to be more sensitive for total biomass production and hence was considered to be the MV for the optimization studies (Dan et al., 2002; Ko and Wang, 2006; Shin and Lim, 2006). Furthermore, the constraints for $F$ were chosen based on the simulation studies (section 5.4.1.1). The two objectives considered in this study are as follows:

(i) maximizing the total biomass ($XV$) at $t_f$:

$$\text{maximize } f(1) = XV \tag{5.1}$$

(ii) minimizing the broth volume ($V$) at $t_f$:

$$\text{minimize } f(2) = V \tag{5.2}$$

Subject to constraints in $F$ (0.01-0.5 L $h^{-1}$): $0.01 \leq F \leq 0.5$

The total biomass ($XV$, g) was obtained by multiplying the biomass concentration $X$ with the corresponding broth volume $V$ (L) at time $t_f$. The bounds for the decision variable and constraints were chosen based on the sensitivity of the decision variable on objective function through simulation studies and experimental limitations. For instance, the volume constraint was fixed according to the maximum reactor volume and minimum working volume required for the reactor operation, such that the results could be validated in the same reactor. Whereas in the case of the substrate flowrate, the validated model was used for simulation studies to observe the reactor performance under the conditions of different flow rates (which could also be alternatively implemented in the experimental setup with the aid of an existing pump). Even after achieving optimal results, it was ensured that none of the bounds was limiting the optimal results.

The objectives *f*(1) and *f*(2) were observed to be of conflicting nature; that is, when the total biomass was subjected to maximization, the respective volume also increased. In the performed simulation studies, where the single objective optimization for *f*(1) and *f*(2) was tested individually, it was observed that either both *f*(1) and *f*(2) were maximized or they were minimized simultaneously (section 5.4.1.1). Therefore, it was concluded that the MOO was necessary to address the two objectives with conflicting nature and achieve both maximum total biomass and minimum broth volume simultaneously (Sarkar and Modak, 2005).

Additionally, both *f*(1) and *f*(2) could be combined by adopting either one of the following two approaches: (i) calculating the ratio between two objectives (in this case, XV and V) or (ii) converting them into a single objective function with relevant cost factors. For this study, it was observed that the optimal solution would be moving toward less reactor volume in the former approach. In the latter one, the optimal solution will vary according to the upstream and downstream processing cost, including raw material and energy cost (equation 5.7). Considering these unknown cost-related factors, in this work, the optimal Pareto between the two objectives was established by formulating the MOO problem using the former approach.

The epsilon method (ε-constraint method) was used to find the optimal Pareto points between *f*(1) and *f*(2), wherein the penalty factor $\lambda$ was tuned in between 0 and 1 (Cui et al., 2017). Optimization was performed by manipulating the substrate feeding profile from time $t_b$ to $t_f$. Various time intervals were tested, and the time interval for *F* was selected accordingly. The combined objective function could thus be defined as follows.

$$\text{Minimize } f = \lambda \frac{1}{(XV)_n} + (1-\lambda) V_n \qquad (5.3)$$

where *f* represents the objective function to be minimized with the first and second terms corresponding to *f*(1) and *f*(2), respectively. As the total biomass needs to be maximized, the objective *f*(1) is defined in the denominator. The subscript *n* represents the normalization of the two objectives to bring both objectives to numerically comparable scales.

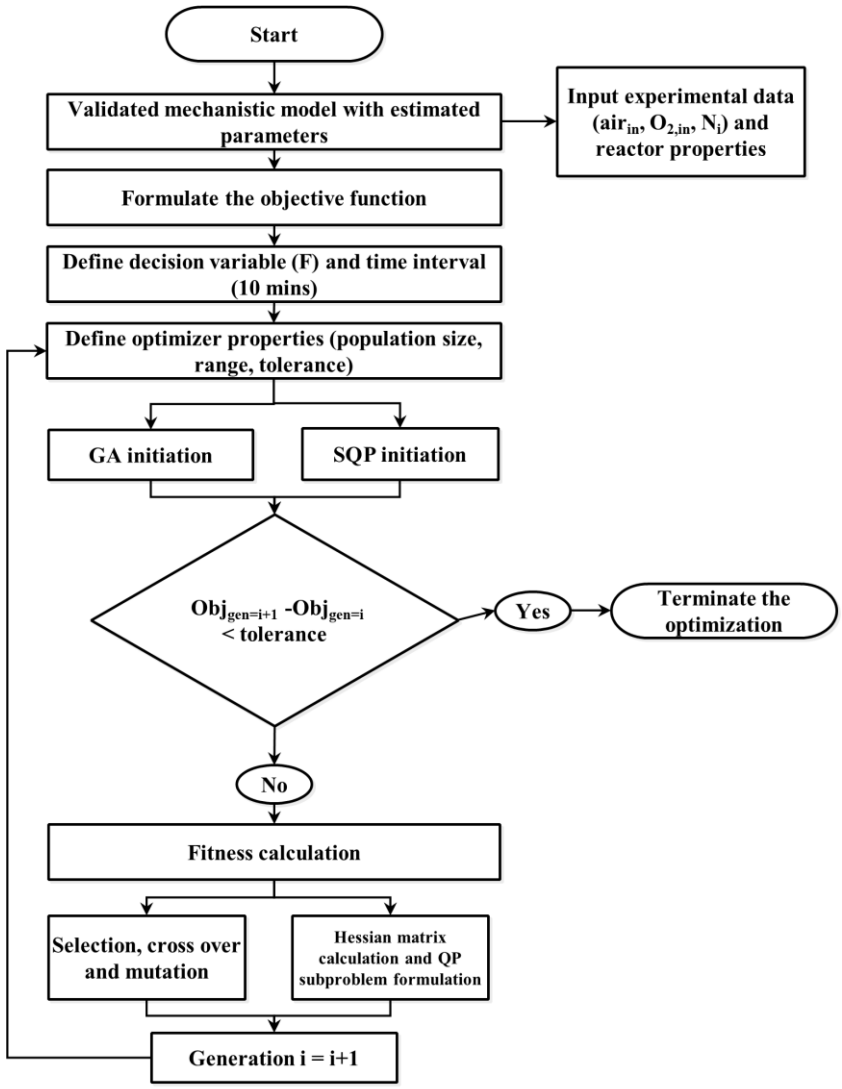



Figure 5.1. Flowchart for implementing optimization studies using Genetic algorithm (GA) and sequential quadratic programming (SQP)

The optimization studies were performed in MATLAB 2019a. Two optimization algorithms, namely, GA and sequential quadratic programming (SQP), were chosen from stochastic and deterministic methods, respectively (Pushpavanam et al., 1999; Sarkar and Modak, 2003). The GA is expected to give a global optimal solution for nonlinear problems, but SQP may converge to a local minimum in case of multiple minima. Therefore, GA was adapted to compare and verify the global optimality of the SQP's solution. However, GA is expected to consume more computational time and may not be suitable for real-time optimization applications. The flowchart representing the optimization procedure is represented in Figure 5.1. The summary of the optimization

problem formulation and the details regarding the algorithms employed to solve the problem is presented in Table 5.2, along with the case study (2).

#### 5.3.3.3 Case studies for testing the effect of fault in actuators

The optimization objective was also tested to study the effect of faults in the actuators in the system. The MV, *F*, was disturbed and made to be zero for different time intervals at different locations throughout the fed-batch phase, and the corresponding response in the optimization function was studied. The different case studies for testing the system response with respect to faults in the actuators are summarized in Table 5.1.

Table 5.1. Summary of the case studies employed to study the effect of disturbance in the system

| **Location of the fault in the actuators in the fed-batch phase** | **Time duration for which the fault in the actuator is applied (*F*=0) (min)** | | |
|---|---|---|---|
| | 120 | 50 | |
| Cases | 1 | 2 | |
| Start | [A]330 - 450 | [A1]330 - 380 | [A2]390 - 440 |
| Mid | [B]460 - 580 | [B1]450 - 500 | [B2]510 - 560 |
| End | [C]590 - 710 | [C1]570 - 620 | [C2]630 - 680 |

[A, A1, A2] fault in the actuator located at the start of the fed-batch phase

[B, B1, B2] fault in the actuator located at the mid of the fed-batch phase

[C, C1, C2] fault in the actuator located at the end of the fed-batch phase

### 5.3.4 Case study (2): Predicting optimal fed-batch harvest time

Case study (2) was carried out as an extension of the case study (1) by including the fed-batch harvest time ($t_f$ or $t_{end}$) as an additional objective. As a general practice during fed-batch cultivation, the batch is harvested based on the offline DCW values, typically when the operator observes a decrease in DCW, which is an approximate indicator of reduction in cell viability. However, predicting the optimal harvest time and observing the effect of the fed-batch harvest time would be beneficial in the long run (Patel and Padhiyar, 2017). Harvesting the batch at an optimal time can overcome the problems of over dilution of the feed and prevent cell lysis. Achieving an optimal harvest time in addition to achieving

maximum total biomass and minimal broth volume would thus be advantageous in tackling the above-said challenges. Moreover, this study could generate stable economics since the target could be achieved at an optimal time, and the resources could be effectively utilized. Therefore, case study (2) proposes the inclusion of harvest time ($t_{end}$) as an objective along with the previously designed MOO.

#### 5.3.4.1 Formulation of the objective function

The overall strategy of the case study (2) is represented in Figure 5.2. Optimizer 1 consists of the MOO objective as formulated in the case study (1). The objective for optimizer 2 was to provide optimal fed-batch harvest time ($t_{end}$). Optimizer 2 will, in turn, provide the values of different harvest time to optimizer 1, whose decision variables was chosen based on the $t_{end}$ values. The decision variable for this case study was also chosen to be the substrate flow rate during the fed-batch phase ($F$), whose constraints were chosen based on simulation studies as explained in 5.3.3.1.

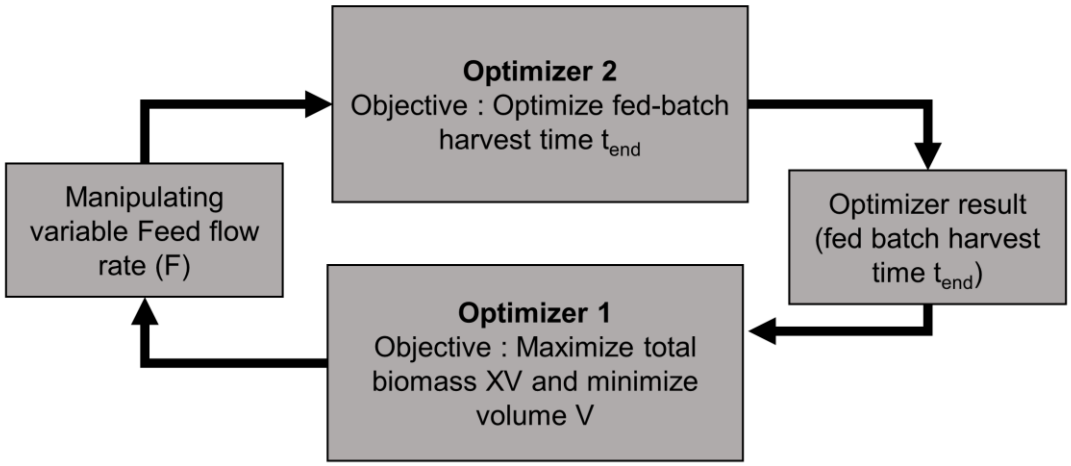



Figure 5.2. Illustration of the overall optimization strategy for case study (2).

Initially, the optimization was performed by choosing different $t_{end}$ values. Similar to the case study (1), the MOO was solved using the epsilon method for different values of $\lambda$ from 0-1. The total number of decision variables were calculated based on the $t_{end}$ values according to the equation 5.4, where $t_{int}$ refers to the time interval for the manipulating variable and $t_b$ refers to the batch time (which is approximately 5 h). The time interval of 10 min was taken forward for case study (2), according to the best interval obtained from the simulation studies of the case study (1). The objective function considered for case

study (2) was the same as the function *f* defined in the case study (1), which was now solved for different $t_{end}$ values. Therefore, the overall objective function for achieving the optimal fed-batch harvest time from the case study (2) is represented by equation 5.5 and 5.6.

$$\text{Total decision variables} = \frac{\left(t_{\text{end}} - t_b\right)}{t_{\text{int}}} \tag{5.4}$$

$$\text{Minimize } f = \lambda \frac{1}{\left(XV\right)_n} + \left(1-\lambda\right)V_n \tag{5.5}$$

Subject to constraints: $0.01 \leq F \leq 0.5$

Solved at varying values of fed-batch harvest time ($t_{end}$)

$$J_t = t_{end} \tag{5.6}$$

The optimization problem was solved in MATLAB 2021a using SQP since it was the best performing algorithm as observed from the case study (1), considering the lesser computational time required. As presented in Figure 5.3, the optimization procedure was repeated for different $t_{end}$ values to obtain the Pareto for different values of fed-batch harvest time. Alternatively, another approach to solve the two objective functions *f*(1) and *f*(2) as defined in equations 5.1 and 5.2 was attempted using the *Pareto search* algorithm from MATLAB 2021a. This methodology is an alternative to the epsilon method, and this was utilized to solve the two objectives and obtain the Pareto for different $t_{end}$ values. The Pareto search algorithm results in an optimal Pareto front and shows the trade-off between the two defined objectives for a multiobjective optimization problem. This algorithm was employed for our MOO problem for the varying fed-batch harvest time. The two approaches employed in the case study (2) for solving the optimization problem at different fed-batch harvest time values are explained in the flowchart presented in Figure 5.3. The overall summary of the optimization problem defined for the case study (2) and optimizer settings are presented in Table 5.2.

Table 5.2. Summary of the optimization problem formulated for the two case studies.

<table>
<tr><th>Attribute</th><th colspan="2">Case study (1): Optimizing total biomass at a minimum broth volume</th><th colspan="2">Case study (2): Predicting optimal fed-batch harvest time</th></tr>
<tr><td rowspan="2">Objective function</td><td colspan="4">Minimize $f = \lambda \frac{1}{(XV)_n} + (1-\lambda)V_n$<br>Subject to constraints in $F$ $0.01 \leq F \leq 0.5$</td></tr>
<tr><td colspan="2">–</td><td colspan="2">Solved for varying values of fed-batch harvest time $t_{end}$<br>$J_t = t_{end}$</td></tr>
<tr><td>Manipulating variable</td><td colspan="4">Substrate flow rate during fed-batch phase ($F$)<br>Time interval ($t_{int}$) (best): 10 min</td></tr>
<tr><td>Final reactor time ($t_f$) or Fed-batch harvest time ($t_{end}$)</td><td colspan="2">12 h (Feeding started after 5 h)</td><td colspan="2">Varying: (10-30 h)</td></tr>
<tr><td>Total decision variables</td><td colspan="2">$\frac{(t_f - t_b)}{t_{int}}$ :(7 h/10 min) = 42</td><td colspan="2">Varying: $\frac{(t_{end} - t_b)}{t_{int}}$</td></tr>
<tr><td rowspan="2">Software/ solver used</td><td colspan="2">MATLAB 2019a</td><td colspan="2">MATLAB 2021a</td></tr>
<tr><td>Sequential quadratic programming (SQP)<br>Options:<br>Max Func. Evaluations: 50000<br>Tol: 1e-06<br>Max Iter: 400</td><td>Genetic algorithm (GA)<br>Options:<br>Population Size: 1000<br>Tol: 1e-03<br>Max Gen: 100</td><td>Sequential quadratic programming (SQP)<br>Options:<br>Max Func. Evaluations: 50000<br>Tol: 1e-06<br>Max Iter: 400</td><td>Pareto search<br>Options:<br>Pareto Set Size: 60<br>Constraint Tol: 1e-06<br>Pareto Set Change Tol: 1e-04<br>Max Iter: 10</td></tr>
</table>

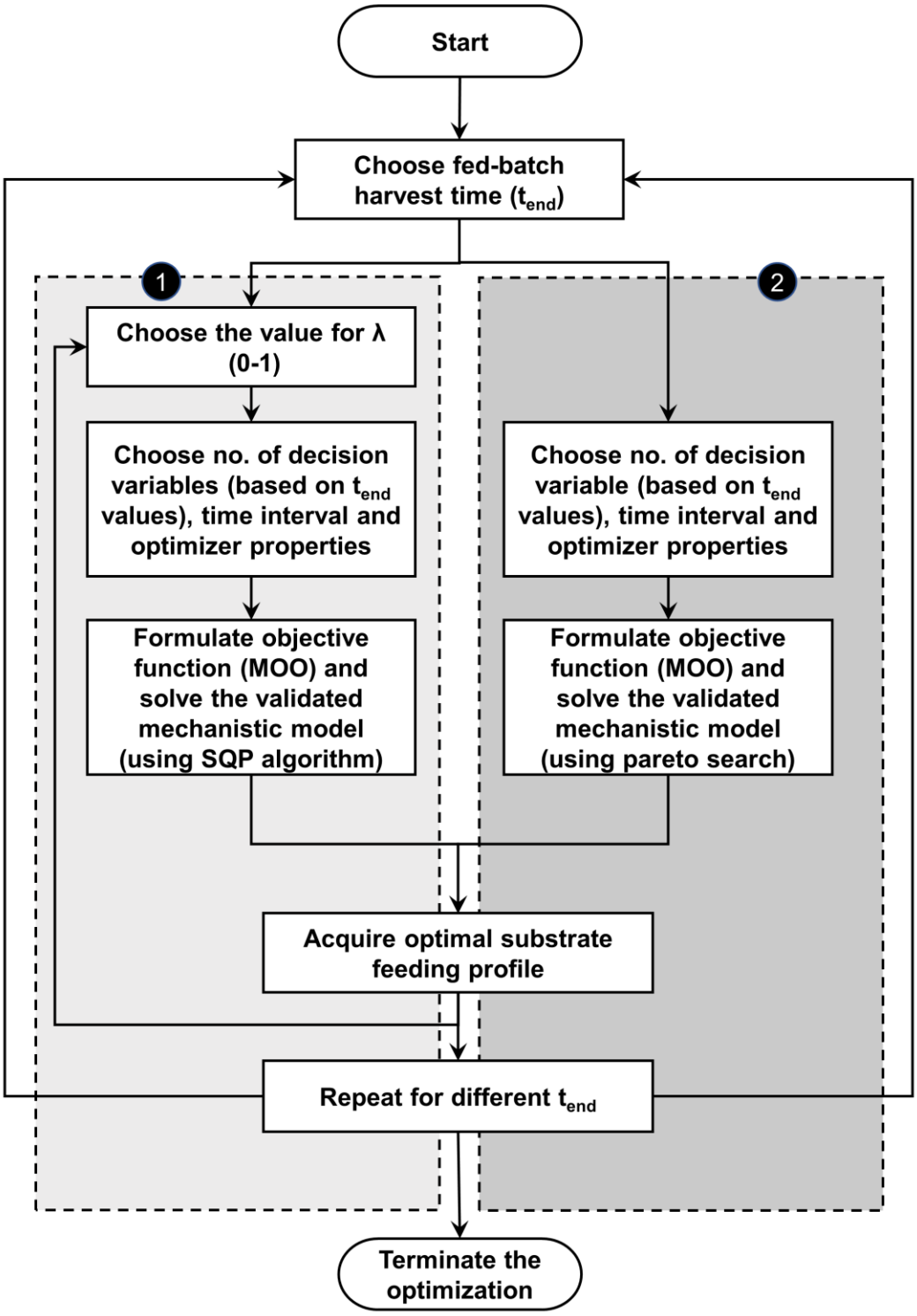



Figure 5.3. Methodology for solving the case study (2) for optimization, with sequential quadratic programming (SQP) algorithm and Pareto search algorithm, represented by the blocks labelled as 1 and 2 respectively.

## 5.4 Results and discussions

### 5.4.1 Case study (1): Optimizing total biomass at a minimum broth volume

#### 5.4.1.1 Simulation studies

The result of the simulation studies with different constant substrate flow rates is presented in Figure 5.4. It was observed that when the flow rate was varied from 0.01 L $h^{-1}$ to 0.5 L $h^{-1}$, the volume of the broth $V$ [$f(2)$] varied from 1.583 to 5 L. Because the minimum working volume for the reactor was 1.5 L and the total reactor volume was 5 L, the constraints for $F$ were fixed accordingly. Additionally, the conflicting nature of the two objectives was observed, wherein the flow rate yielding the maximum value for $f(1)$ also yielded the maximum value for $f(2)$, thereby emphasizing the need to implement the MOO strategy.

Simulation studies were performed to test the single objective optimization for $f(1)$ and $f(2)$ individually. In the case of $f(1)$ as the objective, a maximum of 78.35 g total biomass was achieved while the volume reached its upper bound of 5 L, and for the case of $f(2)$ as the objective, a minimum volume of 1.57 L was achieved while the total biomass was only 12.98 g. Therefore, it was concluded that a multiobjective optimization was necessary to address the conflicting nature of the two objectives and simultaneously achieve maximum total biomass and minimum broth volume.

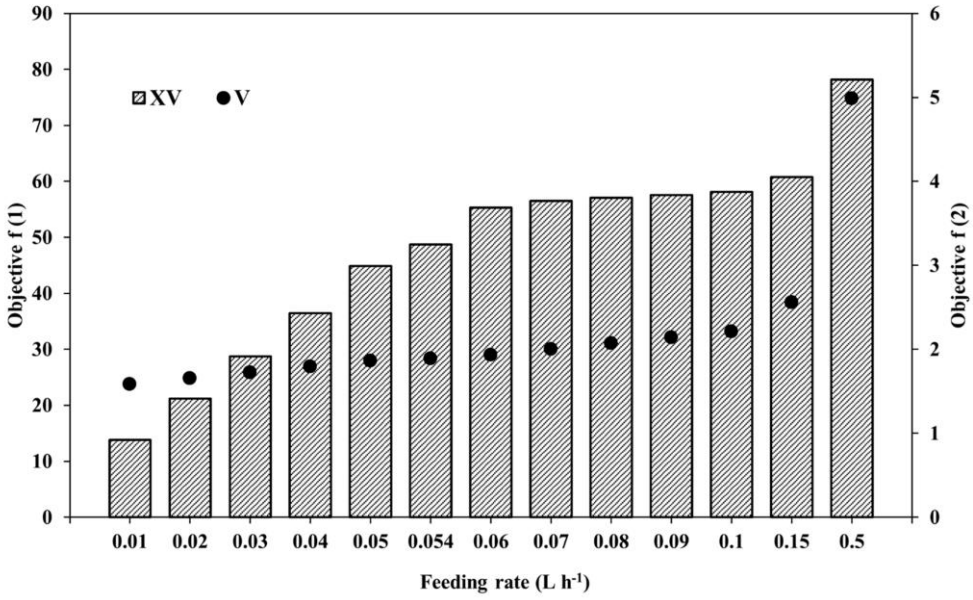



Figure 5.4. Simulation studies for different feeding rates

Among equal time interval cases, minimum *f* was achieved in cases starting from 10 min intervals. Furthermore, in variable-interval cases (with smaller intervals at the start of feeding), no significant improvement (< 2%) was observed compared to equal interval cases. Additionally, it should be noted that rapid changes in the feeding may affect the metabolism of the organism and also might not be practically feasible (Ochoa, 2016). Therefore, an equal time interval of 10 min (E2) was chosen for this case study (1). The time interval could also be included as another optimization objective. The comparison of the objective function, *f*, for all time interval simulation studies and its respective values of *f*(1) and *f*(2) are presented in Figure 5.5.

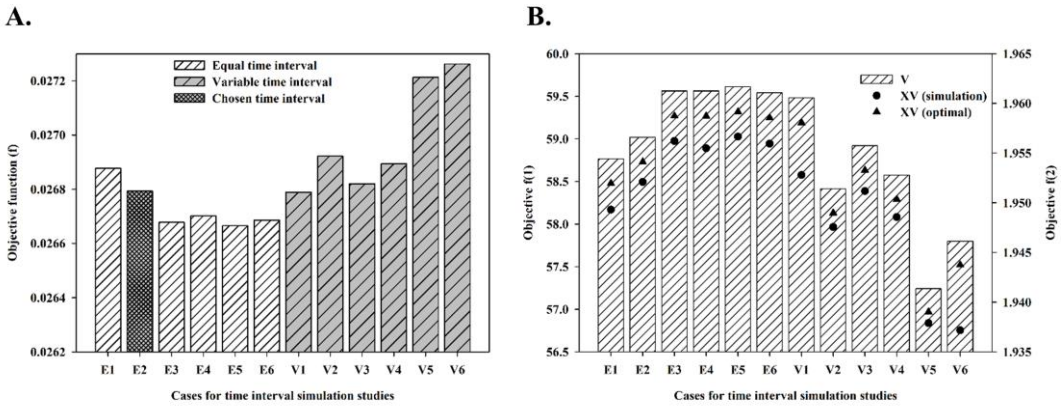



Figure 5.5. Simulation studies for different time intervals: (A). Objective function comparison. (B). Comparison of simulated and optimal total biomass at respective broth volume.

### 5.4.1.2 Multiobjective optimization for enhanced production of therapeutic proteins in *E. coli*

As mentioned in section 5.3.3.2, the MOO function was formulated, and $\lambda$ was tuned to obtain the respective Pareto for the problem. Prior to tuning the $\lambda$ value, different time intervals for *F* were tested, and as mentioned previously, an interval of 10 min was chosen to be optimum in order to give sufficient time for the system to respond. Therefore, the substrate feeding profile *F* was taken as the MV for the MOO, wherein the flow rate was manipulated for every 10 min. According to the simulation studies performed, the constraint for *F* was set to be between 0.01 and 0.5 L $h^{-1}$.

A comparison of the objective function values at different $\lambda$ values was carried out (Figure 5.6. (A)). The objective function *f*(1) reaches an asymptotic value beyond the $\lambda$ value of

0.995 (broth volume > 1.96 L). Although the $\lambda$ value of the Pareto optimization was obtained to be very close to 1, it still yielded desired and better results compared to the standard single-objective optimization, which was because of the fact that objectives *f*(1) and *f*(2) were of conflicting nature. Furthermore, based on the expenses of the downstream processes, the operator can choose any point in the asymptotic region. For instance, if the operating cost of the downstream processes is less for a high–value recombinant therapeutic protein, then one can choose to operate the batch with high volume and vice versa.

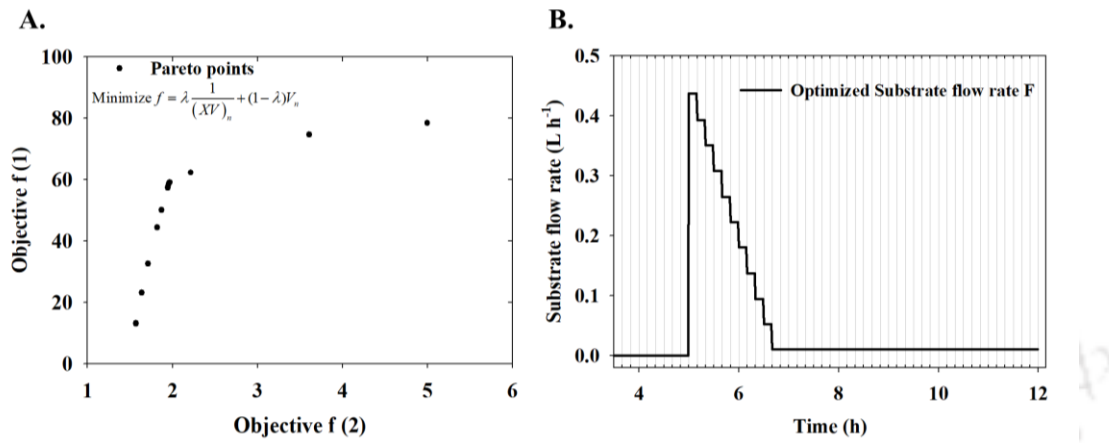



Figure 5.6. (A). Pareto for multiobjective optimization for different $\lambda$ values. (B). Optimal substrate feeding profile at a $\lambda$ value of 0.995

The corresponding optimal substrate feeding profile is represented in Figure 5.6. (B), with the minor gridlines corresponding to the time interval of 10 min. The MOO was solved using both the SQP and GA, and they yielded maximum total biomass (XV) of 58.8 g and 58.55 g for a minimum volume (V) of 1.96 L, respectively. Also, this optimization result showed that the total biomass obtained from the optimal substrate feeding profile was 20.6% higher than the total biomass from the experimental studies. The results indicated that in the case study (1), the local optimal obtained from SQP was found to be very close to the global optimal that resulted from the GA. Hence, either of the algorithms could be applied for solving the MOO developed in this study. However, GA takes a significant amount of computational time (3 h) compared to SQP (1 h) to solve the optimal profile. Consequently, further optimization studies were recommended to be carried out using SQP so that it could be implemented in real-time.

As shown in Figure 5.6 (B), the obtained optimal feeding profile initially has a spike at the beginning of the fed-batch phase, and it gradually decreases in a stepwise manner until it reaches the constraint value of 0.01 L $h^{-1}$, and then, constant feeding was observed till the end of the batch. It could be noted that this is in accordance with the experimental substrate feeding profile (refer to Figure 4.3 A), where a pulse input was given at the beginning of the fed-batch phase in order to settle the DO spike at the earliest. This was done prior to the constant feeding in order to avoid underfeeding at the beginning of the fed-batch phase. Furthermore, the overfeeding of the substrate was prevented by the constraints within the developed model equations with the Monod-type growth equation (equations 4.14 and 4.15). Therefore, it could be concluded that the optimal substrate feeding profile with an initial pulse will be practically feasible. The optimal values for $XV$ [$f(1)$] for the respective broth volume $V$ [$f(2)$] were calculated from the Pareto obtained from the MOO and are also represented in Figure 5.5 (B) for comparison. It was observed that in all the cases, the optimal value of $XV$ was higher than the values obtained from simulation, which further substantiates the results of the optimization studies.

#### 5.4.1.3 Case studies for testing the effect of fault in actuators

A comparison of the objective function values for different cases is presented in Figure 5.7. It is observed that the fault in the actuator introduced at the beginning of the fed-batch phase (case 1 A and case 2 A2) has a significant effect on the objective function. This is because, in these cases, the substrate present in the system was near zero, according to the model. It was also observed that the fault in the actuator immediately after the induction phase (case 2 C1) had the greatest effect on the objective function. In all the tested cases, the objective function increased < 2%, indicating that the developed optimization methodology was capable of handling faults in the actuators in the system, and thus, the developed optimization methodology was validated.

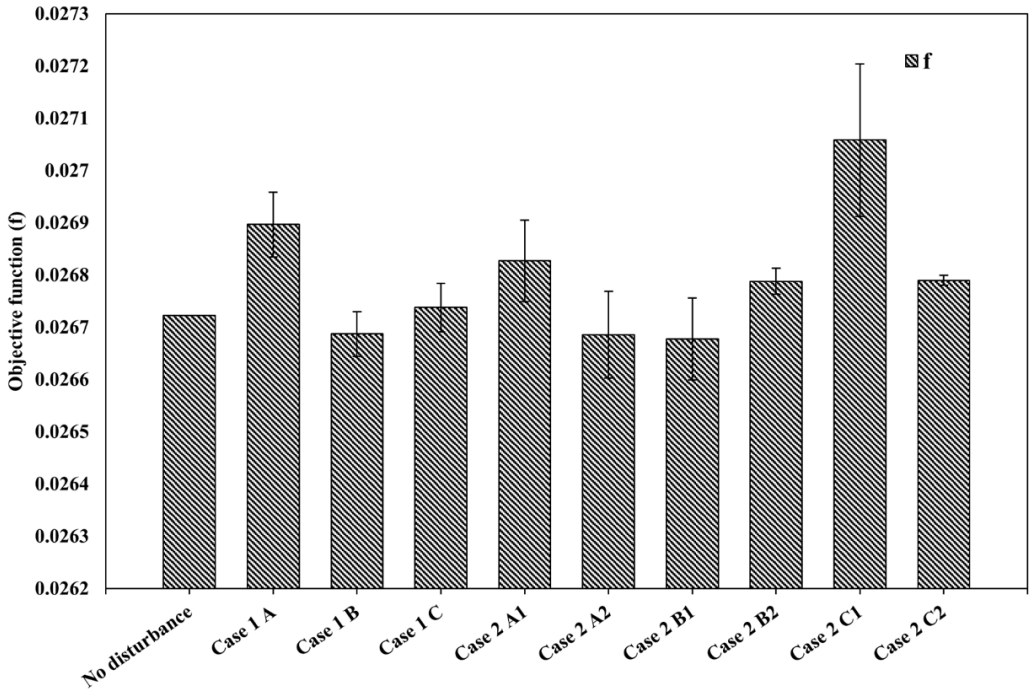



Figure 5.7. Objective function comparison for optimization case studies testing the effect of fault in actuators.

#### 5.4.1.4 Sample procedure to select optimal point for given cost factors

As mentioned previously, the two objective functions, total biomass [$f(1)$] and broth volume [$f(2)$], could be converted to a single objective function by including relevant cost factors, and this approach is explored in this section with the following assumptions.

Assume raw material cost for biomass production ($C_{biomass}$) is in \$/g of biomass and downstream processing cost for therapeutic product production ($C_{Downstreamprocessing}$) is in \$/ L of broth volume. The cost of the therapeutic protein product ($C_{product}$) is in \$/g of product, where the final yield (g) of therapeutic protein is assumed to be directly proportional to the total biomass with $\phi$ as the proportionality factor. Then, the single objective profit function ($f_{profit}$), including all the cost factors, can be calculated as follows,

$$f_{profit} = C_{product}\phi XV - \left(C_{biomass}XV + C_{Downstreamprocessing}V\right) \qquad (5.7)$$

For different values of broth volume, respective total biomass ($XV$) and subsequent product yield ($\phi XV$) could be calculated from the optimal Pareto obtained in this study. A sample representation of optimal point where profit function ( $f_{profit}$ ) is calculated using assumed cost values with respect to broth volume is outlined in Figure 5.8. The assumed cost values were as follows: $C_{product}$ = \$25000/g of product, $C_{biomass}$ = \$0.1/g of biomass, $C_{Downstreamprocessing}$ = \$23/ L of broth volume. It could be observed from Figure 5.8 that for the given cost values, the optimum point with the highest profit function value was observed at a broth volume of 2.9 L.

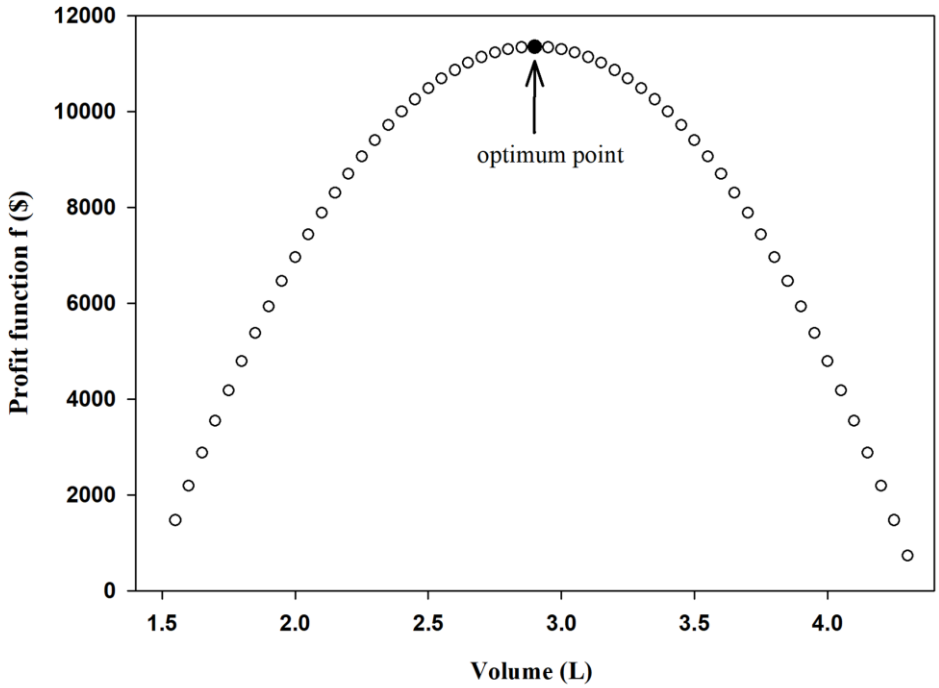



Figure 5.8. Optimal profit function for given cost factors and respective broth volume

### 5.4.2 Case study (2): Predicting optimal fed-batch harvest time

#### 5.4.2.1 Multiobjective optimization with different fed-batch harvest time

As discussed in section 5.3.4, the case study (2) was formulated to explore the influence of different values of fed-batch harvest time upon solving the previously defined multiobjective optimization. The comparison of the objective function values obtained at $\lambda$ values of 0.995 and 0.999 respectively, from different fed-batch harvest time ($t_{end}$) values are presented in Figure 5.9. It could be observed from Figure 5.9 that beyond a $t_{end}$

value of 16 h, there was no significant reduction in the value of objective function *f* for a $\lambda$ value of 0.995, and a similar observation could be observed after $t_{end}$ value of 20 h in the case of 0.999. The corresponding values of the objective functions *f*(1) and *f*(2) at the two $\lambda$ values of 0.995 and 0.999 is presented in Figure 5.10 (A) and (B), respectively. It could be observed that at higher $\lambda$ values, more significance is provided for the total biomass maximization, and therefore, the broth volume reaches the value closer to 5 L. The maximum total biomass, *XV* value, was observed at a $t_{end}$ value of 20 h and 24 h respectively for the $\lambda$ values 0.995 and 0.999, as presented in Figure 5.10. Beyond the value of 24 h, increasing the harvest time further up to 30 h did not improve the objective function *f*, and additionally, it could be observed that the total biomass value declined for a comparable value of broth volume at 24 and 30 h, as visible from Figure 5.10 (B). This emphasizes the significance of implementing optimization studies that include fed-batch harvest time as one of the objective functions.

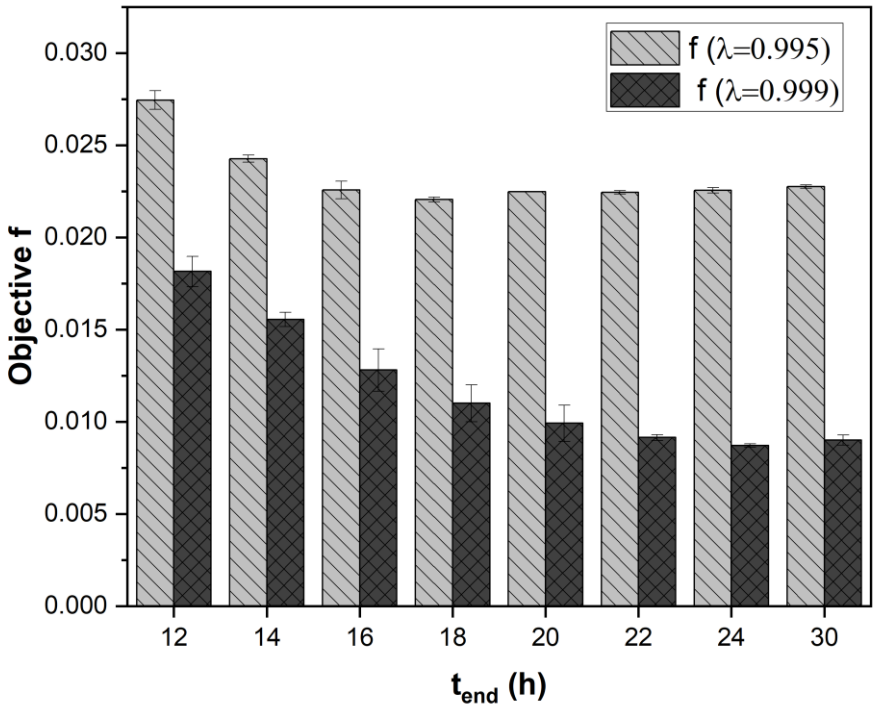



Figure 5.9. Comparison of objective function *f* for different values of fed-batch harvest time ($t_{end}$) at $\lambda$ values of 0.995 and 0.999.

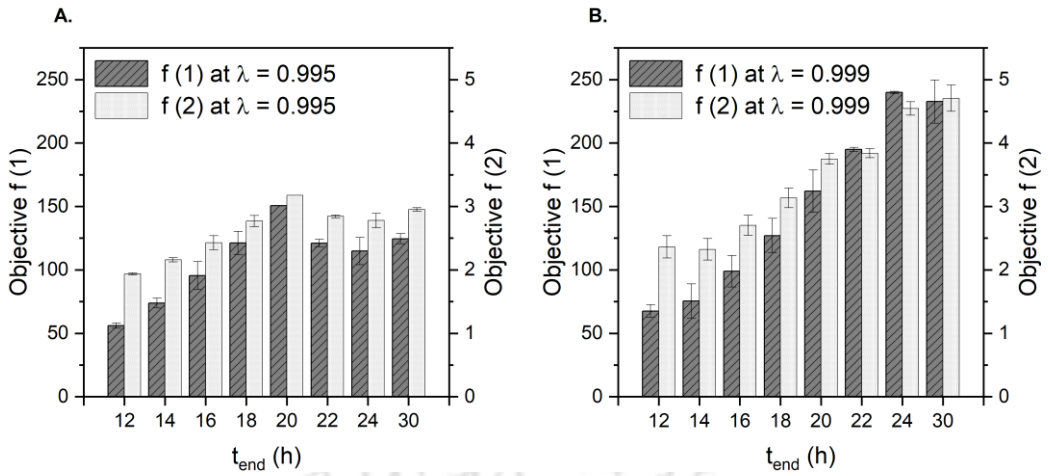



Figure 5.10. Comparison of objective function *f*(1) and *f*(2) for different values of fed-batch harvest time ($t_{end}$). (A). Objective functions at $\lambda$ value of 0.995 and (B). Objective functions at $\lambda$ value of 0.999.

The Pareto points at different $t_{end}$ values are presented in Figure 5.11. It could be observed from Figure 5.11 that for the fed-batch harvest time up to 20 h, the asymptotic value for the Pareto was observed after 0.995, and for higher $t_{end}$ values, asymptote was observed after 0.999. This observation could be interpreted as follows: for higher $t_{end}$ values, the algorithm results in a feeding profile that could increase the total biomass until the $\lambda$ value gets closer to 1. The same could be corroborated since the broth volume increases closer to the bounds at higher $t_{end}$ values due to the availability of more feeding time with increasing harvest time. However, it was observed that at a $t_{end}$ value of 30 h, the Pareto points were lesser, signifying that the objective function might not significantly improve beyond this harvest time. Further investigation into the corresponding feeding profile and the respective state variables will provide the reason behind this decrease.

Accordingly, two significant regions were highlighted from this Pareto plot as represented by two quadrants in Figure 5.11. The two regions correspond to the asymptotic value of the Pareto functions observed at 0.995 and 0.999 for different $t_{end}$ values. It could be concluded that the reactor can be operated at a suitable harvest time from the two shaded regions according to the desired volume of operation and corresponding downstream processing cost, which could be decided by the operator. For instance, if the reactor can be operated at a higher volume close to 5 L, the feeding profiles from the gold-shaded quadrilateral and the corresponding harvest time can be chosen, resulting in enhanced total biomass of up to 250 g. On the other hand, if the major focus is on reducing the

downstream processing cost, it signifies that operating the reactor at minimum broth volume is beneficial, and therefore, the operator can choose a point from the grey-shaded quadrilateral.

Another significant observation from Figure 5.11 is that increasing fed-batch harvest time might not necessarily enhance productivity. It was observed that the Pareto at a $t_{end}$ value of 30 h was below that of 24 h, and no significant improvement in the value of objective function $f$ was observed either. Therefore, it could be concluded that for the case study (2) under assumed constraints, a fed-batch harvest time of 24 h might significantly be beneficial and can yield up to 250 g of total biomass within a total broth volume of 4.5 L as observed from the Pareto points.

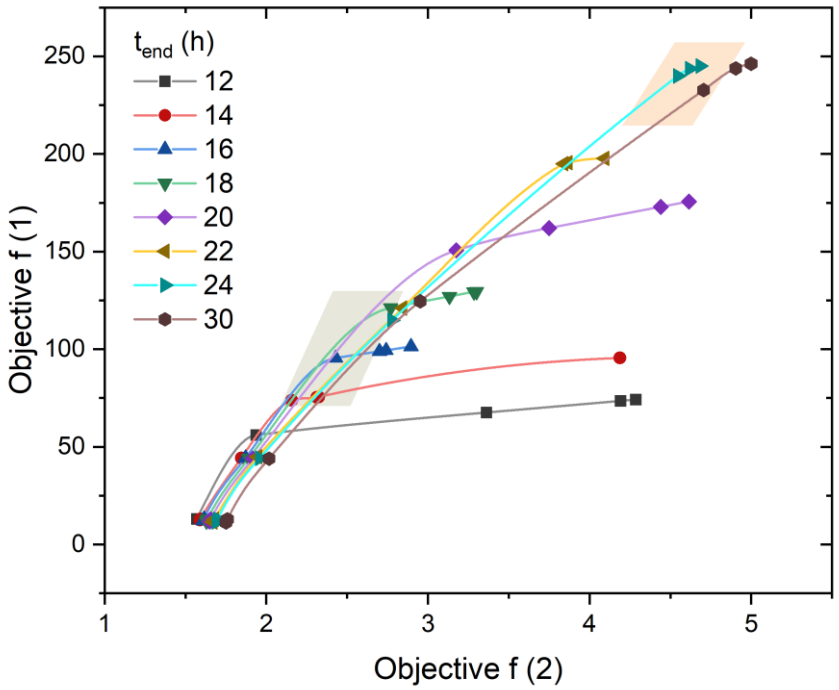



Figure 5.11. Pareto points at different values of fed-batch harvest time ($t_{end}$) obtained from changing the $\lambda$ values. The two highlighted regions (grey and gold quadrilateral) represent two significant regions for operating the reactor at the desired volume.

### 5.4.2.2 Pareto Front for two objectives *f*(1) and *f*(2)

The two objectives $f(1)$ and $f(2)$, which are of conflicting nature as deciphered from the simulation studies presented in section 5.3.3.1, can also be solved using a Pareto search

algorithm. The optimal Pareto representing the trade-off between any two defined objective functions was obtained using the Pareto search algorithm in MATLAB 2021. The Pareto points obtained from the epsilon method was obtained by solving the MOO for different $\lambda$ values by varying them one at a time. However, the Pareto search solves the individual objectives $f(1)$ (total biomass, $XV$) and $f(2)$ (broth volume, V) in a single step to yield the Pareto front. In order to observe the significance of fed-batch harvest time, the Pareto search algorithm was solved for a scaled-up reactor volume of 10 L.

The resulting Pareto front for the two objectives $f(1)$ and $f(2)$ is represented in Figure 5.12. The total Pareto set size was chosen to be the default value of 60, and therefore, it could be observed from Figure 5.12 that every Pareto front at different $t_{end}$ values consists of 60 points. It should be noted that every single Pareto point corresponds to an optimal substrate feeding profile, which was obtained by solving for the objective functions $f(1)$ and $f(2)$. The Pareto fronts depicted a similar pattern until the $t_{end}$ value of 24 h. However, it was observed that beyond 24 h, the Pareto fronts overlapped each other, which denotes that there is no improvement with regard to the desired objectives. For instance, it could be observed that a total biomass value closer to 300 g could be achieved within a broth volume of 5.3 L when the fed-batch is operated for 24 h. At the same time, an almost similar value of total biomass was obtained at an increased broth volume of 6.8 L when the fed-batch was harvested at 30 h. This is one sample representation observed from solving the considered objective functions for the defined feeding rate constraints. Therefore, it could be concluded that increasing the harvest time beyond 24 h might not yield any significant improvement in productivity, considering the current system of investigation and the assumed constraints. This observation also corroborated with the results obtained from the epsilon method, presented in Figure 5.11. Therefore, according to the case study (2), operating the fed-batch for up to 24 h can satisfy the objective functions defined in this study.

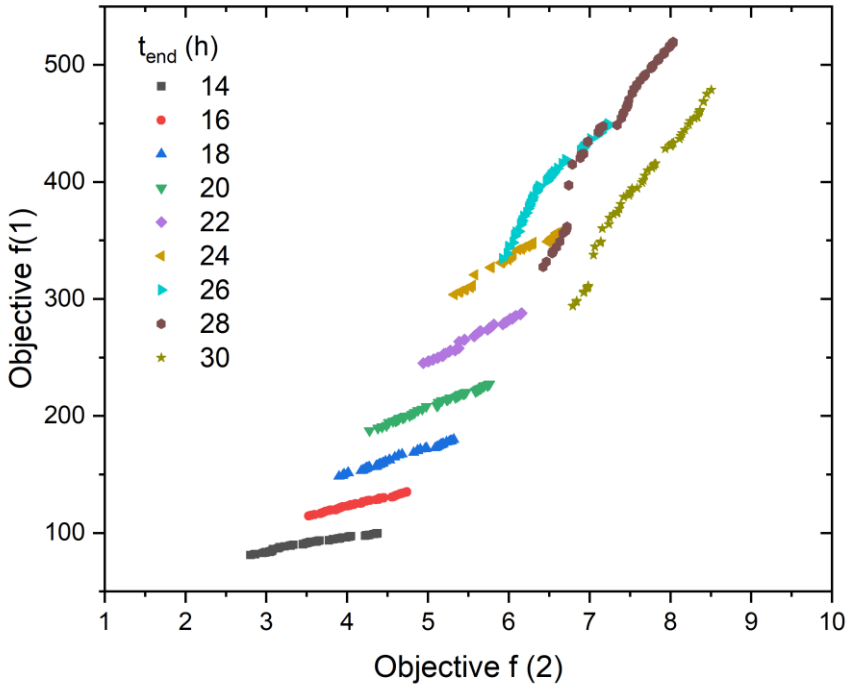



Figure 5.12 Pareto front for the two objectives *f*(1) and *f*(2) obtained from Pareto search algorithm for reactor volume of 10 L.

## 5.5 Summary

This chapter formulated two case studies for exploring the application of optimization studies to enhance the production of the therapeutic protein, Ranibizumab, in fed-batch recombinant *E. coli* cultivation. The previously validated model (as explained in chapter 4) was used to develop the MOO function, and the Pareto front resulting from the MOO revealed that the reactor could be operated above a minimum broth volume of 1.96 L with respect to the downstream operating costs. The total biomass obtained from the optimal substrate feeding profile was 20.6% higher than the total biomass from the experimental constant substrate feeding profile implemented in this study. Among SQP and GA, SQP was found to be capable of solving the optimization solution three times faster, which is required for real-time application.

It can be concluded from the case study (1) that the maximum total biomass at a desired minimal broth volume can be obtained by choosing the appropriate substrate feeding profile based on the Pareto front. The optimization methodology was tested for different

cases of faults in the actuators, and comparisons were made. These case studies indicate that the developed optimization strategy is capable of handling faults in the actuators, which might occur during experiments.

Case study (2) formulated an objective function as an extension of the case study (1), including fed-batch harvest time $t_{end}$ as an additional variable of interest for optimization. The MOO function was solved for different fed-batch harvest time values to observe the influence of fed-batch harvest time upon productivity. The Pareto points were obtained by two methods, namely, the epsilon method and the Pareto search algorithm. Two significant regions were observed from the Pareto obtained from the former method. It was concluded that the reactor could be operated at a suitable harvest time depending on the operation volume. The optimal feeding profile that would be practically feasible can be chosen to operate the reactor for desired harvest time. Additionally, it was observed that a harvest time of 24 h might be beneficial for the current study, according to the Pareto results obtained from both methodologies explored in this work.

The optimization case studies presented herein would be beneficial for operating the reactor under optimal feeding conditions and harvesting the fed-batch at a suitable harvest time. The results obtained from the optimization studies could aid in enhancing the productivity of the desired therapeutic product. The future perspective of this work would be experimental validation of the optimal results obtained from the developed optimization strategies and application of the same for predicting the optimal harvest time in real-time.

# CHAPTER 6

# Conclusions and Future recommendations

## 6.1 Foreword

The purpose of this research was set out to explore the implication of different approaches related to the application of measurement, modelling and optimization strategies in a biotherapeutic production process. The objectives accomplished in this thesis span around three objectives, as presented in chapters 3-5. The significant contributions of the work include real-time estimation of cell physiological properties and the development of a reliable process model for the application of various optimization strategies to enhance the process performance. The production of Ranibizumab as the therapeutic product in a recombinant *E. coli* cultivation was considered as the case study to explore the goals defined in this work. This chapter presents the overall summary and highlights the conclusions from each objective presented in the previous chapters. Further, the chapter concludes with some recommendations for the future and serves as the overall conclusion of the thesis.

## 6.2 Application of dielectric spectroscopy for real-time monitoring of biotherapeutic protein production

The first objective presented in this thesis focused on the application of an established process analyzer, namely, dielectric spectroscopy (DS), for monitoring the fed-batch recombinant *E. coli* cultivation. Biomass concentration is one of the critical process parameters during the production of therapeutic products, and therefore, real-time measurement of the same is extremely beneficial. Real-time estimation of the physiological properties of the organism was accomplished by correlating the DS measurements using a nonlinear Cole-Cole model. The model parameters were estimated using a genetic algorithm (GA) instead of the commonly applied Levenberg-Marquardt algorithm (LMA). A robust methodology for real-time estimation of the physiological properties was proposed as the outcome of this study. The significant conclusions from this objective can be highlighted as follows.

1. Three best-suited pre-processing techniques were applied for the raw capacitance data (moving average (MA), Savitzky-Golay filtering (SG) and smoothing splines (SS)). It was observed that the MA reduced the signal noise by smoothening the raw data up to 46%, as indicated by the Smoothening factor (SF) and comparatively had a smaller SSE value and higher correlation coefficient.

Therefore, it could be inferred that MA filtering eliminates the signal noise without much loss of actual data.

2. The pre-processed capacitance profile indicated physiological changes between the phases, including batch, fed-batch, and induction. The capacitance value rose as high as 5 pF $cm^{-1}$ and then reached a stagnant value during the production phase (after 8-9 h), which could be associated with the intracellular therapeutic recombinant protein production. The decrease or stagnation in the capacitance values toward the end of the fermentation reveals a possible change in cell morphology and/or decrease in cell viability.
3. The cell viability decreased up to 42% towards the end of fermentation, with a proportional increase in the percentage of dead cells by 24%. The measured capacitance values can corroborate well with the viable cell concentration (VCC).
4. A linear model for correlating the capacitance data with offline dry cell weight (DCW) values could be established with an $R^2$ value of 0.966 using the dual-frequency mode.
5. It was observed that the average $R^2$ values of the linear correlation established using VCC values (0.972) were higher than that of those developed using DCW values (0.938). This indicates that the measured capacitance values have a better correlation with the VCC values, which was also observed from steeper slopes with a three-fold increase than the slope values from the DCW correlation.
6. Among the two methods, LMA and GA, it was observed that global optimum values for the respective parameters were obtained from GA, irrespective of the start values. Therefore, a GA algorithm or a hybrid algorithm was better suited to obtain consistent global values for the parameters.
7. There is considerable agreement between the experimental and Cole-Cole model predicted capacitance profile, with an average error value of 15.94%. The error for the Cole-Cole model predictions were 1.03% and 7.72% for the cell diameter and VCC, respectively.
8. A robust methodology for the real-time application of the Cole-Cole model to predict the physiological properties was proposed.

The outcomes of the first objective resulted in achieving reliable monitoring of the biomass concentration using dielectric spectroscopy. Additionally, the results from this study indicated the possibility for real-time estimation of physiological properties

as demonstrated in the fed-batch recombinant *E. coli* cultivation producing Ranibizumab as a therapeutic product. The estimated real-time biomass concentration was further integrated with the developed process model and adopted to develop optimization studies based on the same as presented in the upcoming sections.

## 6.3 Modelling and validation of batch and fed-batch process of the fermenter

The second objective explored in this thesis focused on the development of a mechanistic model to represent both batch and fed-batch phases of the therapeutic protein production process. Developing a reliable process model that can fingerprint the process of interest has numerous benefits and serves as the first step towards a thorough understanding of the process. A mechanistic model representing the two phases of the fermentation process was developed based on mass balance for the state variables. The growth kinetics was represented using a Monod type dual-substrate model with residual oxygen concentration as a limiting substrate along with the carbon source, based on the sensitivity analysis and simulation studies. The parameters defined in the model equations were estimated by error minimization, and the model was later validated using a different experimental dataset. The validated model obtained from this study precisely represented the underlying process, and this could be taken forward to develop control strategies based on the same. Some of the significant results from this chapter can be summarized as follows.

1. The result of positive perturbation in all the parameters concluded that the oxygen yield coefficient in the batch phase was the most sensitive parameter for biomass concentration compared to all other parameters. This, in turn, implies that the oxygen-related parameters are sensitive towards biomass concentration. Thus, the significance of the application of the Monod type dual substrate kinetic model for describing biomass profile was observed from the sensitivity analysis.
2. It was observed that the model could predict the respective experimental values with an average error of 12.64%.
3. The model predicted the experimental DO and $O_2$ within an error value of 28.21 and 3.41%, respectively. The model predicted the experimental results within error values of 11.97%, 7.23%, 15.81%, and 9.20% for $X$, $S_b$, $S_f$ and $P$, respectively.

4. The model predicted the experimental values of the validation dataset with an average error value of 14.97%.

The observations from this objective emphasized the significance of developing a reliable process model to understand the process thoroughly. The procedure for model development and exploring the associated components of the same was thus demonstrated for the production of Ranibizumab from recombinant *E. coli* through this objective. The availability of a validated process model facilitates the opportunity to explore the implementation of optimization and control strategies to enhance the process efficiency, and this was focused on the following objective.

## 6.4 Optimization studies for maximizing biomass production and predicting harvest time

The third objective presented in this thesis focused on implementing different optimization strategies for enhancing the productivity of the protein of interest and improving the various aspects of the production process. The physiological properties of the biomass greatly influence productivity, and therefore, developing optimization strategies based on biomass concentration is significant and can aid in improving the process. Two case studies were formulated; case study (1) focused on maximizing the total biomass (XV) by simultaneously minimizing the reactor's broth volume (V). The two objectives, $f(1)$ and $f(2)$ were combined to formulate a multiobjective optimization (MOO) problem, which was then solved using different algorithms like sequential quadratic programming (SQP) and genetic algorithm (GA). Case study (2) aimed to explore the impact of different values of harvest time ($t_{end}$) along with the MOO from the case study (1). The outcome of this study was extended to demonstrate a procedure to calculate the profit function from the biotherapeutic production process. The significant outcomes from this chapter are highlighted as follows.

1. Simulation studies were carried out to choose the constraints for the manipulating variable and test the single objective optimization for $f(1)$ and $f(2)$ individually. Different time interval cases (equal and variable) were also simulated, and an equal time interval of 10 min (E2) was chosen for the first case study.
2. The MOO was solved using both the SQP and GA, and they yielded maximum total biomass (XV) of 58.8 and 58.55 g for a minimum volume (V) of 1.96 L, respectively.

3. A comparison of the objective function values at different $\lambda$ values was presented, and the corresponding optimal substrate feeding profile was obtained.
4. The optimization results from the case study (1) showed that the total biomass obtained from the optimal substrate feeding profile was 20.6% higher than the total biomass from the experimental studies.
5. Among SQP and GA, SQP was found to be capable of solving the optimization problem three times faster, which is required for real-time application.
6. The developed optimization methodology was capable of handling faults in the actuator, as the objective function increased < 2% in the tested cases, and thus, the developed optimization methodology was validated.
7. A sample procedure to select optimal points for given cost factors was demonstrated.
8. Simulation studies were carried out with different $t_{end}$ values to predict the optimal fed-batch harvest time.
9. The reactor could be operated at a suitable harvest time depending on the operation volume by choosing a practically feasible optimal feeding profile.
10. It was observed that a harvest time of 24 h might be beneficial for the current study, according to the Pareto results obtained from both methodologies (SQP and Pareto search) explored in this work.

The results from the third objective projected the possibility of achieving enhanced productivity by operating the reactor according to the optimal substrate feeding profile achieved from the optimization studies. Further, operating the reactor according to the optimal fed-batch harvest time presented in this study can streamline the therapeutic production process, yielding great economic benefits.

## 6.5 Recommendations for the future

The work presented in this thesis successfully demonstrated the application of measurement, modelling and optimization strategies to improve the process performance of a therapeutic protein production process. The significant accomplishments of this work can be summarized as follows.

- The proposed Cole-Cole model could be applied for real-time estimation of the physiological properties, thereby enabling the operator to take real-time process decisions.

- The dynamic parameter estimation and the MOO implemented in this study would be beneficial for upscaling the therapeutic protein production at an industrial scale.
- The optimization studies for biomass maximization and harvest time predictions can improve process monitoring and product quality.

Based on the observations and outcomes outlined in this thesis, the following few suggestions are provided for carrying out additional research along similar lines. The recommendations which have the scope to be pursued for future research can be briefed as follows.

- Integration and implementation of the proposed Cole-Cole methodology in real-time could be carried out in future studies. The methodology for real-time estimation of physiological properties was proposed based on the validation of the Cole-Cole model presented in this study. This proposal can be implemented in real-time, and the application of dielectric spectroscopy to make real-time decisions based on the observed physiological changes can be explored.
- Relevant kinetic models other than Monod based dual kinetics can be explored for describing the biomass growth kinetics. Additionally, the mechanistic model can be extended to include other relevant state variables like by-product formation, and measurements from other process analyzers can also be incorporated into the process model. The overall dynamic model validation strategy could be extended for other biotherapeutic protein production processes.
- The experimental validation of the optimization results reported from this work can be explored. Reactor studies can be carried out using the optimal dynamic feed profile furnished from the presented optimization results, and the enhancement in total biomass concentration and productivity can be observed. Similarly, the validation studies to harvest the fed-batch cultivation at suitable optimal harvest time can be carried out to verify the practical feasibility of the optimal results achieved.
- The validated process model presented in this study can be utilized for developing appropriate process control strategies wherein the choice of different manipulating variables can be explored. The proposed methodology for real-time biomass estimation can be employed to measure the critical process variable, and

developing a control strategy based on the same can be investigated. The application of various control strategies can significantly enhance productivity, and therefore, the presented measurement and modelling strategies can be extended for the development of the same.

- Strategies and methodologies associated with the measurement using advanced PAT tool, development of reliable process model, and formulation of different optimization strategies described in this work could be taken forward for their implementation in other relevant therapeutic production processes.
- With the growing interest in the development of 'digital twins' to serve as a replica of the underlying bioprocesses, the integration of data from advanced PAT tools can be taken forward to develop better process models that can enable the implementation of advanced real-time optimization and control strategies.
- The advancements in the field of data science can be exploited to enable efficient noise treatment and obtain more accurate process measurements. Additionally, the possibility of developing data-driven and hybrid models for the interpretation of the process measurements and obtaining a reliable representation of the fermentation process can be explored.

# REFERENCES


Abu-Absi, N.R., Kenty, B.M., Cuellar, M.E., Borys, M.C., Sakhamuri, S., Strachan, D.J., Hausladen, M.C., Li, Z.J., 2011. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol. Bioeng. 108, 1215–1221. https://doi.org/10.1002/bit.23023

Amor, K. Ben, Breeuwer, P., Verbaarschot, P., Rombouts, F.M., Akkermans, A.D.L., De Vos, W.M., Abee, T., 2002. Multiparametric flow cytometry and cell sorting for the assessment of viable, injured, and dead bifidobacterium cells during bile salt stress. Appl. Environ. Microbiol. 68, 5209–5216. https://doi.org/10.1128/AEM.68.11.5209-5216.2002

Asami, K., Hanai, T., Koizumi, N., 1980. Dielectric analysis of *Escherichia coli* suspensions in the light of the theory of interfacial polarization. Biophys. J. 31, 215–228. https://doi.org/10.1016/S0006-3495(80)85052-1

Ashoori, A., Moshiri, B., Khaki-Sedigh, A., Bakhtiari, M.R., 2009. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. J. Process Control 19, 1162–1173. https://doi.org/10.1016/j.jprocont.2009.03.006

Åström, K.J., Hägglund, T., 2006. Advanced PID Control. ISA-The Instrumentation, Systems, and Automation Society.

Atasoy, I., Yuceer, M., Berber, R., 2013. Optimisation of operating conditions in FED-batch baker's yeast fermentation. Chem. Process Eng. - Inz. Chem. i Proces. 34, 175–186. https://doi.org/10.2478/cpe-2013-0015

Atlas, R.M., Snyder, J.W., 2006. Handbook Of Media for Clinical Microbiology. CRC Press, Boca Raton. https://doi.org/10.1201/9781420005462

Azimzadeh, F., Galán, O., Barford, J., Romagnoli, J.A., 1999. On-line optimization control for a time-varying process using multiple models: A laboratory scale fermenter application. Comput. Chem. Eng. 23, S235–S239. https://doi.org/10.1016/S0098-1354(99)80058-1

Baeshen, M.N., Al-Hejin, A.M., Bora, R.S., Ahmed, M.M.M.M., Ramadan, H.A.I.I.,

Saini, K.S., Baeshen, N.A., Redwan, E.M., 2015. Production of Biopharmaceuticals in *E. coli* : Current Scenario and Future Perspectives. J. Microbiol. Biotechnol. 25, 953–962. https://doi.org/10.4014/jmb.1412.12079

Biechele, P., Busse, C., Solle, D., Scheper, T., Reardon, K., 2015. Sensor systems for bioprocess monitoring. Eng. Life Sci. 15, 469–488. https://doi.org/10.1002/elsc.201500014

Biopharmaceuticals Market (2021 - 2026) [WWW Document], 2021. URL https://www.mordorintelligence.com/industry-reports/global-biopharmaceuticals-market-industry (accessed 6.24.21).

Bracewell, D.G., Gill, A., Hoare, M., 2002. An in-line flow injection optical biosensor for real-time bioprocess monitoring. Food Bioprod. Process. Trans. Inst. Chem. Eng. Part C 80, 71–77. https://doi.org/10.1205/09603080252938690

Cannizzaro, C., Gügerli, R., Marison, I., von Stockar, U., Gu, R., Marison, I., Stockar, U. Von, 2003. On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy. Biotechnol. Bioeng. 84, 597–610. https://doi.org/10.1002/bit.10809

Casella, F., Donida, F., Åkesson, J., 2011. Object-Oriented Modeling and Optimal Control: A Case Study in Power Plant Start-Up. IFAC Proc. Vol. 44, 9549–9554. https://doi.org/10.3182/20110828-6-IT-1002.03229

Chang, L., Liu, X., Henson, M.A., 2016. Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. J. Process Control 42, 137–149. https://doi.org/10.1016/j.jprocont.2016.04.012

Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., Ierapetritou, M., 2020. Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing : Processes 8, 1–33.

Chéruy, A., 1997. Software sensors in bioprocess engineering. J. Biotechnol. 52, 193–199. https://doi.org/10.1016/S0168-1656(96)01644-6

Chhatre, S., 2012. Modelling Approaches for Bio-Manufacturing Operations, in: Mandenius, C.-F., Titchener-Hooker, N.J. (Eds.), Measurement, Monitoring, Modelling and Control of Bioprocesses. Springer, Berlin, Heidelberg, pp. 85–107.

https://doi.org/10.1007/10_2012_170

Chopda, V.R., Gomes, J., Rathore, A.S., 2016. Bridging the gap between PAT concepts and implementation: An integrated software platform for fermentation. Biotechnol. J. 11, 164–171. https://doi.org/10.1002/biot.201500507

Chopda, V.R., Rathore, A.S., Gomes, J., 2015. Maximizing biomass concentration in baker's yeast process by using a decoupled geometric controller for substrate and dissolved oxygen. Bioresour. Technol. 196, 160–168. https://doi.org/10.1016/j.biortech.2015.07.050

Clementschitsch, F., Jürgen, K., Florentina, P., Karl, B., 2005. Sensor combination and chemometric modelling for improved process monitoring in recombinant E. coli fed-batch cultivations. J. Biotechnol. 120, 183–196. https://doi.org/10.1016/j.jbiotec.2005.05.030

Cole, H., Demont, A., Marison, I., 2015. The Application of Dielectric Spectroscopy and Biocalorimetry for the Monitoring of Biomass in Immobilized Mammalian Cell Cultures. Processes 3, 384–405. https://doi.org/10.3390/pr3020384

Cole, K.S., Cole, R.H., 1941. Dispersion and absorption in dielectrics I. Alternating current characteristics. J. Chem. Phys. 9, 341–351. https://doi.org/10.1063/1.1750906

Craven, S., Whelan, J., Glennon, B., 2014. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. J. Process Control 24, 344–357. https://doi.org/10.1016/j.jprocont.2014.02.007

Cui, Y., Geng, Z., Zhu, Q., Han, Y., 2017. Review: Multi-objective optimization methods and application in energy saving. Energy 125, 681–704. https://doi.org/10.1016/j.energy.2017.02.174

Dabros, M., Amrhein, M., Bonvin, D., Marison, I.W., Von Stockar, U., 2009a. Data reconciliation of concentration estimates from mid-infrared and dielectric spectral measurements for improved on-line monitoring of bioprocesses. Biotechnol. Prog. 25, 578–588. https://doi.org/10.1002/btpr.143

Dabros, M., Dennewald, D., Currie, D.J., Lee, M.H., Todd, R.W., Marison, I.W., von Stockar, U., 2009b. Cole–Cole, linear and multivariate modeling of capacitance data

for on-line monitoring of biomass. Bioprocess Biosyst. Eng. 32, 161–73. https://doi.org/10.1007/s00449-008-0234-4

Dabros, M., Schuler, M.M., Marison, I.W., 2010. Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass. Bioprocess Biosyst. Eng. 33, 1109–1118. https://doi.org/10.1007/s00449-010-0438-2

Dan, W., Ling, G., Jianqiang, L., Yinbo, Q., Jianqun, L., Shiyuan, Y., 2002. Optimization of substrate feeding trajectory for fed-batch culture using genetic algorithm. Gong ye wei sheng wu = Ind. Microbiol. 32, 40—43.

Davey, C.L., Davey, H.M., Kell, D.B., 1992. On the dielectric properties of cell suspensions at high volume fractions. J. Electroanal. Chem. 343, 319–340. https://doi.org/10.1016/0022-0728(92)85097-M

Davey, C.L., Davey, H.M., Kell, D.B., Todd, R.W., 1993. Introduction to the dielectric estimation of cellular biomass in real time, with special emphasis on measurements at high volume fractions. Anal. Chim. Acta 279, 155–161. https://doi.org/10.1016/0003-2670(93)85078-X

Díaz Pacheco, A., Delgado-Macuil, R.J., Díaz-Pacheco, Á., Larralde-Corona, C.P., Dinorín-Téllez-Girón, J., López-Y-López, V.E., 2021. Use of equivalent circuit analysis and Cole–Cole model in evaluation of bioreactor operating conditions for biomass monitoring by impedance spectroscopy. Bioprocess Biosyst. Eng. 44, 1923–1934. https://doi.org/10.1007/s00449-021-02572-0

Doran, P.M., 2012. Bioprocess Engineering Principles 2nd ed. Elsevier, United Kingdom, pp. 411–413. https://doi.org/10.1016/S0892-6875(96)90075-8

Downey, B.J., Graham, L.J., Breit, J.F., Glutting, N.K., 2014. A novel approach for using dielectric spectroscopy to predict viable cell volume (VCV) in early process development. Biotechnol. Prog. 30, 479–487. https://doi.org/10.1002/btpr.1845

Ducommun, P., Kadouri, A., Von Stockar, U., Marison, I.W., 2001. On-line determination of animal cell concentration in two industrial high-density culture processes by dielectric spectroscopy. Biotechnol. Bioeng. 77, 316–323. https://doi.org/10.1002/bit.1197

Ehgartner, D., Hartmann, T., Heinzl, S., Frank, M., Veiter, L., Kager, J., Herwig, C., Fricke, J., 2017. Controlling the specific growth rate via biomass trend regulation in filamentous fungi bioprocesses. Chem. Eng. Sci. 172, 32–41. https://doi.org/10.1016/j.ces.2017.06.020

Ehgartner, D., Sagmeister, P., Herwig, C., Wechselberger, P., 2015. A novel real-time method to estimate volumetric mass biodensity based on the combination of dielectric spectroscopy and soft-sensors. J. Chem. Technol. Biotechnol. 90, 262–272. https://doi.org/10.1002/jctb.4469

FDA, 2004. Guidance for Industry, PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance.

Feng, M., Austin, A.J., Ward, A.C., Glassey, J., 1999. Physiological State Estimation in Recombinant *Escherichia coli* Fermentations. IFAC Proc. Vol. 32, 7527–7532. https://doi.org/10.1016/S1474-6670(17)57285-7

Flores-Cosío, G., Herrera-López, E.J., Arellano-Plaza, M., Gschaedler-Mathis, A., Kirchmayr, M., Amaya-Delgado, L., 2020. Application of dielectric spectroscopy to unravel the physiological state of microorganisms: current state, prospects and limits. Appl. Microbiol. Biotechnol. 104, 6101–6113. https://doi.org/10.1007/s00253-020-10677-x

Freitas, H., Olivo, J., Andrade, C., 2017. Optimization of Bioethanol In Silico Production Process in a Fed-Batch Bioreactor Using Non-Linear Model Predictive Control and Evolutionary Computation Techniques. Energies 10, 1763. https://doi.org/10.3390/en10111763

Fung Shek, C., Betenbaugh, M., 2021. Taking the pulse of bioprocesses: at-line and in-line monitoring of mammalian cell cultures. Curr. Opin. Biotechnol. 71, 191–197. https://doi.org/https://doi.org/10.1016/j.copbio.2021.08.007

Galvanauskas, V., Grigs, O., Vanags, J., Dubencovs, K., Stepanova, V., 2013. Model-based optimization and pO2control of fed-batch Escherichia coli and Saccharomyces cerevisiae cultivation processes. Eng. Life Sci. 13, 172–184. https://doi.org/10.1002/elsc.201200012

Gargalo, C.L., Udugama, I., Pontius, K., Lopez, P.C., Nielsen, R.F., Hasanzadeh, A.,

Mansouri, S.S., Bayer, C., Junicke, H., Gernaey, K. V., 2020. Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes. J. Ind. Microbiol. Biotechnol. 47, 947–964. https://doi.org/10.1007/s10295-020-02308-1

Geethalakshmi, S., Narendran, S., Ramalingam, S., Pappa, N., 2011. Optimization of Fed-Batch Process for Recombinant Protein Production in *Escherichia coli* Using Genetic Algorithm, in: 2011 International Conference on Process Automation, Control and Computing. IEEE, pp. 1–5. https://doi.org/10.1109/PACC.2011.5978907

Glassey, J., Gernaey, K. V., Clemens, C., Schulz, T.W., Oliveira, R., Striedner, G., Mandenius, C.F., 2011. Process analytical technology (PAT) for biopharmaceuticals. Biotechnol. J. 6, 369–377. https://doi.org/10.1002/biot.201000356

Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2008. Control of cultivation processes for recombinant protein production: A review. Bioprocess Biosyst. Eng. 31, 21–39. https://doi.org/10.1007/s00449-007-0163-7

Gnoth, S., Jenzsch, M., Simutis, R., Lübbert, A., 2007. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. J. Biotechnol. 132, 180–186. https://doi.org/10.1016/j.jbiotec.2007.03.020

Gnoth, S., Kuprijanov, A., Simutis, R., Lübbert, A., 2010. Simple adaptive pH control in bioreactors using gain-scheduling methods. Appl. Microbiol. Biotechnol. 85, 955–964. https://doi.org/10.1007/s00253-009-2114-5

Godawat, R., Konstantinov, K., Rohani, M., Warikoo, V., 2015. End-to-end integrated fully continuous production of recombinant monoclonal antibodies. J. Biotechnol. 213, 13–19. https://doi.org/10.1016/j.jbiotec.2015.06.393

Goldfeld, M., Christensen, J., Pollard, D., Gibson, E.R., Olesberg, J.T., Koerperick, E.J., Lanz, K., Small, G.W., Arnold, M.A., Evans, C.E., 2014. Advanced near-infrared monitor for stable real-time measurement and control of Pichia pastoris bioprocesses. Biotechnol. Prog. 30, 749–759. https://doi.org/10.1002/btpr.1890

Gomes, J., Chopda, V.R., Rathore, A.S., 2015. Integrating systems analysis and control for implementing process analytical technology in bioprocess development. J. Chem. Technol. Biotechnol. 90, 583–589. https://doi.org/10.1002/jctb.4591

Goudar, C., Biener, R., Boisart, C., Heidemann, R., Piret, J., de Graaf, A., Konstantinov, K., 2010. Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [13C, 1H] COSY NMR spectroscopy. Metab. Eng. 12, 138–149. https://doi.org/10.1016/j.ymben.2009.10.007

Gregory, M.E., Turner, C., 1993. Open-loop control of specific growth rate in Fed-batch cultures of recombinant *E.coli*. Biotechnol. Tech. 7, 889–894.

Gustavsson, R., Mandenius, C.F., 2013. Soft sensor control of metabolic fluxes in a recombinant Escherichia coli fed-batch cultivation producing green fluorescence protein. Bioprocess Biosyst. Eng. 36, 1375–1384. https://doi.org/10.1007/s00449-012-0840-z

Habegger, L., Crespo, K.R., Dabros, M., 2018. Preventing overflow metabolism in crabtree-positive microorganisms through on-line monitoring and control of fed-batch fermentations. Fermentation 4, 79. https://doi.org/10.3390/fermentation4030079

Hakemeyer, C., Strauss, U., Werz, S., Jose, G.E., Folque, F., Menezes, J.C., 2012. At-line NIR spectroscopy as effective PAT monitoring technique in Mab cultivations during process development and manufacturing. Talanta 90, 12–21. https://doi.org/10.1016/j.talanta.2011.12.042

Harris, C.M., Kell, D.B., 1985. The estimation of microbial biomass. Biosensors 1, 17–84. https://doi.org/10.1016/0265-928X(85)85005-7

Harris, C.M., Todd, R.W., Bungard, S.J., Lovitt, R.W., Morris, J.G., Kell, D.B., 1987. Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass. Enzyme Microb. Technol. 9, 181–186. https://doi.org/10.1016/0141-0229(87)90075-5

Hausmann, R., Henkel, M., Hecker, F., Hitzmann, B., 2017. 25 - Present Status of Automation for Industrial Bioprocesses, in: Larroche, C., Sanromán, M.Á., Du, G., Pandey, A. (Eds.), Current Developments in Biotechnology and Bioengineering.

Elsevier, pp. 725–757. https://doi.org/https://doi.org/10.1016/B978-0-444-63663-8.00025-2

He, C., Ye, P., Wang, H., Liu, X., Li, F., 2019. A systematic mass-transfer modeling approach for mammalian cell culture bioreactor scale-up. Biochem. Eng. J. 141, 173–181. https://doi.org/10.1016/j.bej.2018.09.019

Heinzle, E., Oeggerli, A., Dettwiler, B., 1990. On-line fermentation gas analysis: Error analysis and application of mass spectrometry. Anal. Chim. Acta 238, 101–115. https://doi.org/10.1016/S0003-2670(00)80528-0

Holzer, M., 2017. Is Continuous Downstream Processing Becoming a Reality? Bioprocess Int. 15, 20–28.

Höpfner, T., Bluma, A., Rudolph, G., Lindner, P., Scheper, T., 2010. A review of non-invasive optical-based image analysis systems for continuous bioprocess monitoring. Bioprocess Biosyst. Eng. 33, 247–256. https://doi.org/10.1007/s00449-009-0319-8

Horta, A.C.L., Sargo, C.R., da Silva, A.J., de Carvalho Gonzaga, M., dos Santos, M.P., Gonçalves, V.M., Zangirolami, T.C., de Campos Giordano, R., 2012. Intensification of high cell-density cultivations of r*E. coli* for production of *S. pneumoniae* antigenic surface protein, PspA3, using model-based adaptive control. Bioprocess Biosyst. Eng. 35, 1269–1280. https://doi.org/10.1007/s00449-012-0714-4

Horta, A.C.L., Silva, A.J. da, Sargo, C.R., Cavalcanti-Montaño, I.D., Galeano-Suarez, I.D., Velez, A.M., Santos, M.P., Gonçalves, V.M., Giordano, R.C., Zangirolami, T.C., 2015. On-line monitoring of biomass concentration based on a capacitance sensor: assessing the methodology for different bacteria and yeast high cell density fed-batch cultures. Brazilian J. Chem. Eng. 32, 821–829. https://doi.org/10.1590/0104-6632.20150324s00003534

Ignova, M., Montague, G.A., Glassey, J., Ward, A.C., Kornfeld, G., Thomas, C.R., 1998. Hybrid Modeling and Optimisation of Industrial Fed-Batch Fermentation Process. IFAC Proc. Vol. 31, 271–276. https://doi.org/10.1016/S1474-6670(17)40197-2

Isidro, I.A., Pais, D.A.M., Domingues, M., Abecasis, B., Almeida, J.I., Linares, N.Z., Madoz, J.R.R., Prosper, F., Aspegren, A., Alves, P.M., 2021. Online monitoring of

hiPSC expansion and hepatic differentiation in 3D culture by dielectric spectroscopy 3610–3617. https://doi.org/10.1002/bit.27751

Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006a. Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. J. Biotechnol. 127, 84–94. https://doi.org/10.1016/j.jbiotec.2006.06.004

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A., 2006b. Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. Bioprocess Biosyst. Eng. 29, 315–321. https://doi.org/10.1007/s00449-006-0080-1

Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A., 2006c. Estimation of biomass concentrations in fermentation processes for recombinant protein production. Bioprocess Biosyst. Eng. 29, 19–27. https://doi.org/10.1007/s00449-006-0051-6

Jenzsch, M., Simutis, R., Luebbert, A., 2006d. Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. J. Biotechnol. 122, 483–493. https://doi.org/10.1016/j.jbiotec.2005.09.013

Jungbauer, A., 2013. Continuous downstream processing of biopharmaceuticals. Trends Biotechnol. 31, 479–492. https://doi.org/10.1016/j.tibtech.2013.05.011

Junker, B.H., Wang, H.Y., 2006. Bioprocess monitoring and computer control: Key roots of the current PAT initiative. Biotechnol. Bioeng. 95, 226–261. https://doi.org/10.1002/bit.21087

Justice, C., Brix, A., Freimark, D., Kraume, M., Pfromm, P., Eichenmueller, B., Czermak, P., 2011. Process control in cell culture technology using dielectric spectroscopy. Biotechnol. Adv. 29, 391–401. https://doi.org/10.1016/j.biotechadv.2011.03.002

Kager, J., Tuveri, A., Ulonska, S., Kroll, P., Herwig, C., 2020. Experimental verification and comparison of model predictive, PID and model inversion control in a Penicillium chrysogenum fed-batch process. Process Biochem. 90, 1–11. https://doi.org/10.1016/j.procbio.2019.11.023

Kahraman, C., Öztaysi, B., Onar, S.C., 2016. A Comprehensive Literature Review of 50

Years of Fuzzy Set Theory. Int. J. Comput. Intell. Syst. 9, 3. https://doi.org/10.1080/18756891.2016.1180817

Kaiser, C., Pototzki, T., Ellert, A., Luttmann, R., 2008. Applications of PAT-Process Analytical Technology in Recombinant Protein Processes with *Escherichia coli* . Eng. Life Sci. 8, 132–138. https://doi.org/10.1002/elsc.200720232

Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng. 82, 35. https://doi.org/10.1115/1.3662552

Kapadi, M.D., Gudi, R.D., 2004. Optimal control of fed-batch fermentation involving multiple feeds using Differential Evolution. Process Biochem. 39, 1709–1721. https://doi.org/10.1016/j.procbio.2003.07.006

Katla, S., Mohan, N., Pavan, S.S., Pal, U., Sivaprakasam, S., 2019. Control of specific growth rate for the enhanced production of human interferon α2b in glycoengineered *Pichia pastoris* : process analytical technology guided approach. J. Chem. Technol. Biotechnol. 94, 3111–3123. https://doi.org/10.1002/jctb.6118

Kawohl, M., Heine, T., King, R., 2007. Model based estimation and optimal control of fed-batch fermentation processes for the production of antibiotics. Chem. Eng. Process. Process Intensif. 46, 1223–1241. https://doi.org/10.1016/j.cep.2006.06.023

Kiviharju, K., Salonen, K., Moilanen, U., Eerikäinen, T., 2008. Biomass measurement online: The performance of in situ measurements and software sensors. J. Ind. Microbiol. Biotechnol. 35, 657–665. https://doi.org/10.1007/s10295-008-0346-5

Kiviharju, K., Salonen, K., Moilanen, U., Meskanen, E., Leisola, M., Eerikäinen, T., 2007. On-line biomass measurements in bioreactor cultivations: Comparison study of two on-line probes. J. Ind. Microbiol. Biotechnol. 34, 561–566. https://doi.org/10.1007/s10295-007-0233-5

Knabben, I., 2011. Linear correlation between online capacitance and offline biomass measurement up to high cell densities in *Escherichia coli* fermentations in a pilot-scale pressurized bioreactor. J. Microbiol. Biotechnol. 21, 204–211. https://doi.org/10.4014/jmb.1004.04032

Ko, C.-L., Wang, F.-S., 2006. Run-to-run fed-batch optimization for protein production using recombinant *Escherichia coli* . Biochem. Eng. J. 30, 279–285.

https://doi.org/10.1016/j.bej.2006.05.010

Ko, C.L., Wang, F.S., 2007. On-line estimation of biomass and intracellular protein for recombinant *Escherichia coli* cultivated in batch and fed-batch modes. J. Chinese Inst. Chem. Eng. 38, 197–203. https://doi.org/10.1016/j.jcice.2007.04.005

Konstantinov, K.B., Cooney, C.L., 2015. White paper on continuous bioprocessing May 20-21, 2014 continuous manufacturing symposium. J. Pharm. Sci. 104, 813–820. https://doi.org/10.1002/jps.24268

Kornaros, M., Lyberatos, G., 1997. Kinetics of aerobic growth of a denitrifying bacterium, Pseudomonas denitrificans, in the presence of nitrates and/or nitrites. Water Res. 31, 479–488. https://doi.org/10.1016/S0043-1354(96)00288-6

Kotidis, P., Jedrzejewski, P., Sou, S.N., Sellick, C., Polizzi, K., del Val, I.J., Kontoravdi, C., 2019. Model-based optimization of antibody galactosylation in CHO cell culture. Biotechnol. Bioeng. 116, 1612–1626. https://doi.org/10.1002/bit.26960

Kourlas, H., Abrams, P., 2007. Ranibizumab for the treatment of neovascular age-related macular degeneration: A review. Clin. Ther. 29, 1850–1861. https://doi.org/10.1016/j.clinthera.2007.09.008

Kuprijanov, A., Gnoth, S., Simutis, R., Lübbert, A., 2009. Advanced control of dissolved oxygen concentration in fed batch cultures during recombinant protein production. Appl. Microbiol. Biotechnol. 82, 221–229. https://doi.org/10.1007/s00253-008-1765-y

Kuprijanov, A., Schaepe, S., Simutis, R., Lübbert, A., 2013. Model predictive control made accessible to professional automation systems in fermentation technology. Biosyst. Inf. Technol. 2, 26–31. https://doi.org/10.11592/bit.131101

Kuystermans, D., Mohd, A., Al-Rubeai, M., 2012. Automated flow cytometry for monitoring CHO cell cultures. Methods 56, 358–365. https://doi.org/10.1016/j.ymeth.2012.03.001

Lee, J., Lee, S.Y., Park, S., Middelberg, A.P.J., 1999. Control of fed-batch fermentations. Biotechnol. Adv. 17, 29–48. https://doi.org/10.1016/S0734-9750(98)00015-9

Levisauskas, D., Simutis, R., Borvitz, O., Lijbbert, A., Borvitz, D., Lübbert, A., Borvitz, O., Lijbbert, A., 1996. Automatic control of the specific growth rate in fed-batch

cultivation processes based on an exhaust gas analysis. Bioprocess Eng. 15, 145–150. https://doi.org/10.1007/s004490050248

Lira-Parada, P.A., Tuveri, A., Seibold, G.M., Bar, N., 2021. Comparison of noninvasive, in-situ and external monitoring of microbial growth in fed-batch cultivations in Corynebacterium glutamicum. Biochem. Eng. J. 170, 107989. https://doi.org/10.1016/j.bej.2021.107989

Logan, D., Carvell, J., 2011. A biomass monitor for disposable bioreactors. Bioprocess Int.

Lopes, M.B., Martins, G., Calado, C.R.C., 2014. Kinetic modeling of plasmid bioproduction in *Escherichia coli* DH5α cultures over different carbon-source compositions. J. Biotechnol. 186, 38–48. https://doi.org/10.1016/j.jbiotec.2014.06.022

Loske, A.M., Tello, E.M., Vargas, S., Rodriguez, R., 2014. *Escherichia coli* viability determination using dynamic light scattering: a comparison with standard methods. Arch. Microbiol. 196, 557–563. https://doi.org/10.1007/s00203-014-0995-x

Lüder, C., Lindner, P., Bulnes-Abundis, D., Lu, S.M., Lücking, T., Solle, D., Scheper, T., 2014. In situ microscopy and MIR-spectroscopy as non-invasive optical sensors for cell cultivation process monitoring. Pharm. Bioprocess. 2, 157–166. https://doi.org/10.4155/pbp.14.13

Luedeking, R., Piret, E.L., 2000. Kinetic study of the lactic acid fermentation. Batch process at controlled pH. Biotechnol. Bioeng. 67, 636–644. https://doi.org/10.1002/(SICI)1097-0290(20000320)67:6<636::AID-BIT3>3.0.CO;2-U

Luo, Y., Kurian, V., Ogunnaike, B.A., 2021. Bioprocess systems analysis, modeling, estimation, and control. Curr. Opin. Chem. Eng. 33, 100705. https://doi.org/10.1016/j.coche.2021.100705

Luttmann, R., Bracewell, D.G., Cornelissen, G., Gernaey, K. V., Glassey, J., Hass, V.C., Kaiser, C., Preusse, C., Striedner, G., Mandenius, C.F., 2012. Soft sensors in bioprocessing: A status report and recommendations. Biotechnol. J. 7, 1040–1048. https://doi.org/10.1002/biot.201100506

Ma, F., Zhang, A., Chang, D., Velev, O.D., Wiltberger, K., Kshirsagar, R., 2019. Real-time monitoring and control of CHO cell apoptosis by in situ multifrequency scanning dielectric spectroscopy. Process Biochem. 80, 138–145. https://doi.org/10.1016/j.procbio.2019.02.017

Mahato, S., Singh, A., Rangan, L., Jana, C.K., 2016. Synthesis, in silico studies and in vitro evaluation for antioxidant and antibacterial properties of diarylmethylamines: A novel class of structurally simple and highly potent pharmacophore. Eur. J. Pharm. Sci. 88, 202–209. https://doi.org/10.1016/j.ejps.2016.03.004

Mandenius, C.-F., 2004. Recent developments in the monitoring, modeling and control of biological production systems. Bioprocess Biosyst. Eng. 26, 347–351. https://doi.org/10.1007/s00449-004-0383-z

Marison, I., Hennessy, S., Foley, R., Schuler, M., Sivaprakasam, S., Freeland, B., 2013. The choice of suitable online analytical techniques and data processing for monitoring of bioprocesses. Adv. Biochem. Eng. Biotechnol. 132, 249–80. https://doi.org/10.1007/10_2012_175

Markx, G.H., Davey, C.L., 1999. The dielectric properties of biological cells at radiofrequencies: applications in biotechnology. Enzyme Microb. Technol. 25, 161–171. https://doi.org/10.1016/S0141-0229(99)00008-3

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. https://doi.org/10.1137/0111030

Maskow, T., Röllich, A., Fetzer, I., Ackermann, J.-U., Harms, H., 2008a. On-line monitoring of lipid storage in yeasts using impedance spectroscopy. J. Biotechnol. 135, 64–70. https://doi.org/10.1016/j.jbiotec.2008.02.014

Maskow, T., Röllich, A., Fetzer, I., Yao, J., Harms, H., 2008b. Observation of non-linear biomass-capacitance correlations: Reasons and implications for bioprocess control. Biosens. Bioelectron. 24, 123–128. https://doi.org/10.1016/j.bios.2008.03.024

McCready, C., 2017. Model Predictive Control for Bioprocess Forecasting and Optimization. Bioprocess Int. 15, 14–20.

Mears, L., Stocks, S.M., Albaek, M.O., Sin, G., Gernaey, K. V., 2017a. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. Trends

Biotechnol. 35, 914–924. https://doi.org/10.1016/j.tibtech.2017.07.002

Mears, L., Stocks, S.M., Sin, G., Gernaey, K. V., 2017b. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. J. Biotechnol. 245, 34–46. https://doi.org/10.1016/j.jbiotec.2017.01.008

Menawat, A., Mutharasan, R., Coughanowr, D.R., 1987. Singular optimal control strategy for a fed-batch bioreactor: Numerical approach. AIChE J. 33, 776–783. https://doi.org/10.1002/aic.690330510

Metze, S., Ruhl, S., Greller, G., Grimm, C., Scholz, J., 2019. Monitoring online biomass with a capacitance sensor during scale-up of industrially relevant CHO cell culture fed-batch processes in single-use bioreactors. Bioprocess Biosyst. Eng. 43, 193–205. https://doi.org/10.1007/s00449-019-02216-4

Modak, J.M., Lim, H.C., 1989. Optimal operation of fed-batch bioreactors with two control variables. Chem. Eng. J. 42, B15–B24. https://doi.org/10.1016/0300-9467(89)85008-7

Mohan, N., Pavan, S.S., Achar, A., Swaminathan, N., Sivaprakasam, S., 2019. Calorespirometric investigation of *Streptococcus zooepidemicus* metabolism: Thermodynamics of anabolic payload contribution by growth and hyaluronic acid synthesis. Biochem. Eng. J. 152, 107367. https://doi.org/10.1016/j.bej.2019.107367

Mohan, N., Pavan, S.S., Jayakumar, A., Rathinavelu, S., Sivaprakasam, S., 2022. Real-time metabolic heat-based specific growth rate soft sensor for monitoring and control of high molecular weight hyaluronic acid production by Streptococcus zooepidemicus. Appl. Microbiol. Biotechnol. 106, 1079–1095. https://doi.org/10.1007/s00253-022-11760-1

Mohan, N., Sivaprakasam, S., 2017. Heat Compensation Calorimeter as a Process Analytical Tool to Monitor and Control Bioprocess Systems. Ind. Eng. Chem. Res. 56, 8416–8427. https://doi.org/10.1021/acs.iecr.7b01367

Monod, J., 1949. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 3, 371–394. https://doi.org/10.1146/annurev.mi.03.100149.002103

Montague, G.A., Glassey, J., Ignova, M., Paul, G.C., Kent, C.A., Thomas, C.R., Ward, A.C., 2002. Hybrid modelling for on-line penicillin fermentation optimisation. IFAC

Proc. Vol. 35, 395–400. https://doi.org/10.3182/20020721-6-ES-1901.01375

Narayanan, H., Sokolov, M., Morbidelli, M., Butté, A., 2019. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnol. Bioeng. 116, 2540–2549. https://doi.org/10.1002/bit.27097

Nasir, N., Al Ahmad, M., 2020. Cells Electrical Characterization: Dielectric Properties, Mixture, and Modeling Theories. J. Eng. 2020, 1–17. https://doi.org/10.1155/2020/9475490

Noll, P., Henkel, M., 2020. History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance. Comput. Struct. Biotechnol. J. 18, 3309–3323. https://doi.org/10.1016/j.csbj.2020.10.018

O'Shea, N., O'Callaghan, T.F., Tobin, J.T., 2019. The application of process analytical technologies (PAT)to the dairy industry for real time product characterization - process viscometry. Innov. Food Sci. Emerg. Technol. 55, 48–56. https://doi.org/10.1016/j.ifset.2019.05.003

Ochoa, S., 2016. A new approach for finding smooth optimal feeding profiles in fed-batch fermentations. Biochem. Eng. J. 105, 177–188. https://doi.org/10.1016/j.bej.2015.09.004

Ohadi, K., Aghamohseni, H., Legge, R.L., Budman, H.M., 2014. Fluorescence-based soft sensor for at situ monitoring of chinese hamster ovary cell cultures. Biotechnol. Bioeng. 111, 1577–1586. https://doi.org/10.1002/bit.25222

Opel, C.F., Li, J., Amanullah, A., 2010. Quantitative modeling of viable cell density, cell size, intracellular conductivity, and membrane capacitance in batch and fed-batch CHO processes using dielectric spectroscopy. Biotechnol. Prog. 26, 1187–1199. https://doi.org/10.1002/btpr.425

Park, S.Y., Park, C.H., Choi, D.H., Hong, J.K., Lee, D.Y., 2021. Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing. Curr. Opin. Chem. Eng. 33, 100702. https://doi.org/10.1016/j.coche.2021.100702

Patel, N., Padhiyar, N., 2017. Multi-objective dynamic optimization study of fed-batch bio-reactor. Chem. Eng. Res. Des. 119, 160–170.

https://doi.org/10.1016/j.cherd.2017.01.002

Priyanka, Kumar, J., Gomes, J., Rathore, A.S., 2019. Implementing Process Analytical Technology for the Production of Recombinant Proteins in *Escherichia coli* Using an Advanced Controller Scheme. Biotechnol. J. 1800556. https://doi.org/10.1002/biot.201800556

Priyanka, P., Rathore, A.S., 2021. A novel strategy for efficient expression of an antibody fragment in Escherichia coli : ranibizumab as a case study . J. Chem. Technol. Biotechnol. https://doi.org/10.1002/jctb.6883

Priyanka, Roy, S., Chopda, V., Gomes, J., Rathore, A.S., 2018. Comparison and implementation of different control strategies for improving production of rHSA using Pichia pastoris. J. Biotechnol. 290, 33–43. https://doi.org/10.1016/J.JBIOTEC.2018.12.002

Proß, S., Bachmann, B., 2012. Hybrid Modelling and Process Optimization of Biological Systems. {IFAC} Proc. Vol. 45, 1041–1046. https://doi.org/http://dx.doi.org/10.3182/20120215-3-AT-3016.00184

Pushpavanam, S., Rao, S., Khan, I., 1999. Optimization of a biochemical fed-batch reactor using sequential quadratic programming. Ind. Eng. Chem. Res. 38, 1998–2004. https://doi.org/10.1021/ie9805483

Qian, G., Mahdi, A., 2020. Sensitivity analysis methods in the biomedical sciences. Math. Biosci. 323. https://doi.org/10.1016/j.mbs.2020.108306

Raftery, J.P., DeSessa, M.R., Karim, M.N., 2017. Economic improvement of continuous pharmaceutical production via the optimal control of a multifeed bioreactor. Biotechnol. Prog. 33, 902–912. https://doi.org/10.1002/btpr.2433

Rajamanickam, V., Babel, H., Montano-Herrera, L., Ehsani, A., Stiefel, F., Haider, S., Presser, B., Knapp, B., 2021. About model validation in bioprocessing. Processes 9, 1–17. https://doi.org/10.3390/pr9060961

Randek, J., Mandenius, C.-F.F., 2020. In situ scanning capacitance sensor with spectral analysis reveals morphological states in cultures for production of biopharmaceuticals. Sensors Actuators B Chem. 313, 128052. https://doi.org/10.1016/j.snb.2020.128052

Randek, J., Mandenius, C.F., 2018. On-line soft sensing in upstream bioprocessing. Crit. Rev. Biotechnol. 38, 106–121. https://doi.org/10.1080/07388551.2017.1312271

Rathore, A., Kateja, N., Agarwal, H., Sharma, A.K., 2016. Continuous processing for the production of biopharmaceuticals. BioPharm Int. 29, 1–5.

Rathore, A.S., 2016. Quality by Design ( QbD ) -Based Process Development for Purification of a Biotherapeutic. Trends Biotechnol. 34, 358–370. https://doi.org/10.1016/j.tibtech.2016.01.003

Rathore, A.S., 2014. QbD/PAT for bioprocessing: Moving from theory to implementation. Curr. Opin. Chem. Eng. 6, 1–8. https://doi.org/10.1016/j.coche.2014.05.006

Rathore, A.S., Agarwal, H., Sharma, A.K., Pathak, M., Muthukumar, S., 2015. Continuous Processing for Production of Biopharmaceuticals. Prep. Biochem. Biotechnol. 45, 836–849. https://doi.org/10.1080/10826068.2014.985834

Rathore, A.S., Bhambure, R., Ghare, V., 2010. Process analytical technology (PAT) for biopharmaceutical products. Anal. Bioanal. Chem. 398, 137–154. https://doi.org/10.1007/s00216-010-3781-x

Rathore, A.S., Kateja, N., 2016. A coiled flow inversion reactor enables continuous processing. BioPharm Int. 29, 32–35.

Rathore, A.S., Mishra, S., Nikita, S., Priyanka, P., 2021. Bioprocess control: Current progress and future perspectives. Life 11. https://doi.org/10.3390/life11060557

Read, E.K., Park, J.T., Shah, R.B., Riley, B.S., Brorson, K.A., Rathore, A.S., 2010. Process analytical technology (PAT) for biopharmaceutical products: Part I. concepts and applications. Biotechnol. Bioeng. 105, 276–284. https://doi.org/10.1002/bit.22528

Reardon, K.F., 2021. Practical monitoring technologies for cells and substrates in biomanufacturing. Curr. Opin. Biotechnol. 71, 225–230. https://doi.org/10.1016/j.copbio.2021.08.006

Reichelt, W.N., Thurrold, P., Brillmann, M., Kager, J., Fricke, J., Herwig, C., 2016. Generic biomass estimation methods targeting physiologic process control in induced bacterial cultures. Eng. Life Sci. 16, 720–730.

https://doi.org/10.1002/elsc.201500182

Sander, R., 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmos. Chem. Phys. 15, 4399–4981. https://doi.org/10.5194/acp-15-4399-2015

Santos, L.O., Dewasme, L., Coutinho, D., Wouwer, A. Vande, 2012. Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. Comput. Chem. Eng. 39, 143–151. https://doi.org/10.1016/j.compchemeng.2011.12.010

Sarkar, D., Modak, J.M., 2005. Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. Chem. Eng. Sci. 60, 481–492. https://doi.org/10.1016/j.ces.2004.07.130

Sarkar, D., Modak, J.M., 2003. Optimisation of fed-batch bioreactors using genetic algorithms. Chem. Eng. Sci. 58, 2283–2296. https://doi.org/10.1016/S0009-2509(03)00095-2

Schwan, H.P., 1957. Electrical Properties of Tissue and Cell Suspensions, in: Advances in Biological and Medical Physics. ACADEMIC PRESS INC., pp. 147–209. https://doi.org/10.1016/B978-1-4832-3111-2.50008-0

Shin, H.S., Lim, H.C., 2006. Cell-mass maximization in fed-batch cultures. Bioprocess Biosyst. Eng. 29, 335–347. https://doi.org/10.1007/s00449-006-0082-z

Shiue, Y.-L., Wang, F.-S., Lee, W.-C., 1995. Parameter estimation and sensitivity analysis for batch fermentation of recombinant cells. Biotechnol. Tech. 9, 891–896. https://doi.org/10.1007/BF00158542

Shuler, M.L., Kargi, F., 2002. Bioprocess Engineering Basic Concepts. Prentice Hall PTR, Upper Saddle River, NJ, pp. 155–218.

Simutis, R., Lübbert, A., 2015. Bioreactor control improves bioprocess performance. Biotechnol. J. 10, 1115–1130. https://doi.org/10.1002/biot.201500016

Slouka, C., Wurm, D., Brunauer, G., Welzl-Wachter, A., Spadiut, O., Fleig, J., Herwig, C., 2016. A Novel Application for Low Frequency Electrochemical Impedance Spectroscopy as an Online Process Monitoring Tool for Viable Cell Concentrations. Sensors 16, 1900. https://doi.org/10.3390/s16111900

Sokolov, M., Morbidelli, M., Butté, A., Souquet, J., Broly, H., 2018. Sequential Multivariate Cell Culture Modeling at Multiple Scales Supports Systematic Shaping of a Monoclonal Antibody Toward a Quality Target. Biotechnol. J. 13, 1700461. https://doi.org/10.1002/biot.201700461

Sokolov, M., von Stosch, M., Narayanan, H., Feidl, F., Butté, A., 2021. Hybrid modeling — a key enabler towards realizing digital twins in biopharma? Curr. Opin. Chem. Eng. 34, 100715. https://doi.org/10.1016/j.coche.2021.100715

Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., Striedner, G., 2017. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. Biotechnol. J. 12, 1–7. https://doi.org/10.1002/biot.201600546

Sonnleitner, B., 2012. Automated Measurement and Monitoring of Bioprocesses: Key Elements of the M3C Strategy. pp. 1–33. https://doi.org/10.1007/10_2012_173

Streefland, M., Martens, D.E., Beuvery, E.C., Wijffels, R.H., 2013. Process analytical technology (PAT) tools for the cultivation step in biopharmaceutical production. Eng. Life Sci. 13, 212–223. https://doi.org/10.1002/elsc.201200025

Tibayrenc, P., Preziosi-Belloy, L., Ghommidh, C., 2011. On-line monitoring of dielectrical properties of yeast cells during a stress-model alcoholic fermentation. Process Biochem. 46, 193–201. https://doi.org/10.1016/j.procbio.2010.08.007

Turner, C., Gregory, M.E., Thornhill, N.F., 1994. Closed-loop control of fed-batch cultures of recombinant Escherichia coli using on-line HPLC. Biotechnol. Bioeng. 44, 819–829. https://doi.org/10.1002/bit.260440707

Vaidyanathan, U., Moshirfar, M., 2021. Ranibizumab, StatPearls. Treasure Island (FL).

Vinet, L., Zhedanov, A., 2013. Measurement, Monitoring, Modelling and Control of Bioprocesses, Journal of Physics A: Mathematical and Theoretical, Advances in Biochemical Engineering/Biotechnology. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36838-7

Vojinović, V., Cabral, J.M.S., Fonseca, L.P., 2006. Real-time bioprocess monitoring: Part I: In situ sensors. Sensors Actuators, B Chem. 114, 1083–1091. https://doi.org/10.1016/j.snb.2005.07.059

von Stosch, M., Hamelink, J.-M., Oliveira, R., 2016a. Hybrid modeling as a QbD/PAT

tool in process development: an industrial E. coli case study. Bioprocess Biosyst. Eng. 39, 773–784. https://doi.org/10.1007/s00449-016-1557-1

von Stosch, M., Hamelink, J.M., Oliveira, R., 2016b. Toward intensifying design of experiments in upstream bioprocess development: An industrial Escherichia coli feasibility study. Biotechnol. Prog. 32, 1343–1352. https://doi.org/10.1002/btpr.2295

Walsh, G., 2018. Biopharmaceutical benchmarks 2018. Nat. Biotechnol. 36, 1136–1145. https://doi.org/10.1038/nbt.4305

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32, 992–1000. https://doi.org/10.1038/nbt.3040

Wang, J., Ye, J., Yin, H., Feng, E., Wang, L., 2012. Sensitivity analysis and identification of kinetic parameters in batch fermentation of glycerol. J. Comput. Appl. Math. 236, 2268–2276. https://doi.org/10.1016/j.cam.2011.11.015

Warikoo, V., Godawat, R., Brower, K., Jain, S., Cummings, D., Simons, E., Johnson, T., Walther, J., Yu, M., Wright, B., Mclarty, J., Karey, K.P., Hwang, C., Zhou, W., Riske, F., Konstantinov, K., 2012. Integrated continuous production of recombinant therapeutic proteins. Biotechnol. Bioeng. 109, 3018–3029. https://doi.org/10.1002/bit.24584

Wechselberger, P., Sagmeister, P., Herwig, C., 2013. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. Bioprocess Biosyst. Eng. 36, 1205–1218. https://doi.org/10.1007/s00449-012-0848-4

Wechselberger, P., Seifert, A., Herwig, C., 2010. PAT method to gather bioprocess parameters in real-time using simple input variables and first principle relationships. Chem. Eng. Sci. 65, 5734–5746. https://doi.org/10.1016/j.ces.2010.05.002

WHO: biologicals, 2018.

Yamane, T., Kume, T., Sada, E., Takamatsu, T., 1977. A Simple Optimization Technique for Fed-batch Culture : Kinetic Studies on Fed-batch Cultures (VI). J. Ferment. Technol. 55, 587–598.

Yardley, J.E., Kell, D.B., Barrett, J., Davey, C.L., 2000. On-Line, Real-Time

Measurements of Cellular Biomass using Dielectric Spectroscopy. Biotechnol. Genet. Eng. Rev. 17, 3–36. https://doi.org/10.1080/02648725.2000.10647986

Zafira, I.Z., Nandong, J., 2019. Optimal feeding strategy of Cephalosporin C fermentation. IOP Conf. Ser. Mater. Sci. Eng. 495. https://doi.org/10.1088/1757-899X/495/1/012107

Zhang, D., Del Rio-Chanona, E.A., Petsagkourakis, P., Wagner, J., 2019. Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. Biotechnol. Bioeng. 116, 2919–2930. https://doi.org/10.1002/bit.27120

Zhong, J.J., Tang, Y.J., 2004. Submerged cultivation of medicinal mushrooms for production of valuable bioactive metabolites. Adv. Biochem. Eng. Biotechnol. 87, 25–59. https://doi.org/10.1007/b94367

Ziegler, J.G., Nichols, N.B., 1942. Optimum Sttings for Automatic Controllers. Transacction of the A.S.M.E 64, 759–768. https://doi.org/10.1115/1.2899060

Zitzmann, J., Weidner, T., Eichner, G., Salzig, D., Czermak, P., 2018. Dielectric Spectroscopy and Optical Density Measurement for the Online Monitoring and Control of Recombinant Protein Production in Stably Transformed Drosophila melanogaster S2 Cells. Sensors 18, 900. https://doi.org/10.3390/s18030900

Zürcher, P., Sokolov, M., Brühlmann, D., Ducommun, R., Stettler, M., Souquet, J., Jordan, M., Broly, H., Morbidelli, M., Butté, A., 2020. Cell culture process metabolomics together with multivariate data analysis tools opens new routes for bioprocess development and glycosylation prediction. Biotechnol. Prog. 36. https://doi.org/10.1002/btpr.3012

Zydney, A.L., 2015. Perspectives on integrated continuous bioprocessing - opportunities and challenges. Curr. Opin. Chem. Eng. 10, 8–13. https://doi.org/10.1016/j.coche.2015.07.005

# RESEARCH OUTPUT

*Peer-reviewed Journals*

- **Swaminathan, N**., Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2022. Cole-Cole modeling of real-time capacitance data for estimation of cell physiological properties in recombinant Escherichia coli cultivation. Biotechnol. Bioeng. 119, 922–935. https://doi.org/10.1002/bit.28028
- **Swaminathan, N.,** Priyanka, P., Rathore, A.S., Sivaprakasam, S., Subbiah, S., 2020. Multiobjective Optimization for Enhanced Production of Therapeutic Proteins in Escherichia coli: Application of Real-Time Dielectric Spectroscopy. Ind. Eng. Chem. Res. 59, 21841–21853. https://doi.org/10.1021/acs.iecr.0c04010

*Patent filed*

- Senthilmurugan Subbiah, Senthilkumar Sivaprakasam, Anurag S. Rathore, **Nivedhitha Swaminathan**, and Priyanka, "A robust system for the real-time estimation of physiological properties of biomass using dielectric spectroscopy". (*Indian Patent Application No. 202131018107*, Filing date: April 19, 2021)

*Conferences*

- **Swaminathan, N.,** Priyanka, P., Rathore, A. S., Sivaprakasam, S., & Subbiah, S., Multiobjective optimization for enhanced biomass production from *E. coli* fermentation, Bioprocessing India 2019, CFTRI, Mysuru, India. (Poster)
- **Swaminathan, N.,** Advanced process control for continuous bioprocessing of biotherapeutic protein production, Research & Industrial Conclave 2022, IITG, Guwahati, India. (Oral)